# Structural investigation of peptaibols using accelerated molecular dynamics simulations

**PhD dissertation**


## Chetna Tyagi

**Supervisors**:

**Dr. Ferenc Ötvös**
Institute of Biochemistry,
Biological Research Centre

**Dr. László Kredics**
Department of Microbiology, TTIK,
University of Szeged

## Doctoral School of Biology

**Department of Microbiology**

**Faculty of Science and Informatics**

**University of Szeged**


**2020**

**Szeged**

# Table of contents

# 1. ABBREVIATIONS

accelerated MD (**aMD**)
acetylated N-terminus (**Ac**)
alamethicin (**Alm**)
antiamoebin I (**Aam-I**)
antimicrobial peptides (**AMPs**)
α-aminoisobutyric acid (**Aib**)
cartesian coordinate PCA (**cPCA**)
classical molecular dynamics (**cMD**)
dihedral principal component analysis (**dPCA**)
D-isovaline (**Div**)
diffusion coefficient (**DC**)
1,2-dimyristoyl-sn-glycero-3-phosphocholine (**DMPC**)
1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (**DOPE**)
1,2-dioleoyl-*sn*-glycero-3-[phospho-rac-(1-glycerol)] (**DOPG**)
electrostatic potential (**ESP**)
free energy landscape (**FEL**)
graphical processing unit (**GPU**)
hydroxyproline (**Hyp**)
Kullback-Leibler divergence (**KLD**)
leucinol (**Leuol**)
lipopolysaccharide (**LPS**)
Liquid Chromatography-Mass Spectrometry (**LC-MS**)
molecular dynamics (**MD**)
non-ribosomal peptide synthetases (**NRPSs**)
nuclear magnetic resonance (**NMR**)
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (**POPC**)
phenylalaninol (**Pheol**)
phosphatidylethanolamine (**PE**)
phosphatidylglycerol (**PG**)
potential-of-mean-force (**PMF**)
principal component (**PC**)
principal component analysis (**PCA**)
radius of gyration (**RoG**)
Research Collaboratory for Structural Bioinformatics: Protein DataBank (**RCSB:PDB**)
RESP ESP charge derive (**RED**)
restrained electrostatic potential (**RESP**)
root-mean-square-deviation (**RMSD**)
root-mean-square-fluctuation (**RMSF**)
trikoningin KA-V (**TRK-V**)
tripleurin XIIc (**TPN XIIc**)
valinol (**Vol**)
Visual molecular dynamics software (**VMD**)
zervamicin IIb (**Zrv-IIb**)

# 2. INTRODUCTION

With the growing instances of antibiotic resistance all over the world, the scientific community is, more than ever, desperate to identify novel, fail-safe ways of treatments for a plethora of devastating diseases. This search is not limited to human pathogens but has also been extended to the problem of agricultural pathogens. The class of antimicrobial peptides (AMPs), defined as host-defense molecules found in all life forms, is a promising solution. Many AMPs have already crossed over to clinical trials as novel therapeutics, immunity modulators and wound healing promoters (Mahlapuu et al., 2016; Zhang et al., 2019). Bioactive peptides like peptaibols, a special class of fungal AMPs containing non-standard residues like α-aminoisobutyric acid (Aib) or hydroxyproline (Hyp), have piqued the interests of microbiologists owing to their antibacterial, antifungal, antiviral, antihelminth, and antitumor properties (Yun et al., 2000; Schiell et al., 2001; Berg et al., 2003; Leitgeb et al., 2007; Ayers et al., 2012; Mohamed-Benkada et al., 2016) as well as their abilities to elicit plant defense responses (Engelberth et al., 2000; Chakraborty et al., 2020).

In our studies we have focused on peptaibols, produced by a specific group of fungi, the genus *Trichoderma*, which comprises of more than 300 species (Bissett et al., 2015; Zhang and Zhuang, 2018) and are known for antagonistic behavior against naturally competing fungi (Reino et al., 2008). They deploy a number of mechanisms to wipe out competing fungal phytopathogens, which mainly depend on the production of cell wall degrading enzymes (Elad et al., 1980; Javeria et al., 2020) and other secondary metabolites like peptaibols, meanwhile promoting plant root growth (Olabiyi & Ruocco, 2013; Kushwaha & Verma, 2014). Iron competetition was also observed as a mode of action of *Trichoderma asperellum* which produces iron-binding siderophores which controls Fusarium wilt (Segarra et al., 2010). The scientific community has recognized this property to develop *Trichoderma* strains into successful biocontrol agents. For example, several plant pathogens like *Nectria galligena* in apples, *Sclerotium rolfsii* in tobacco, bean and iris, *Rhizoctonia solani* in radish, strawberry, cucumber, potato and tomato, *Chondrosterum purpureum* in stone-fruit and other crops, or *Botrytis cinerea* in apple have been controlled using commercially available strains of *Trichoderma* (Cutler & Cutler 1999; Reino et al., 2008). *Trichoderma koningiopsis* has been studied as a biocontrol agent against a specific mushroom, *Leucoagaricus gongylophorus*, the main food source of leaf-cutting ants (Castrillo et al., 2016). *Trichoderma atroviride* showed most promising results in terms of disease reduction (*Rhizoctonia solani*, *Pythium ultimum* and *Sclerotinia trifoliorum*) and plant growth promotion (Kandula et al.,

2015). Similarly, Chen et al. (2016) successfully showed the biocontrol efficacy of *Trichoderma gamsii* against the root-rot pathogenic fungi of *Panax notoginseng*. It has also been proven to be effective against Fusarium head blight disease of wheat (Baroncelli et al., 2016). Rinu et al. (2014) isolated *T. gamsii* from the roots of lentil (*Lens esculenta* Moench) and tested its efficacy against seven phytopathogens, out of which it was effective against six. The economic importance of the genus *Trichoderma* lies also in being a source of antibiotics, plant growth promoters, commercial biofungicides, as well as commercial enzymes used in recycling cellulosic waste – achieved mainly by *Trichoderma reesei* belonging to the clade Longibrachiatum (Harman & Kubicek, 1998; Kubicek et al., 2009). Moreover, a product marketed as TUSAL made from *Trichoderma harzianum* and *Trichoderma viride* was developed by the phytopathology research group of the University of Salamanca and Newbiotechnic S.A. Corporation against a leaf-falling disease affecting multiple crops (Reino et al., 2008).

On the other hand, *Trichoderma* species can be harmful against cultivated mushrooms such as *Agaricus bisporus* and *Pleurotus ostreatus*, commonly known as the white button mushroom and oyster mushroom, respectively. The production of both crops can be seriously affected by the so-called green mold diseases caused by certain members of the mycoparasitic fungal genus *Trichoderma* such as *Trichoderma aggressivum*, *Trichoderma pleuroticola* and *Trichoderma pleuroti* (Hatvani et al., 2008; Kredics et al., 2010). Moreover, *Trichoderma* strains, mostly from clade Longibrachiatum, have also been reported in an increasing number of cases as etiologic agents in human infections (Molnár-Gábor et al., 2013; Hatvani et al. 2013). *T. harzianum* was detected in the postmortem examination of a renal transplant recipient (Guarro et al., 1999). These *Trichoderma* strains are present in environmental habitats including agricultural systems and they can be potential sources of human infections in immunocompromised patients. The idea to use *Trichoderma longibrachiatum* as a biocontrol strain appeared in tropical countries, as members of the Longibrachiatum clade are able to grow at higher temperature unlike other *Trichoderma* species (Zhang et al. 2014; 2015). Although instead of using the whole organism, an alternative option would be to use only their metabolites with plant growth promotion or biocontrol activities.

## 2.1. Peptaibols

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 Aib, Hyp, D-isovaline (Div), and C-terminal alcohol residues like phenylalaninol (Pheol), valinol (Vol), etc., along with an acetylated N-terminus (Ac) are characteristic for these peptides ranging 7-20 amino acid residues in length (Benedetti et al., 1982; Brückner & Graf, 1983; Degenkolb & Brückner, 2008). 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 (Zocher & Keller, 1996; Marahiel et al., 1997; Wei et al., 2005). Due to the relaxed specificity of NRPSs for residues, the peptaibols produced by a certain fungal strain may show considerable heterogeneity (Miller & Gulick, 2016). The diversity in their sequence length, hydrophobicity, antimicrobial properties and producing fungal species contribute to a plethora of peptaibols that are yet to be discovered and studied. The increasing gap between known peptaibol sequences and their three-dimensional structures can be reduced using computational modeling and molecular dynamics simulations. The knowledge of peptide structural dynamics is a key to unraveling the mechanisms of their antimicrobial action.

### 2.1.1. Mode of action

Peptaibols show antibiotic properties by forming pores within biological membranes via aggregation, which has been widely studied (Fox & Richards, 1982; Karle et al., 1991; Shenkarev et al., 2002). Apart from studies on bioactivity, substantial focus has been given to their three-dimensional structures and folding dynamics (Condamine et al., 1998; Snook et al., 1998; Balashova et al., 2000; Kronen et al., 2003). Only 17 peptaibols have been studied using experimental methods in different solvents which includes alamethicin [Protein Databank (PDB) ID: 1AMT] (Fox & Richards, 1982), antiamoebin I (Aam-I) [PDB IDs: 2M1F, 1GQ0, 1JOH] (Snook et al., 1998; Galbraith et al., 2003; Shenkarev et al., 2013), bergofungin A [PDB ID: 5MAS] (Gessmann et al., 2017), cephaibol A, B and C [PDB IDs: 1OB4, 1OB6, 1OB7, respectively] (Bunkóczi et al., 2003), chryospermin C [PDB ID: 1EE7] (Anders et al., 2000), gichigamin [PDB ID: 4Z0W] (Du et al., 2017), harzianin HK-VI [PDB IDs: 5M9Y, 5MF3, 5MF8], paracelsin-X [PDB ID: 4BY8], samarosporin I [PDB IDs: 4G13, 4G14] (Gessmann et al., 2012a), trichotoxin [PDB ID: 1M24] (Chugh et al., 2002), trichovirin [PDB ID: 3SBN] (Gessmann et al., 2012b), zervamicin IIB (Zrv-IIb) [PDB IDs: 1DLZ, 1IH9, 1R9U] (Balashova et al., 2000; Shenkarev et al., 2002; Ovchinnikova et al., 2003) (**Figure 1**).

Alamethicin (Alm) is the most studied peptaibol so far, closely followed by Aam-I and Zrv-IIb. The discovery of Alm is credited to Meyer & Reusser (1967), who referred it as "antibiotic U-22324" obtained from '*T. viride'* and classified as a cyclic peptide due to its inability to react with ninhydrin. The antibacterial activity against Gram-positive strains was highlighted. The correct producer was reidentified later to be *Trichoderma arundinaceum* from the Brevicompactum clade of genus *Trichoderma* (Degenkolb et al., 2008). The experimentally observed (X-ray diffraction) three-dimensional structure of Alm F30/3 is available in the RCSB PDB database with accession ID 1AMT (**Figure 1**).

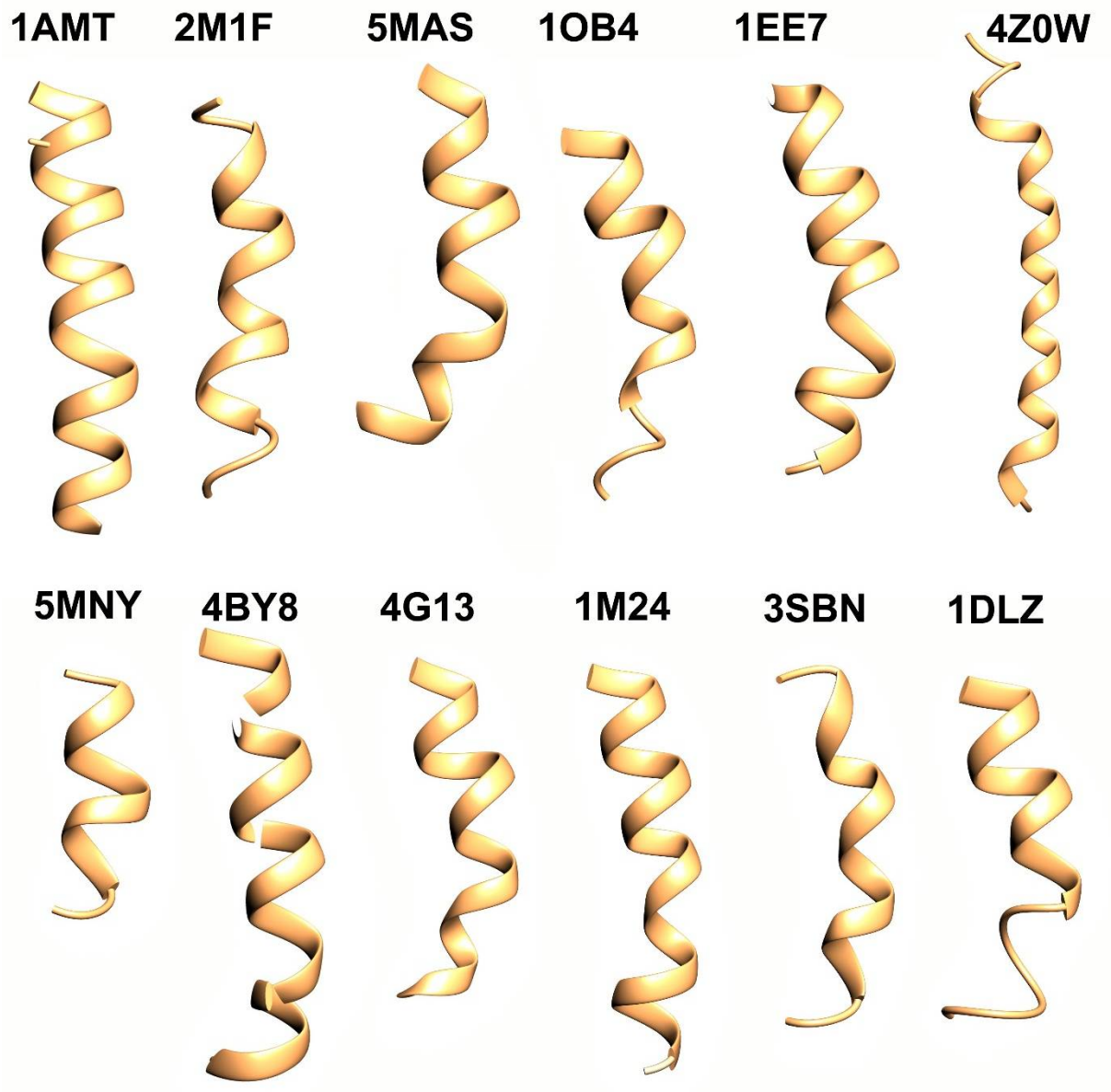


**Figure 1.** The experimentally determined three-dimensional structures of various peptaibols obtained from the RCSB PDB database. Their respective PDB ID codes are labelled.

The first possible primary structure of Alm was reported by Payne et al. (1970), who described it as a cyclic molecule by linking the γ-carboxylate group of the glutamic acid residue to the first proline in the sequence, which was reported as *Pro-Mea-Ala-Mea-Ala-Gln-Mea-Val-Mea-Gly-Leu-Mea-Pro-Val-Mea-Mea-Glu-Gln*, where Mea is Aib. They hypothesized a stack-like tunnel formation by the cyclic Alm structure with hydrophobic interior. Martin & Williams (1976) later corrected it by describing the structure of Alm as a linear polypeptide by nuclear magnetic resonance (NMR) spectroscopy. The new sequence assignment included an acetylated Aib residue at the N-terminus and a phenylalaninol as a side-chain of the 18$^{th}$ Gln residue. They stressed upon the importance of linear Alm structure to be long enough to stretch across a lipid bilayer and rejected the idea of stacked-ring pores. The presence of proline in the 14$^{th}$ position introduces a slight bend in the structure as shown by X-ray crystallography (Fox & Richards, 1982), NMR (Haris & Chapman, 1988) and optical spectroscopy (Nagao et al., 2015). The helix is formed in such a manner that the polar residues are arranged on one side. Alm is classified as amphipathic due to distinct hydrophobic and hydrophilic faces, which renders the ability to either interact with a membrane horizontally or form voltage-gated ion channels with a vertical insertion. They have been shown to induce lipid flip-flop even in a surface-bound state by disordering lipids in the membrane (Taylor et al., 2019). The recent studies to understand Alm conformation and pore formation have been reviewed by Leitgeb et al. (2007), Kredics et al. (2013).

Aam-I and Zrv-IIb acquire very similar structural topologies in membrane-mimicking environments, although significantly differ at the N-termini (1-8 residues) but show markedly different dynamic properties that makes Aam-I to be highly soluble in water but less bioactive than Zrv-IIb (Shenkarev et al., 2013). The crystal structures were determined for Bergofungin A produced after replacing the amino-terminal phenylalanine in Samarosporin I to a valine which results in an inactive peptide in place of a moderately active Samarosporin I (Gessmann et al., 2017). A single residue substitution may, thus, result in increase or decrease in bioactivity. Similarly, the crystal structure of 22-mer gichigamin was determined to be an unusual right-handed $3_{11}$ helical structure which confers it the ability to penetrate mammalian membranes. This potency could be increased by introducing a semisynthetic modification of C-terminus linked coumarin derivative (Du et al., 2017). Cephaibols A, B and C contain all isovalines in the D configuration and form helical structures with a sharp bend of 55˚ at the central hydroxypoline (Bunkóczi et al., 2003) and shows pronounced anti-helminthic action and activity against ectoparasites (Schiell et al., 2001).

## 2.2. Molecular dynamics simulations

The attempts to computationally model biomolecular systems were carried out to complement experimental observations, providing missing details, verification of experimental results, and generating new hypotheses and experiments based on these new results. It has been 43 years since the first molecular dynamics (MD) simulation of a macromolecule of biological interest, bovine pancreatic trypsin inhibitor, was published (McCammon et al., 1977). With passing time and requirement, the need to accurately represent a biological system and simulate for significantly long timescales has been met with technological advancement (Dror et al., 2012; Lane et al., 2013). MD simulations are carried out to recreate and visualize various biomolecular processes like protein folding/misfolding (Snow et al., 2005; Scheraga et al., 2007), protein-drug interaction and stability (Swegat et al., 2003; Tyagi et al., 2015), protein-protein binding (Grünberg et al., 2006), DNA/RNA folding (Cheatham et al., 1995; Cheatham et al., 1997), bilayer membranes and their interaction with transmembrane proteins/peptides or receptors (Ayton & Voth, 2009; Balali-Mood et al., 2009) etc., with the help of potentials formulated on the basis of physical laws that govern all chemical entities. It is hoped to observe and understand the properties of assemblies of molecules in terms of their structure and the microscopic interactions between them. The MD simulation technique is distinct in a way that it calculates the time evolution of properties of the macro-molecular system which is called as a trajectory (Lindorff-Larsen et al., 2012).

### 2.2.1. Principle

MD simulations are based on Newton's second law of motion, $F = ma$, where $F$ is the force exerted on the particle, $m$ is the mass and $a$ is its acceleration. It is possible to determine the acceleration of each atom in a system if the forces acting on these atoms are known. Numerical integration of the equation of motion yields a trajectory that describes the positions, velocities and accelerations of particles over time (Haile, 1992; Petrenko & Meller, 2001; Petrenko & Meller, 2010). To calculate such a trajectory, one only needs the initial position of atoms and the initial distribution of velocities (Frenkel & Smit, 2001). This initial information can either be obtained from experimental structural data like X-ray crystallography and NMR, a computationally modeled protein structure based on template homology or a randomly generated string of atoms as can be done using the *leap* module of AmberTools18.

MD simulations run on the accuracy of potentials that may be calculated using Born-Oppenheimer approximation or quantum mechanics (Vanommeslaeghe et al., 2010). The Born-Oppenheimer approximation states that the motion of the nuclei and electrons is so vastly different owing to the great difference between their masses and the resulting acceleration that it is possible to describe the electrons in a molecule while neglecting the motion of atomic nuclei. The second approximation treats nuclei as point particles that follow Newtonian dynamics. Force fields or potentials are empirical in nature and can be classified as bonding potentials (bonded forces) including bond lengths, bond angles and dihedral angles and non-bonding interactions comprising of short-range interactions (van der Waal potentials) and long-range interactions (Columbic potentials). A potential can be described as (Paquet & Viktor, 2015):

$$\begin{aligned}\mu(r^N) = &\sum_d k_d(d-d_0)^2 + \sum_s k_s(S-S_0)^2 \\ &+ \sum_\theta k_\theta(\theta-\theta_0)^2 \\ &+ \textstyle\sum_\chi k_\chi\big(1+cos\,(n\chi-\delta)\big) \\ &+ \sum_\varphi k_\varphi(\varphi-\varphi_0)^2 \\ &+ \textstyle\sum_{i,j} \varepsilon_{ij}\left(\left(\frac{r_{ij}^0}{r_{ij}}\right)^{12} - \left(\frac{r_{ij}^0}{r_{ij}}\right)^{6} + \frac{q_i}{\varepsilon_l}\frac{q_j}{r_{ij}}\right)\end{aligned} \qquad \text{Eq. 1}$$

where $d$ is the bound length, $S$ is the Urey-Bradley bound length, $\theta$ is the bond angle, $\chi$ is the dihedral angle, $\varphi$ is the improper dihedral angle, $r_{ij}$ is the distance in between atom $i$ and $j$, $k_d$, $k_s$, $k_0$, $k_\chi$, and $k_\varphi$ are constants, $d_0$, $S_0$, $\theta_0$, $\varphi_0$, and $r_{ij}$ are equilibrium positions, $\varepsilon_{ij}$ is related to the Lennard-Jones well depth, and $\varepsilon_l$ is the effective dielectric constant. Finally, $q_i$ is the partial atomic charge associated with atom $i$: the partial charge comes from the asymmetrical distribution of the electrons in the chemical bounds. The first term on the last line is the van der Waal interaction (or Lennard-Jones potential as described by Smit & Frenkel, 1991), and the last term in the last line is the Columbic interaction. This equation including the atom and bond specific parameters are known as “force field” in the computational modeling world. The calculation of long range/non-bonded interactions like Columbic interaction is generally computationally exhaustive therefore, to rectify the cost,

multiple numerical approximations such as shifted cutoff radii, reaction field algorithms, particle mesh Ewald summation (PME; Essmann et al., 1995), etc. are employed in the force fields. The force fields also contain parameters for different types of atoms, bonds, torsions, angles, partial charges for individual atoms etc, which are calculated either empirically (Class I force fields) or derived by quantum mechanical calculations (Class II force fields). The current force fields assign one static value of charge for an atom irrespective of its electrostatic environment while the work for accurate polarizable force fields is still underway (Halgren & Damm, 2001).

### 2.2.2. The Amber force field ff14SB

It is crucial to understand the force field used for all calculations in this work. We used ff14SB (Maier et al., 2015) which is a continuing evolution of the ff99SB force field. The older ff94 and ff99 parameter sets did not provide a good energy balance between helical and extended regions of peptide and protein backbones. Another problem is that many of the ff94 variants had incorrect treatment of glycine backbone parameters. ff99SB improved this behavior, presenting a careful reparameterization of the backbone torsion terms in ff99 and achieves much better balance of four basic secondary structure elements (PP II, β, αL, and αR) (Hornak et al., 2006). In addition, it corrected the glycine sampling and should also perform well for β-turn structures, two things which were especially problematic with most previous Amber force field variants. The changes mainly involve torsional parameters for the backbone and side chains. For backbones, experimental scalar coupling data for small solvated peptides became available (Graf et al., 2007) against which ff99SB was compared. Ff14SB has been shown to enhance reproduction of experimentally indicated geometries over ff99SB due to minimum dependence of ff14SB side chain parameters on particular backbone conformations. Therefore, ff14SB provides the latest updates to side-chain dihedral corrections for lysine, arginine, glutamate, glutamine, methionine, serine, threonine, valine, tryptophan, cysteine, phenylalanine, tyrosine, and histidine.

### 2.2.3. Accelerated Molecular Dynamics

As learned from previous sections, the MD simulation technique is an unparalleled predictive technique in theoretical chemistry, physics and biology amongst others. But for studying most biological systems of interest, classical MD simulations fall short of the time scale required to obtain slow conformational transitions separated by high energy barriers. In other words, the energy landscape of a biological molecule may have multiple minimum energy wells separated by high energy barriers as described in **Figure 2,** that may only be accessed

by simulation timescales run over several orders of magnitude. The blue regions denote energy minima with the corresponding conformation of a peptide, while the red regions are the high energy barriers that must be crossed to reach to different minima. These zones have been defined as funnel-shaped or trough-like energy minimum wells, while the barriers between them can be described as crest-like. During classical MD simulation, a system can get trapped in a local minimum for long period of times which requires powerful computers to process. A notable example of such a supercomputer designed for MD simulations is the "Anton" by D. E. Shaw and his group, a massively parallelized machine that has been succesfully used to simulate biomolecules on micro- to millisecond timescales (Shaw et al., 2007).

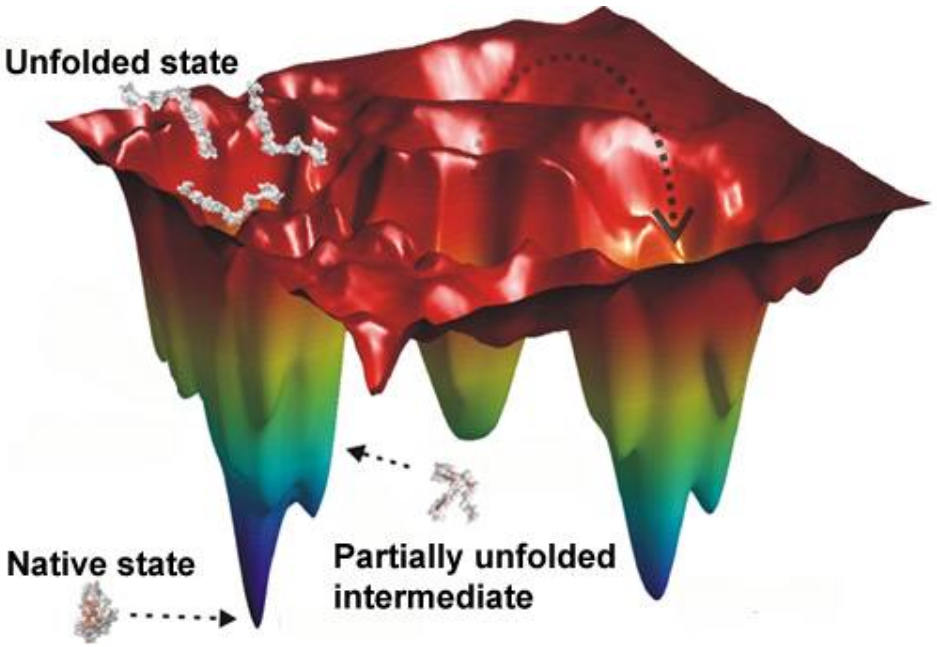


**Figure 2.** The free energy landscape (FEL) of a peptide with potential energy basins or wells in blue that corresponds with the native conformation. Each basin corresponds to a distinct or intermediate conformation of the peptide. The system may get stuck in one basin for long periods of time during classical MD and requires a "boost" to either raise the energy basins or lower the energy barriers. (adapted from Quintas, 2013)

In order to enhance the sampling of conformational landscape of a biological system of interest, many methods have been developed that follow the scheme of modifying the Hamiltonian by adding a bias potential, for example, umbrella sampling (Torrie & Valleau, 1977), parallel replica dynamics (Voter, 1998), hyper-dynamics (Voter, 1997) or temperature-accelerated dynamics (Sörensen & Voter, 2000). Apart from the strategy to develop algorithms, the use of graphical processing units (GPUs) instead of central

processing units (CPUs) through implementation of various molecular modeling software has greatly reduced the time required to achieve significant sampling (Götz et al., 2012; Salomon-Ferrer et al., 2013; Stone et al., 2010). One such method, the aMD has been discussed in detail.

In 2004, building on the work of Voter, 1997, Hamelberg et al. (2004) published a molecular dynamics approach based on altering the amount of computational time a system spends in potential energy minima to move over potential barriers. It works by adding a bias potential $\Delta V(r)$ to the true potential and modifying the potential energy surface $V(r)$ in such a way that the surfaces near the minima are raised but those near the barriers remain unaffected. The technique promotes sampling of infrequent events of biomolecular systems without any prior knowledge of the location of energy wells or barriers.

The bias is then removed from the statistics calculated with the bias potential which is called *reweighting*. The general idea behind aMD can be understood as depicted in **Figure 3**. To quote "A continuous non-negative bias boost potential function $\Delta V(r)$ is defined in such a way that when the true potential $V(r)$ is below a chosen threshold value $E$, the boost energy, the simulation is performed on a modified potential $V^*(r) = V(r) + \Delta V(r)$, represented by dashed lines. When $V(r)$ is greater than $E$, the simulation is performed on the true potential $V^*(r) = V(r)$."

$$V(r) = \begin{cases} V(r), V(r) \geq E \\ V(r) + \Delta V(r), V(r) < E \end{cases} \qquad \text{Eq. 2}$$

Where $\Delta V(r)$ is given by,

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))} \qquad \text{Eq. 3}$$

Where α is a parameter to determine the depth of modified potential energy basin i.e. if α is zero, the modified potential $V^*(r) = E$ i.e. follows the true potential. The choice of $E$ and α parameters are very crucial to determine how aggressively the simulation will be accelerated. The acceleration threshold energy, E, determines which portion of the energy surface is affected by aMD and the acceleration factor, α, determines how smooth the modified potential surface becomes (Hamelberg et al., 2004).

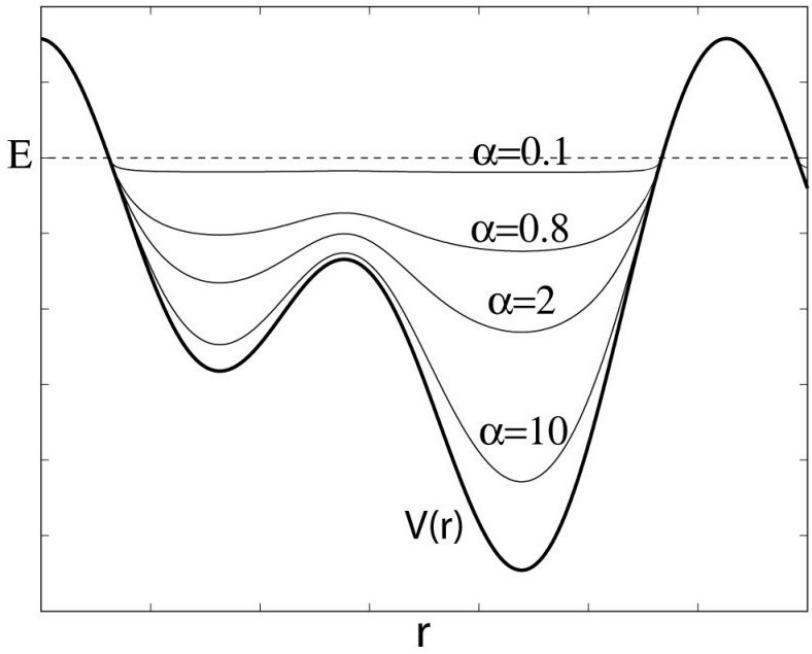


**Figure 3.** A 2D representation of energy basins and barriers between two minima to explain basic principle behind accelerated MD technique. The different values of α parameter raise the energy minima to enable faster transition between states till it remains lower than the threshold energy E. The thin lines depict modified energy potential and smaller barrier between energy basins. (adapted from https://www.ks.uiuc.edu/Research/namd/2.9/ug/node63.html)

The implementation in AmberTools18 to calculate these parameters to set up an aMD is simply based on a set of simple equations. As discussed elsewhere, the aMD implementation includes three kinds of boost to potential, a) boost to total potential of the system, b) boost to dihedral potential of the biological system, and c) dual boost combining the previous two.

The parameters $E_{dihed}$, $\alpha_{dihed}$, $E_{total}$ and $\alpha_{total}$ were calculated as required by Equation 4:

$$E_{dihed} = V_{avg\_dihed} + a_1 \times N_{res}, \qquad \alpha_{dihed} = a_2 \times N_{res}/5;$$

$$E_{total} = V_{avg\_total} + b_1 \times N_{atoms}, \qquad \alpha_{total} = b_2 \times N_{atoms} \qquad \text{Eq. 4}$$

where $N_{res}$ is the number of peptide residues, $N_{atoms}$ is the total number of atoms in the system. $V_{avg\_dihed}$ and $V_{avg\_total}$ are average dihedral and total potential energies obtained from the short classical MD simulations.

aMD has been applied to study a variety of systems like agonist-antagonist binding to a G-protein coupled receptor while placed in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer membrane (Kappel et al., 2015), the dynamic transition between open and closed forms of a crucial Gram-negative bacterial protein, maltose binding protein (Bucher et al., 2011), third IgG-binding domain of Protein G (GB3) in combination with NMR experimental data (Markwick et al., 2007), the simple alanine dipeptide to show that aMD effectively enhances the sampling of the phi/psi conformational phase space and the free energy plots revealed that all minima regions are accurately sampled and the canonical distribution is recovered (De Oliveira et al., 2007), backbone conformation of serine-proline motifs in proteins (Hamelberg et al., 2005) etc.

### 2.2.4. Implicit and explicit solvents

Along with the biomolecule of interest, the immediate environment of the system including water molecules and ions that generally surround them must be modeled accurately but at the same time it should not increase the computational burden excessively. MD simulations can broadly be carried out in vacuum (no solvent), in implicit solvent where the solvent molecules are represented as a continuous potential which reduces the computational burden manifolds, however, it is not considered to be the most accurate representation, and in explicit solvent where discrete solvent molecules surround the biomolecular system in a solvation box. The explicit water solvent is mostly represented as a rigid three-point water model like TIP3P (Jorgensen et al., 1983) or SPC (Berendsen et al., 1981). Many times, a solvent other than water can be used, for example, chloroform can be used to simulate a hydrophobic environment as in a bilayer membrane. With the advent of modern processing machines, it is now feasible and expected to employ explicit representation of the solvents.

### 2.2.5. Characterization of secondary structure through Ramachandran plots

To characterize the secondary structure evolution of a peptide through the course of simulation and in different solvents, Ramachandran plots were produced. Ramachandran, (1963) published a study describing different protein secondary structural conformations as a function of dihedral angles of constituting residues. The scatter plot of these values enabled them to define “allowed” and “disallowed” regions on a plot which later became famous as the Ramachandran plot. In their definition, they discussed α-helix and β-sheet regions, left-handed helix regions, π-helices, right and left γ-helices (sterically defined as unlikely to occur in polypeptides) and ribbon structure/$2.2_7$ helices (first described by Donohue, 1953).

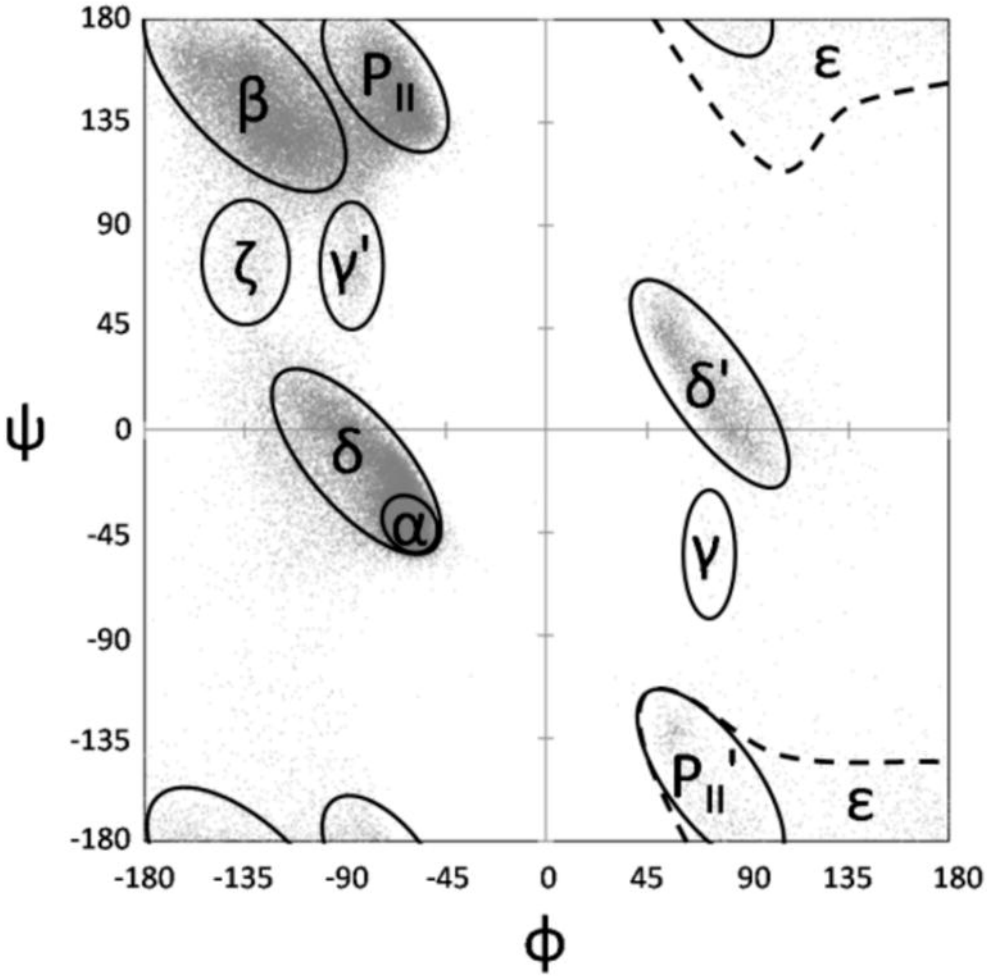


**Figure 4.** The various secondary struuctural regions from Hollingsworth & Karplus nomenclature mapped on a classical Ramachandran Plot.

Fast forward to few decades, Hollingsworth & Karplus (2010) described five main conformation clusters on the plot, namely, α-helices, β-strands, polyproline II ($P_{II}$)-spirals, γ-turns and γ′-turns along with an additional ε-region (later named $P_{II}$' region) generally populated by glycine residues and the bridge region between α-helices and β-strands (δ and δ′) (**Figure 4**, **Table 1**). The mirrored δ′ corresponds to the left-handed α-helix region in this work. Finally, an additional region ϕ, ψ = (−130, +80), named as ζ is mentioned mostly for residues preceding proline in the sequence. However, when applying these rules to linear groups, i.e., series of residues with same repeating conformation, it became clear that only three broad types occur including α- and $3_{10}$-helices, a group that is largely made of β-strands and a group that adopt a $P_{II}$ spiral conformation, ϕ, ψ = (−65, +145). They completely dismissed the presence of $2.2_7$ helix regions in putative linear groups. Left-handed helices and $3_{10}$ helices may exist as short segments. While listing fewer regular structures, they mentioned β-turns, $P_{II}$-spirals, γ- and γ′-turns, ζα and ζ$P_{II}$ regions. The bridge region or the δ

region constitutes various types of turns like Type I, II', III etc. which have been classically called as β-turns. The term "β-turns" to describe the δ region structure is used throughout this text and should not be confused.

**Table 1.** Average Φ and Ψ angle values for various conformations on the Ramachandran plot (according to Hollingsworth & Karplus, 2010).

| **Type** | **ϕ, ψ angles** |
|---|---|
| α (alpha helix) | −63, −43 |
| β (beta region) | –157.2, 161.9 |
| $P_{II}$ -spirals | −65, +145 |
| γ-turns | +80, −80 |
| γ′-turns | −80, +80 |
| δ regions | Extending at 45˚ angle to the left of α-helix region |
| δ′ regions | Mirror image of δ region |
| ζ (pre-proline region) | −130, +80 |

### 2.2.6. Addressing convergence

#### 2.2.6.1. *Kullback-Leibler Divergence method*

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. Because they are now known to display varying degrees of conformational dynamics instead of adopting a single conformation (McClendon et al., 2012) which arises the question whether all possibe conformations have been attained. The principal components can give an idea of the conformational states that have been visited during the simulation. A quantitative measure of extent of overlap between any probability distribution is the Kullback-Leibler divergence (KLD) method taken from information theory which can indicate satisfactory sampling (Kullback & Leibler, 1951). By measuring the overlap of PC histograms as a function of simulation time, we can assess the convergence of dynamic properties of simulations. KLD is defined as:

$$KLD(t) = \sum_{i=0}^{M} ln\left(\frac{hPC1_N(t,i)}{hPC2_N(t,i)}\right) hPC1_N(t,i) \qquad \text{Eq.5}$$

where $hPCX_N(t,i)$denotes bin *i* of the histogram from trajectory *X* for the projection of PC *N* using data from frames 0 to *t,* and *M* is the total number of histogram bins (400 in this case). In other words, two independent simulations started from different configurations should

eventually begin to sample the same conformations (Galindo-Murillo et al., 2015). The extent of overlap between such simulations, quantified as the KLD value as a function of time, will indicate whether satisfactory sampling has been attained.

#### *2.2.6.2. Good-Turing method*

To address the question of convergence, a relatively new technique based on probability theory reported by Koukos et al. (2014) was applied to compare convergence of peptaibol folding. They argued that any meaningful way to deduce convergence must be based on probability theory. They devised Good-Turing formalism for estimation of frequency of unobserved species (of conformations) in a trajectory. This formalism draws its conclusion on the classical root-mean-square-deviation (RMSD) matrix by answering the following question "What is the probability that a molecular configuration with an RMSD (from all other already observed configurations) higher than a given threshold has not actually been observed?". In simple terms, this algorithm starts by quantifying structural distance in form of RMSD by creating an RMSD matrix (original matrix [N x N]) from the trajectory. This maximum RMSD is the highest RMSD observed between successive structures in the original matrix, i.e. sampling factor of 1. A sub-matrix formed of size [N/2 x N/2] from the original matrix by taking every second row and column and maximum RMSDs determined on its superdiagonal means that sampling factor of 2 was applied. For real time application, the trajectories should be spaced in a way that the sampling factor stays close to 1. The resulting matrix is considered as a distance matrix and a dendrogram is constructed using hierarchical clustering methods (Shao et al., 2007). This dendrogram is used to produce clusters at different RMSD cutoffs along with their frequencies. Finally, the Good-Turing formalism is applied to these frequencies to calculate the probability of unobserved species (P_unobs.) that could be observed if the simulations were extended. In simple terms, this analysis indicates the probability of observing a molecular configuration not observed before (Serafeim et al., 2016).

## 2.3. Biological Membranes

It is crucial to discuss the composition and properties of biological membranes, especially those of bacterial and fungal origin, as they are the main targets of peptaibol activity. Most of the bacteria and all fungi are surrounded by a cell wall. Amongst bacteria, the Gram-positive bacteria possess a thicker cell wall than Gram-negative bacteria, which compensates for the lack of outer membrane in the former. Many *Trichoderma* species are known to produce cell wall lysing enzymes amongst other secondary metabolites (Gajera et al., 2012; Gruber & Seidl-Seiboth, 2012). Gram-negative bacteria have two membranes called as cytoplasmic or inner membrane and the outer membrane. The top monolayer of the outer membrane consists of a lipid layer only found in Gram-negative bacteria and known as lipopolysaccharide (LPS) (Beutler, 2002). The bottom monolayer of the outer membrane, the cytoplasmic membrane of Gram-negative bacteria and the membrane of Gram-positive bacteria are all composed of lipids, phosphatidylglycerol (PG), phosphatidylethanolamine (PE) and cardiolipin (Epand et al., 2016). The main role of PE in bacterial membrane is to spread out the negative charge, assembling of membrane proteins and help the membrane proteins to fold correctly. Another important lipid species, PG is found on average as 25% of the total lipid composition of bacteria ranging as high as 80% in *S. aureus* to 20% in *E. coli* (Dowhan, 1997). Because of the presence of an extra outer layer in Gram-negative bacteria, they are less susceptible to the effect of membrane-acting antimicrobial agents. Although, here the outer membrane shows higher permeability than the cell membrane due to the presence of porin proteins that are permeable for molecules with masses of 500 Da or less. The closest to this size is the shortest peptaibol recorded as peptaibolin with 5 amino acid residues (Huelsmann et al., 1998; Crisma et al., 2001). Damage to the outer membrane by removal of $Mg^{2+}$ ions may even result in passing of larger molecules (Lam et al., 2014). The understanding of differences between the morphology of Gram-positive and Gram-negative bacterial membranes is crucial to develop potent antibacterial peptaibol agents. The most common mechanism of targeting cell membrane by antimicrobial agents is through altering its bulk properties (Hu et al., 2015; Mularski et al., 2015; Voievoda et al., 2015) like membrane curvature (Matsuzaki et al., 1998; Bozelli Jr et al., 2012; Koller & Lohner, 2014; Perrin et al., 2015), lipid clustering (Epand et al., 2008; Epand et al., 2009; Epand & Epand, 2009; 2011), packing defects (Epand et al., 2016) or direct target of a bilayer component (Machaidze et al., 2002; Pogliano et al., 2012). The different mechanisms are depicted in **Figure 5**.

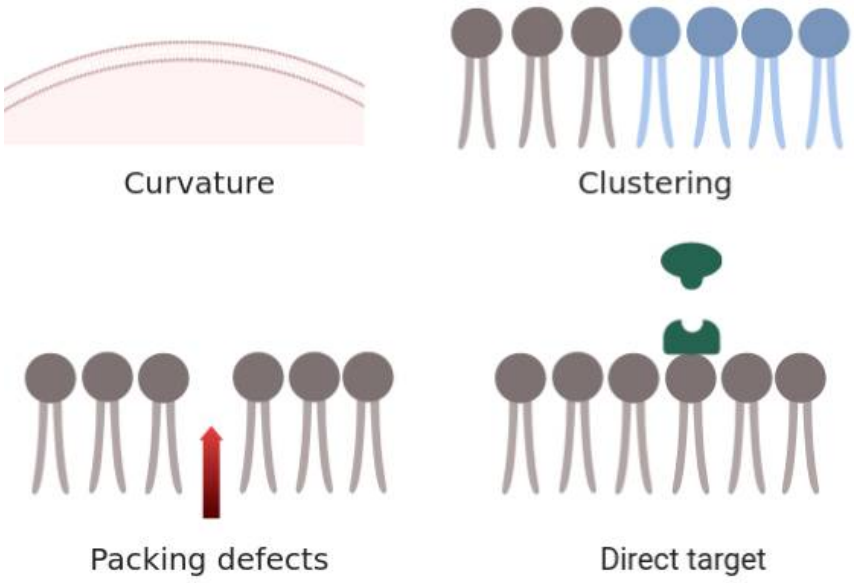


**Figure 5.** Various mechanisms by which antimicrobial compounds may target lipid membranes (prepared using Biorender)

While all listed mechanisms are important in their own way, the last mechanism of "direct action" against specific lipids of the membrane remains the most important avenue to explore for rational antimicrobial discovery. By rational, we mean that an antimicrobial agent must be so designed to have highest efficacy against the targeted bacteria while being the least toxic to the human host. The agents targeting membranes must target those lipid species found in abundance in the bacterial membrane while absent or meagerly present in mammalian membranes. For example, targeting the anionic lipids and PE which are mostly found exposed on the outer surface of bacterial membranes in abundance while found on the cytoplasmic surface of eukaryotic membranes with a low concentration of 5% (Epand et al., 2016). The first peptides discovered to target PE were members of class I, type B lantibiotics (a class of polycyclic peptide antibiotics that contain the characteristic thioether amino acids lanthionine or methyllanthionine, as well as the unsaturated amino acids dehydroalanine, and 2-aminoisobutyric acid) like duramycin, and the resistance to it in *Bacillus subtilis* could be associated with the decrease in amount of PE and cardiolipin in the membrane (Dunkley Jr et al., 1988). For exploiting different avenues of membrane disruption caused by antimicrobial peptides, we must understand the relation between their structure and activities. Therefore, MD simulation techniques are fast, cost-effective and increasingly accurate approaches to elucidate mechanisms of peptaibol folding and to uncover their modes of action against biological membranes.

# 3. AIMS AND OBJECTIVES

The understanding of peptaibol structure and dynamic folding is important to subsequently understand and correlate with their antimicrobial mechanism of action. Peptaibols are known to generally possess a spiral or helical shape and to disrupt biological membranes or form voltage-gated ion channels, as it was proven for alamethicin. However, some of them fold into strict α-helices while others form loose β-bend ribbons. Although most of them exhibit a backbone bend to some degree, a few peptaibols show a highly curved backbone. These differences account for the variability in their ability to interact with cell membranes and therefore, their antimicrobial activity. Our main aim is to develop a reliable strategy for elucidating complete structural ensembles of peptaibol compounds by applying time- and cost-effective *in silico* techniques known as molecular dynamics simulations. The various steps taken in lieu of the main goal are as follows:

**1)** To develop an accurate representation of various non-standard amino acid residues like Aib and Div found in peptaibols, and the C-terminal aminoalcohols like Pheol or Leuol in terms of their atomic partial charges, geometry and ability to incorporate to peptide sequences for computational modeling.

**2)** To test the effect of various solvents and timescales on peptaibol folding using classical MD methods.

**3)** To optimize accelerated MD simulation parameters required to obtain complete conformational landscapes of peptaibols by comparing with a known peptaibol structure.

**4)** To test accelerated MD simulations for correctly modeling biological membranes and to reproduce experimental results.

# 4. MATERIALS AND METHODOLOGY

## 4.1. Sequence selection

For the first part, we carried out classical MD simulations on two sequences: Trikoningin KA V (TRK-V) (McMullin et al., 2017) and Tripleurin XIIc (TPN XIIc), an 18-residue long sequence with five ambiguous Val/Div and one Leu/Ile position was selected. Their primary structure is given in **Table 2**. 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% by liquid chromatography-mass spectrometry (LC-MS) analysis extracted at *m/z* of sodiated molecular ions $M+Na^+$, respectively (Marik et al., 2017). The ambiguous residue positions were predicted based on the sequence of NRPS proteins using the antiSMASH database server (Blin et al., 2016). For the second part, we focused on the use of aMD simulations to obtain complete conformational ensemble and structural dynamics of Alm F30/3. The same technique was later applied to TPN XIIc, Par-B & H, and Brev-I & IV (Marik et al., 2019) and TRK-V (**Table 2**).

**Table 2.** The primary sequences of all peptaibols studied in this work. The ambiguous residues have been underlined.

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Trikoningin KA V** | AcAib[1] | Gly[2] | Ala[3] | Aib[4] | Ile[5] | Gln[6] | Aib[7] | Aib[8] | Aib[9] | Ser[10] | Leu[11] | Aib[12] | Pro[13] | Val[14] | Aib[15] | Ile[16] | Gln[17] | Gln[18] | Leuol[19] | |
| **Tripleurin XIIc** | AcAib[1] | Ser[2] | Ala[3] | Aib[4] | **Vxx**[5] | Gln[6] | **Vxx**[7] | Aib[8] | **Vxx**[9] | Ala[10] | **Vxx**[11] | Aib[12] | Pro[13] | **Lxx**[14] | Aib[15] | **Vxx**[16] | Gln[17] | Pheol[18] | | |
| **Alamethicin F30/3** | AcAib[1] | Pro[2] | Aib[3] | Ala[4] | Aib[5] | Ala[6] | Gln[7] | Aib[8] | Val[9] | Aib[10] | Gly[11] | Leu[12] | Aib[13] | Pro[14] | Val[15] | Aib[16] | Aib[17] | Glu[18] | Gln[19] | Pheol[20] |
| **Brevicelsin I** | AcAib[1] | Ala[2] | Aib[3] | Ala[4] | Aib[5] | Gln[6] | Aib[7] | Leu[8] | Aib[9] | Gly[10] | Aib[11] | Aib[12] | Pro[13] | Val[14] | Aib[15] | Aib[16] | Gln[17] | Gln[18] | Pheol[19] | |
| **Brevicelsin IV** | AcAib[1] | Ala[2] | Aib[3] | Ala[4] | Aib[5] | Gln[6] | Aib[7] | Leu[8] | Aib[9] | Gly[10] | Aib[11] | Aib[12] | Pro[13] | Val[14] | Aib[15] | Val[16] | Gln[17] | Gln[18] | Pheol[19] | |
| **Paracelsin B** | AcAib[1] | Ala[2] | Aib[3] | Ala[4] | Aib[5] | Ala[6] | Gln[7] | Aib[8] | Leu[9] | Aib[10] | Gly[11] | Aib[12] | Aib[13] | Pro[14] | Val[15] | Aib[16] | Aib[17] | Gln[18] | Gln[19] | Pheol[20] |
| **Paracelsin H** | AcAib[1] | Ala[2] | Aib[3] | Ala[4] | Aib[5] | Ala[6] | Gln[7] | Aib[8] | Leu[9] | Aib[10] | Gly[11] | Aib[12] | Aib[13] | Pro[14] | Val[15] | Aib[16] | Val[17] | Gln[18] | Gln[19] | Pheol[20] |

## 4.2. Force field library generation for non-standard residues

As famously known, fungal peptaibols are characterized by their unusual amino acid content. In the selected sequences, Aib, Div, Pheol, and leucinol (Leuol) are the non-standard residues. A graphical representation of their 2D structures is provided in **Figure 6**. Both Aib and Div are derivatives of the parent alanine amino acid residue, in which one methyl group is present as the side-chain. An additional methyl group (-$CH_3$) is attached to the Cα carbon atom as the side-chain in Aib, while an ethyl group (-$CH_2CH_3$) is attached in the Div residue. For calculation of their partial charges and creating force field libraries, the R.E.D server was used (Vanquelef et al., 2011). R.E.D stands for RESP ESP charge derive (Dupradeau et al., 2010). RESP (restrained electrostatic potential) was used to calculate the charges with a HF/6-311G(d) basis set and Gaussian09 as quantum mechanical program interface (Frisch et al., 2009, 2014). The charges for Aib and Div were calculated along with other standard amino acids like alanine, glycine, serine, and valine. The charges calculated for standard residues were used to confirm with existing libraries in AmberTools 18 based on their respective atomic charges. For each residue, two conformations i.e. α-helix ($\Phi = -63.8$, $\Psi = -38.3$) and β-sheet or $C_5$ ($\Phi = -157.2$, $\Psi = 161.9$) were used. These were modified and generated using the Avogadro 3D molecular editor program (Hanwell et al., 2012) based on the strategy described by Cieplak et al. (1995). A slightly different strategy was used to calculate the charges for Pheol and Leuol, where two molecules, ethyl alcohol with phenylalanine and with leucine were used to form the Pheol and Leuol units, respectively. The results include the charges calculated in the molecule files and a script to make force field libraries for these forces (provided in **Appendix**). The sequence was built by supplying residue units from scratch using “tleap” after sourcing the library files of non-standard amino acids.

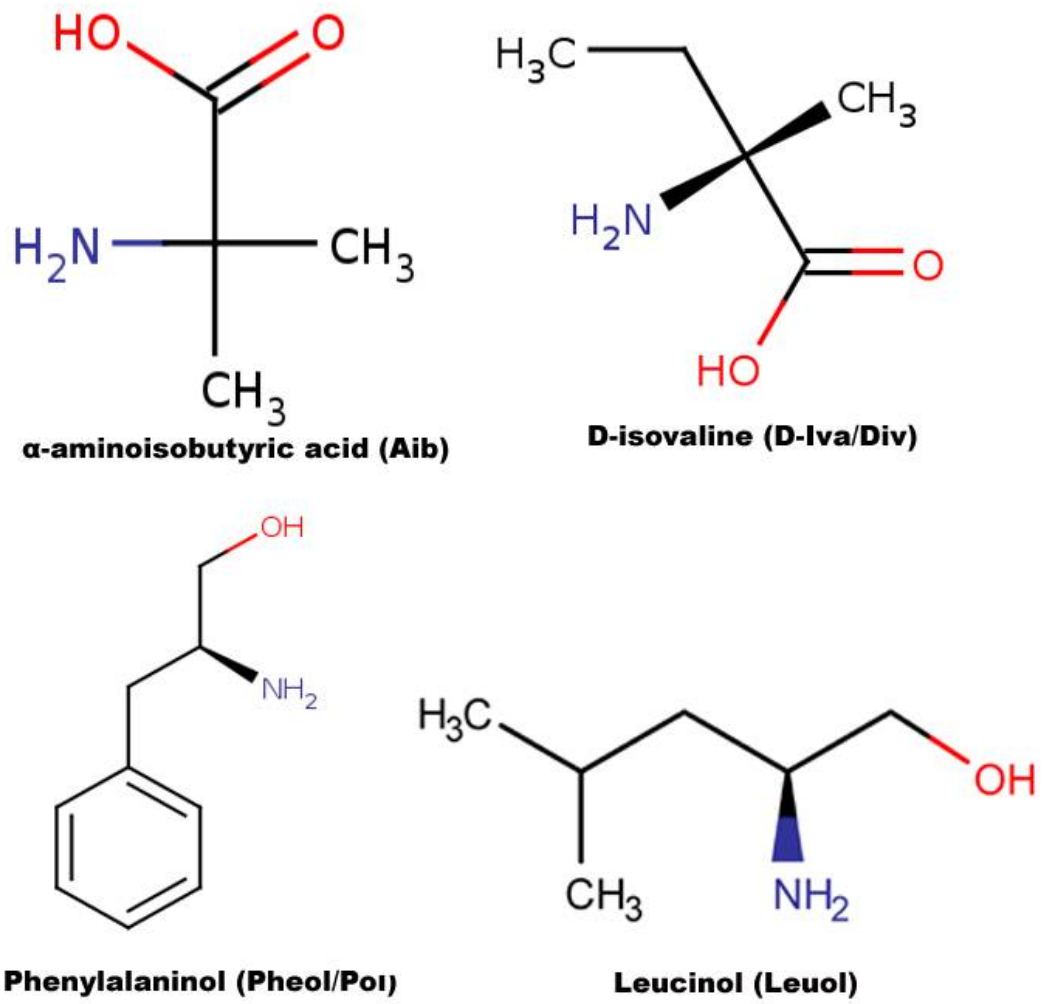


**Figure 6.** Graphical representation of 2D molecular structures of Aib, Div, Pheol and Leuol.

## 4.3. Classical molecular dynamics simulations

### 4.3.1. In implicit water solvent

For the first part, all MD calculations were carried out with AmberTools16 (Case et al., 2016) with ff14SB force field using generalized born implicit solvent method. Every MD simulation starts with preparing the whole system to ensure that it remains stable during the entire course of production run. The first step is energy minimization to stabilize the system. The maximum number of cycles was set at 10000 (maxcyc) with a convergence criterion of 0.01. The steepest descent algorithm was used for the first 100 cycles (ncyc) and then switched to conjugate-gradient algorithm for the remaining cycles. The energy minimization outputs were used for setting up the production run with 50000000 steps which correspond to 10,000 trajectory (frames) and 100 ns of total simulation time. All systems were maintained

at 300 K using Langevin thermostat (ntt = 3, gamma_ln =1.0). The time step was set to 2 fs and no cutoff was applied for non-bonding interactions. The resultant trajectories were visualized in Visual molecular dynamics software (VMD) (Humphrey et al., 1996). Secondary structure analysis was done by the *cpptraj* module (Roe & Cheatham, 2013).

#### 4.3.2. In explicit water and methanol solvents

The average structure of the last 30000 steps from the previous 100 ns long simulation was used as a starting structure for two 30 ns long simulations in explicit water and methanol solvent. The system was energy minimized for 20,000 steps and after a 50,000 step heating run, the production run was started at temperature 300 K under pressure regulation, ntp=1 condition. The periodic boundary conditions were applied, long-range effects were treated using the PME method (Darden et al., 1993). The methanol solvent simulation was carried out the same way with constant pressure and temperature regulation. In case of TPN XIIc, 2391 TIP3P water molecules were added for the explicit water simulation resulting in 7145 atoms in the system with the simulation box of size 42.113 × 55.536 × 42.040 Å and the volume of 98322.397 $Å^3$. Similarly, addition of 1336 methanol residues in a separate TPN XIIc system comprising of 8288 atoms with the size of 48.311 × 58.140 × 66.580 Å and the volume of 187008.295 $Å^3$.

### 4.4. Accelerated molecular dynamics simulations

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. The bias potential function, introduced by Hamelberg et al. (2004, 2007), was applied to make the simulation “jump over” high energy barriers and to sample rare events. A detailed discussion is provided in **section 2.2.3.** of *Introduction.*

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 the peptide under restraint, (b) water movement at 300K under isothermal and isobaric (NTP) conditions while keeping the peptide under restraint, (c) minimization of the whole system for 20,000 cycles, (d) heating from 0 K to 300 K under isothermal and isovolumetric (NVT) conditions while keeping the peptide under restraints, (e) relax the system at 300 K for 0.5 ns while keeping the heavy atoms of the peptide under restraint, and (f) relax system at 300 K under NTP conditions for 5 ns with no restraints. 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 (kcal $mol^{-1}$) 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.

All simulations were carried out at 300 K temperature, 2 fs time step, and energies and boost information was written at every 1000 steps. The electrostatic interactions were calculated using PME (Darden et al., 1993) and long-range interactions were also calculated with 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 NIIF High Performance Computing supercomputer at the University of Debrecen on the partition prod-gpu-k40-Leo nodes with 3 × Nvidia K40X CUDA8 were utilized for all aMD simulations. All simulations were carried out using *pmemd.cuda* implementation of Amber14, 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 FEL by 5-fold in comparison to classical MD, but the dual boost option provides a better reweighting distribution (Wereszczynski & McCammon, 2010).

The extra parameters $E_{dihed}$, $\alpha_{dihed}$, $E_{total}$ and $\alpha_{total}$ were calculated as required in Equation 4 (also provided in **section 2.2.3**. of *Introduction*):

$$E_{dihed} = V_{avg\_dihed} + a_1 \times N_{res}, \qquad \alpha_{dihed} = a_2 \times N_{res}/5;$$
$$E_{total} = V_{avg\_total} + b_1 \times N_{atoms}, \qquad \alpha_{total} = b_2 \times N_{atoms}$$

where $N_{res}$ is the number of peptide residues, $N_{atoms}$ is the total number of atoms in the system (Provided in **Table 3** for each system). $V_{avg\_dihed}$ and $V_{avg\_total}$ are average dihedral and total potential energies obtained from the 100 ns long cMD run in each solvent.

**Table 3.** The values of $N_{res}$, $N_{atoms}$ and the boost coefficients a1, a2 and b1, b2 for in each simulated system.

| Peptaibol compounds | **$N_{res}$** | **$N_{atoms}$** | **a1, a2** | **b1, b2** |
|---|---|---|---|---|
| **Tripleurin XIIc** | 18 | 8243 in water, 7462 in chloroform | See **Table 8** and **9** | See **Table 8** and **9** |
| **Alamethicin F30** | 20 | 14391 | See **Table 12** | See **Table 12** |
| **Trichoningin KA V** | 20 | 19467 | 4.5 | 0.20 |
| **Paracelsin B** | 20 | 12013 | 4.0 | 0.16 |
| **Paracelsin H** | 20 | 10957 | 4.0 | 0.16 |
| **Brevicelsin I** | 19 | 14448 | 4.0 | 0.16 |
| **Brevicelsin IV** | 19 | 13884 | 4.0 | 0.16 |

All systems were solvated using TIP3P water model for aqueous solvent except in case of TPN XIIc solvation in chloroform solvent using *tleap* module of AmberTools18 (Case et al., 2018). The number of residues added in each system with corresponding box size and volume is provided in **Table 4**.

**Table 4.** The solvation information for each system studied in this work.

| **Unfolded peptaibols** | | **No. of residues added** | **Periodic box size (Å)** | **Volume of box ($Å^3$)** |
|---|---|---|---|---|
| Tripleurin XIIc | Water solvent (TIP3P) | 2657 | 44.34 × 59.11 × 42.85 | 112332.6 |
| | Chloroform | 1021 | 49.60 × 59.51 × 67.98 | 200719.8 |
| Alm F30/3 | | 4701 | 48.79 × 71.80 × 54.41 | 190680.5 |
| Trikoningin KA V | | 2324 | 45.96 × 44.35 × 46.88 | 95587.0 |
| Paracelsin B | | 3910 | 55.05 × 46.82 × 62.33 | 160676.0 |
| Paracelsin H | | 3557 | 55.05 × 42.11 × 63.40 | 147021.3 |
| Brevicelsin I | | 4725 | 67.57 × 50.93 × 54.97 | 189190.3 |
| Brevicelsin IV | | 4536 | 68.52 × 45.96 × 58.30 | 183623.0 |

## 4.5. Accelerated molecular dynamics simulations applied on a bilayer membrane system: alamethicin F30/3 hexamer pore

The hexamer pore of Alm F30/3 peptide was obtained through M-ZDOCK server (http://zdock.umassmed.edu/m-zdock/) (Pierce et al., 2005). This is a Fast Fourier Transform

based protein docking program that predicts the structures of cyclically symmetric multimers. The hexameric pore was then embedded into a 3:1 mixture of 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE) and 1,2-dioleoyl-*sn*-glycero-3-[phospho-rac-(1-glycerol)] (DOPG) bilayer membranes which mimics a bacterial (*Escherichia coli*) membrane constitution. This system can be easily prepared in an Amber-ready format by using the '*packmol-memgen*' (Schott-Verdugo & Gohlke, 2019) workflow available with AmberTools18 that uses '*Memembed*' (Nugent & Jones, 2013) to obtain pre-oriented protein conformation with respect to the membrane. This system was solvated in 4410 water residues. The Alm F30/3 hexamer channel embedded in DOPE:DOPG bilayer system was prepared for aMD simulations starting with minimization followed by two steps of system heating and 10 steps of equilibration. This system was energy minimized for 10,000 steps which switches to conjugate gradient method after 5000 steps of the steepest descent method. The minimization was done at constant volume, no SHAKE algorithm was applied, and the non-bonded cutoff was set to 10.0 Å. The minimized system was set for two rounds of gradual 'heating' to reach the 'production run' temperature. In the first step, the system was heated to 100 K while the second step reaches a temperature of 303 K. The 10-step equilibration was carried out at 303 K temperature for 500 ps each. A short 25 ns long production run at 303 K was carried out with constraints on bond distances to calculate the aMD boost parameters followed by two consecutive aMD simulations of 300 ns each with 2 ps time step. Each aMD simulation was carried out with dual boost (iamd = 3) option at 300 K temperature regulated using a Langevin thermostat. A weak external static electric field was also applied along the z direction (across membrane) with efz values (intensity in kcal $(\text{mol} \times \text{A} \times \text{e})^{-1}$) of 0.180 and 0.080 for the first and second simulation, respectively (Escalona et al., 2016). The values of efz were chosen in a way that the magnitude of resulting electric potential is slightly higher than the voltage across plasma membrane. A membrane with 35 Å thickness has a potential of ~70 millivolts (mV) which is 0.07 V per $3.5 \times 10^{-7}$ cm or 0.02 V $\text{nm}^{-1}$ (Lodish et al., 2000). A 0.180 kcal $(\text{mol} \times \text{A} \times \text{e})^{-1}$ translates to an electric potential of 0.07 V × $\text{nm}^{-1}$ while 0.080 kcal $(\text{mol} \times \text{A} \times \text{e})^{-1}$ translates to 0.03 V × $\text{nm}^{-1}$. A distance restraint was applied for all glutamine amino acid residues owing to their importance in Alm F30/3 pore stability. The average pore radius was calculated using the HOLE utility (Smart et al., 1996) and water density available through MDAnalysis (Gowers et al., 2019; Michaud-Agrawal et al., 2011).

# 5. RESULTS AND DISCUSSION

## 5.1. Classical molecular dynamics

### 5.1.1. Implicit solvent simulations on the unfolded structure

#### *5.1.1.1. Secondary structural populations of the 100 ns long run of TRK-V in implicit water: vision from Ramachandran plot*

TRK-V was previously identified as a peptaibol produced by *Trichoderma koningiopsis* along with two other 11-residue sequences, trikoningin KB I and KB II (McMullin et al., 2017). Another study by our group identified novel peptaibols which were named as "koningiopsins" with TRK-V as the closest sequence (Marik et al., 2018). TRK-V, positionally isomeric with sequences Pept-Vb, -VIb, and -VII of *T. gamsii* is a 19-residue peptaibol with seven Aib residues in its sequence. Aib is an achiral residue, which has been shown to promote helix formation and can exist in both right- and left-handed helix regions on the Ramachandran plot. To determine the propensities of each residue for a given secondary structure type, their relative free energies were calculated which clearly describe an energetically favourable conformation (**Figure 7**). The spread of dihedral angle scatter during the simulation indicates that the system underwent through all the conformations. The darkest regions indicate energetically preferable conformations. We also characterized individual amino acid conformation and the overall conformation using the new nomenclature given by Hollingsworth and Karplus (2010) as discussed in the *Introduction*, **Section 2.2.5**.

Unexpectedly, a strong preference was found for the left-handed helix region of Φ-ψ plots during this simulation, specifically for residues in the central region flanked by Gln6-Aib7-Aib8-Aib9 and Aib12-Val14-Aib15-Ile16. Except for Aib1 and Aib4, all other Aib residues show free energy minimum in the left-handed helix region. Most standard (proteinogenic) amino acid residues, Gly2, Ala3, Ser10, Leu11, Gln17, and Gln18, display an energy minimum in the right-handed α-helix region. Ile5, Ser10, Leu11, Pro13, and Ile16 also show preference for the poly-proline II region. This behaviour of Leu and Ile to occupy the β-space on the Ramachandran plot is expected due to their β-branched side-chains. It is known that due to heavier side chains, they show less propensity to exist in a helix and, therefore, prefer to form β-strands. The presence of three consecutive Aib residues in positions 7, 8, and 9 seems to drive its conformation towards a left-handed helix, while the rest shows clear preference for right-handedness. This resulted in an overall unwinding of the helix and does not seem to form a continuous spiral shape.

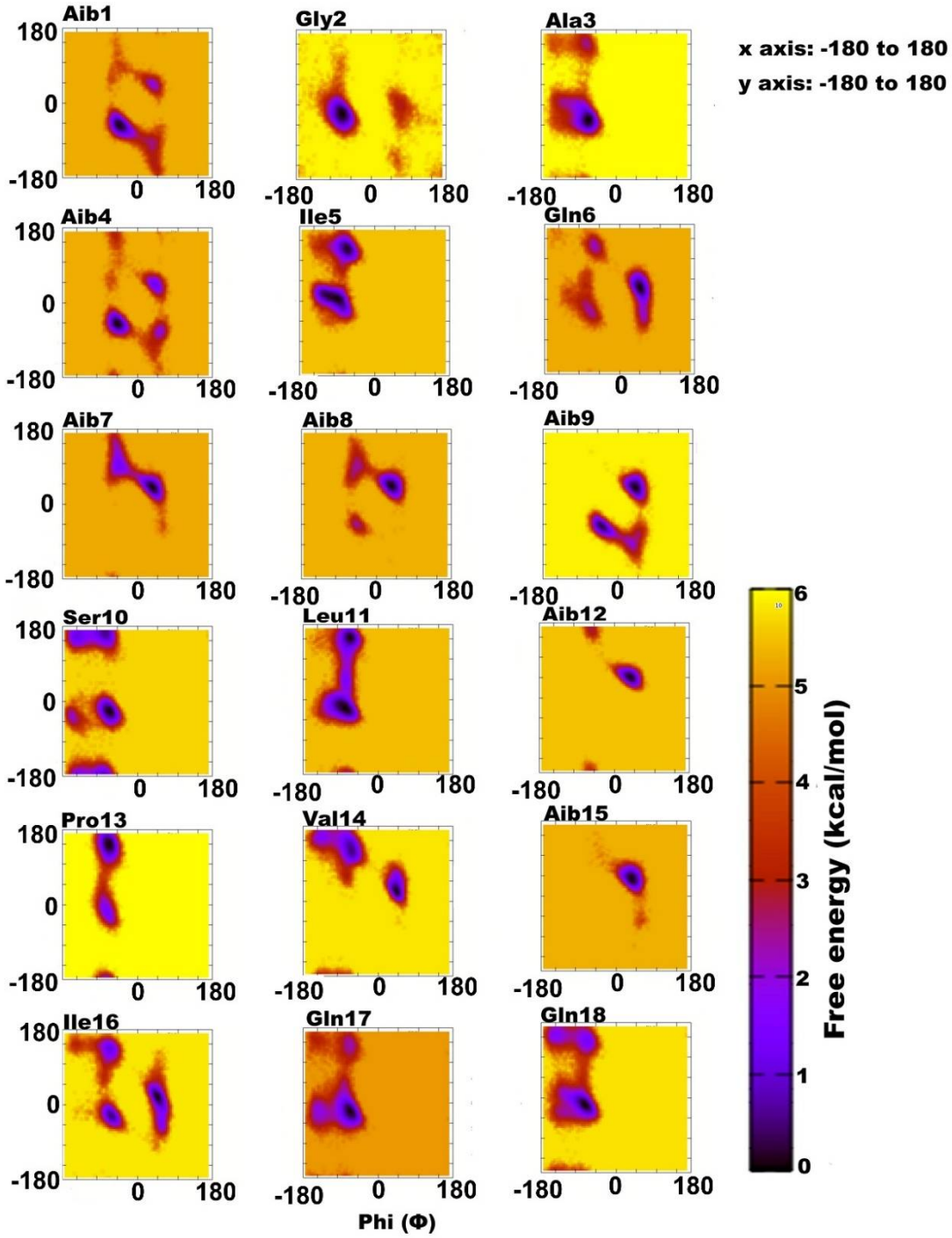


**Figure 7.** Free energy distribution of phi-psi torsion angles of each residue of TRK-V. The deepest violet regions denote energy minima regions.

The RMSD values based on the coordinates of peptide backbone atoms C, CA, and N for each frame with respect to the average structure has been calculated. Similarly, the radius of gyration (RoG) values, which is the RMSD of peptide components from their centre of mass, was calculated for each frame. The preliminary investigation revealed that the overall conformation (obtained from the trajectory with RMSD value between 12 Å to 14 Å, denoted by structures 3 and 4) resembles a hairpin structure with turns that never assumes a spiral

shape (**Figure 8A**). Structures 2 and 5 with an almost unfolded structure show lower RMSD values than 12 Å, which is not energetically favoured. The FEL as a function of RMSD and RoG is shown in **Figure 8B**, which clearly indicates that structures with RoG value of less than 8 Å and RMSD values between 12–14 Å are energetically favoured.

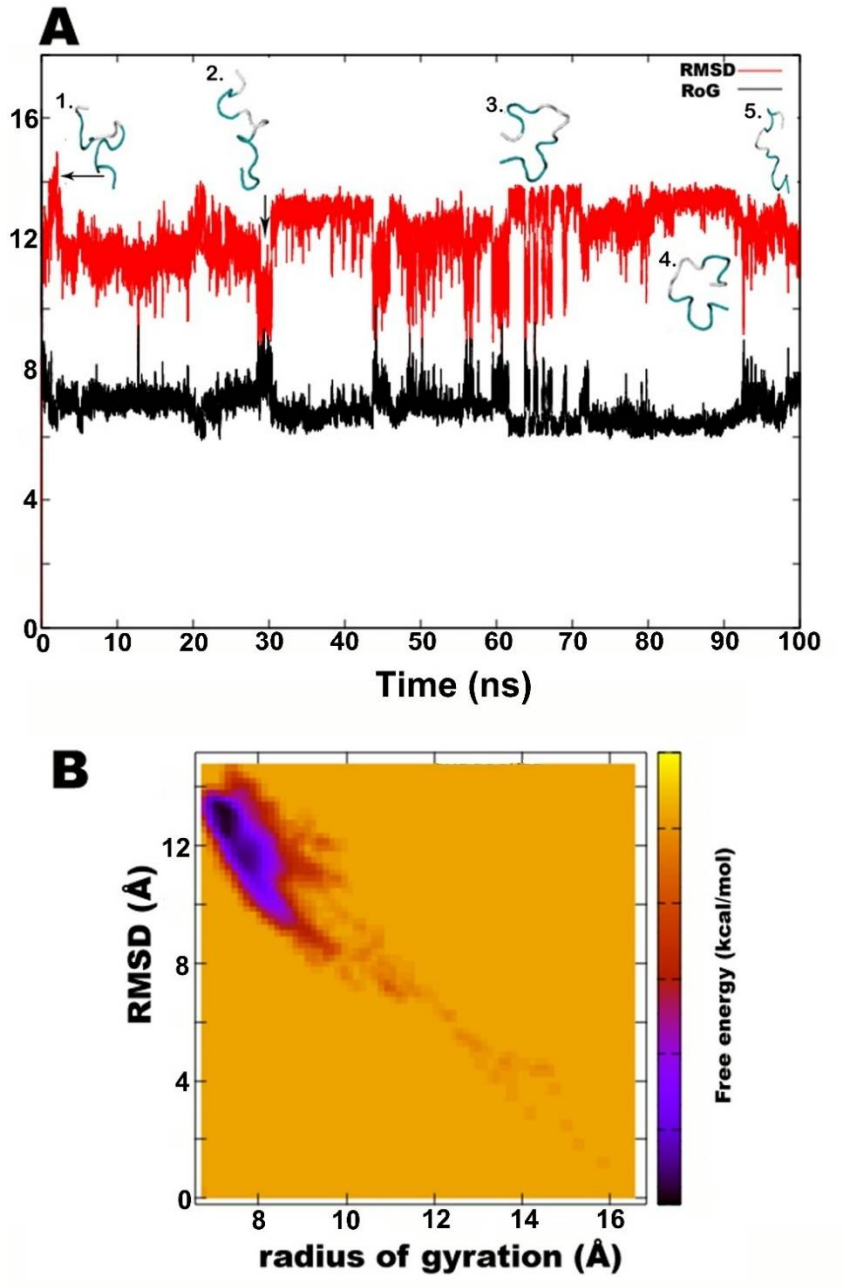


**Figure 8. (A)** The RMSD (red colour) and RoG (black) with corresponding three-dimensional structures of TRK-V. **(B)** FEL as a function of RMSD and RoG shows energetically favoured conformations with RMSD between 12–14 Å and RoG value less than 8 Å.

When compared with the hydrogen bonding pattern within the backbone, mostly i+3→i H-bonds were found that denote $3_{10}$ helices probably in left-handed conformation as indicated by Φ-ψ plots (**Table 5**). Ile5→Gly2, Aib8→Ile5, Aib9→Gln6, Ser10→Aib7, Ile16→Pro13, and Gln17→Val14 are examples of left-handed $3_{10}$ helix bonds while Gln6→Ala3, Leuol19→Ile16, Leu11→Aib8 are examples of right-handed $3_{10}$ helix. Few γ-turn populations are also seen by Aib7→Ile5, Aib12→Ser10, and Gln17→Aib15 as energetically stable. Moreover, few unusual bonds between Ile16→Gly2, Aib15→Gly6 and Gln17→Gly2 indicate a backbone hairpin turn that may put these residues in vicinity to each other and form a H-bond. This means that the highly bent structure resembling a β-hairpin with the N- and C-terminals in close proximity to each other is energetically favoured in comparison with a linear backbone.

**Table 5**. Backbone H-bonds of TRK-V along with their frequency of occurrence given by fraction, average distance, and angle.

| Acceptor | Donor | Fraction | Average Distance (Å) | Average angle (˚) |
|---|---|---|---|---|
| Gly2 | Aib7 | 0.20 | 2.89 | 158.38 |
| Gln6 | Aib9 | 0.18 | 2.89 | 156.70 |
| Ala3 | Gln6 | 0.17 | 2.89 | 151.58 |
| Pro13 | Ile16 | 0.17 | 2.89 | 158.27 |
| Ile16 | Leu19 | 0.16 | 2.89 | 158.30 |
| Aib8 | Leu11 | 0.14 | 2.89 | 154.75 |
| Aib7 | Ser10 | 0.12 | 2.89 | 155.29 |
| Val14 | Gln17 | 0.09 | 2.90 | 159.02 |
| Ile5 | Aib8 | 0.07 | 2.91 | 160.79 |
| Aib15 | Gln17 | 0.06 | 2.81 | 149.04 |
| Aib15 | Gln18 | 0.05 | 2.90 | 156.85 |
| Ile16 | Gly2 | 0.04 | 2.87 | 153.47 |
| Aib15 | Gln6 | 0.04 | 2.88 | 159.95 |
| Ser10 | Aib12 | 0.03 | 2.84 | 148.77 |
| Ile5 | Aib7 | 0.03 | 2.81 | 149.83 |
| Gly2 | Ile5 | 0.03 | 2.91 | 151.88 |
| Gln18 | Val14 | 0.03 | 2.85 | 150.00 |
| Gln17 | Gly2 | 0.03 | 2.87 | 153.97 |
| Aib8 | Aib12 | 0.03 | 2.90 | 156.30 |

#### *5.1.1.2. Secondary structural populations of the 100 ns long run of TPN XIIc in implicit water: vision from Ramachandran plot*

Tripleurins are a newly identified group of peptaibols produced by the fungus *Trichoderma pleuroti*, which causes green mold disease in the cultivation of oyster mushroom (*Pleurotus ostreatus*) (Marik et al., 2017). They were also reported as potential growth inhibitors of oyster mushroom mycelia. LC-MS 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 as Vxxol and Lxxol. From a mixture of various tripleurins, TPN XIIc has been selected based on its high yield and lowest number of ambiguous positions.

Comparison of implicit solvent simulated TPN XIIc plots revealed the presence of β- and some γ-turn populations on the plot especially for non-standard residues like Aib and Div (**Figure 9**). All Aib residues, Ala3 and Div5 show high scatter populations in the first quadrant of Ramachandran plots which is classically known to be left-handed α-helix conformation. However, according to the new classification, this region is marked as δ' region as a mirror image of the δ region. Aib residues also show fluctuation between left- and right-handed α regions owing to their achiral nature. Aib1, Ala3, Aib8 and Aib15 lose their left-handed character as the simulation proceeds towards a stable structure, while Aib12 almost entirely maintains its left-handedness (δ' region). Ala3 and Leu14 also seen in the $P_{II}$ scatter even towards the end of simulation almost like Pro13. All other standard residues populated in the α and β regions (including $P_{II}$ region) except Ala10, which is mostly populated in the α/δ-region with some pre-proline (ζ) character. Ala3, Gln6, Val9 and Leu14 also show large scatter in the β/$P_{II}$-regions, while Pro13 fluctuates between the α- and $P_{II}$ regions towards the end of simulation.

The δ-region has been classically recognized as β-turn conformations. β-turns comprise eight distinct types (Type I, I', II, II', VIa1, VIa2, VIb, and VIII) based on the dihedral angles of their central residues, i+1 and i+2, with one more miscellaneous category that contains all conformations other than the main eight classes, namely the IV class. The average dihedral angle values for various β-turn types were taken from Hutchinson & Thornton (1994). β-turn type I ( $\varphi_{i+1} = -60$, $\psi_{i+1} = -30$, $\varphi_{i+2} = -90$, and $\psi_{i+2} = 0$) was mostly observed amongst residues involving i+3→i H-bonds. Especially, Aib4 →Aib1, Gln6 →Ala3, and Ala10 →Div7 bonds falls in this category in all simulations. Highly constrained residues like Aib

can be accommodated at both i+1 and i+2 positions of type I and I' β turns (Mahalakshmi & Balaram, 2006).

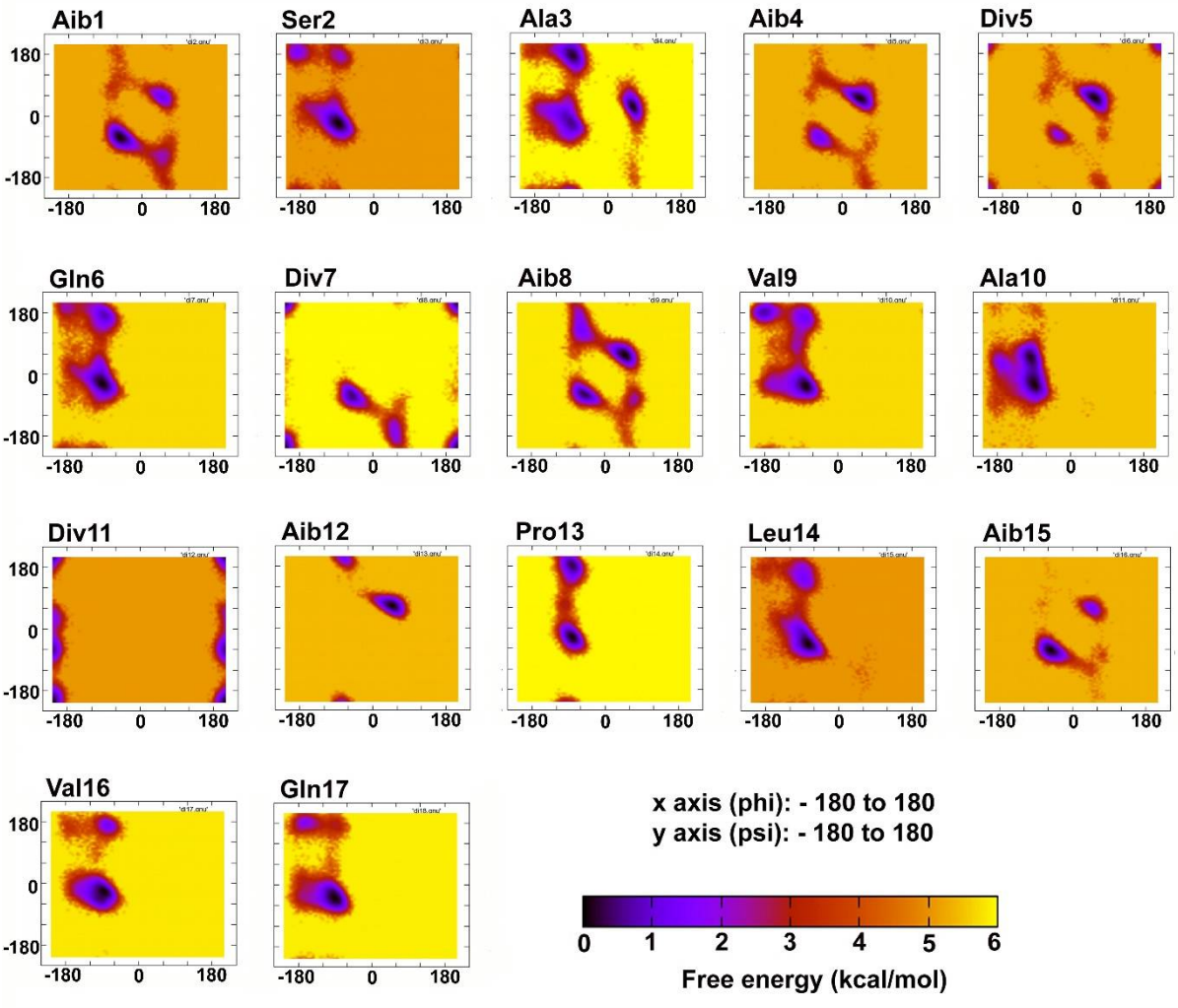


**Figure 9.** Free energy distribution of phi-psi torsion angles of each residue of TPN XIIc in implicit solvent.

Not surprisingly, the implicit solvent simulation samples many types of β-turns during its folding to attain the native conformation (**Figure 9**). The γ'-region (reverse γ turn) has been defined around (ϕ, ψ) = (−80, +80) for i+1 residue (Némethy & Printz, 1972). This conformation shows i+2→i (N→O) backbone hydrogen bond which was seen with high frequency at the Div11→Val9 bond during the simulation. The corresponding plots for Val9 and Ala10 also highly populated the γ'-region geometry, which extends from the α region to the β region. This explains why does not TPN XIIc form a linear helix but shows bending or backbone reversal around this point. All residues, except Div11 and Aib12 show high population in α (peak at ϕ, ψ = (−63, −43)) and δ regions. Div11, interestingly, populates the

scarcely populated corners of the plot (thereby not conforming to “allowed” regions), out of which, corners of the $I^{st}$ and $IV^{th}$ quadrants are defined as the ε-regions (mostly populated by glycine residues). Div11-Aib12-Pro13 residues seem to break TPN XIIc helix continuity. Although the reason behind helix-unfolding is the proline side-chain, its effects can be seen in the plots for Div11. These plots indicate preference for β-turn formation for most of the sequence except at the C-terminal. This is confirmed by the timeline secondary structure calculation plot (**Figure S1**, see Appendix).

Effect of Div stereoisomer: TPN XIIc contains three Div residues that may have characteristic effects on the screw sense of the resulting helix, on bioactivity and even on the probability to be degraded by proteolytic enzymes. Many AMPs fail to pass as therapeutic compounds due to their high toxicity profiles against mammalian cells and susceptibility to proteases. Amongst several strategies employed to overcome this bottleneck, for example, N- or C-terminal modification, cyclization, inclusion of non-peptidic backbones and multimerization, the incorporation of D-amino acids has been arguably the most successful strategy (Falciani et al., 2012; Kim et al., 2014; Ong et al., 2014; Khara et al., 2016).

Formaggio et al. (1995) reported that isovaline (Iva)-rich peptides favour β-bends and $3_{10}$ helices and their chirality affects the helix screw sense in such a way that D-/R-amino acids produce left-handed helices. As our peptide of interest, TPN XIIc has three Div and two Val residues. Div11 seems to populate only the extreme corners (β- and ε-regions) on the plot, most probably due to steric hindrance from Pro13. The values traverse from one extreme to another for both x- and y-axes, which shows that the continuous helix “flattens” out at that position and takes left-handedness at Aib12-Pro13. Due to the presence of other strong right-handed helix promoting residues the peptide returns back to right-handed screw sense at the C-terminal. Other important details are the large angle values (ranging from 110˚ to 160˚ for all simulations compared to an average of 110˚) between C-Div11-NPro12-C, indicating a strong steric repulsion between delta-methylene of Pro and the alkyl side chain of Iva residue. A similar observation was made by Kawai et al. (1993) for Div containing dipeptides.

#### *5.1.1.3. Folding dynamics of TPN XIIc in implicit solvent*

Upon analysing the FEL of the first implicit 100 ns long simulation at 298 K from the unfolded state, at least seven energy wells were revealed and numbered in a descending order by their population. The darkest violet regions on the FEL map show the lowest energy conformation clusters. The peptide undergoes large conformational changes from unfolded

conformation 6 to folded ones (**Figure 10**). The energy barrier between unfolded and folded conformations is quite large (~2-3 kcal $mol^{-1}$) as it is also between intermediate conformations 1, 2, 4, 5, and 7 and the final conformation denoted by 3 (difference is ~3.5 kcal $mol^{-1}$). A number of intermediate conformations occur but the energy barrier amongst them is easily traversible. The path of peptide folding is as follows, 6→ 1→ 4→ 2→ 7→ 2→ 5→ 3. Cluster 3 is formed from conformers belonging to the last 40 ns of simulation where it loses the α-helix at the N-terminal otherwise seen in representative structures of 1st, 2nd, 4th, and 7th clusters. Cluster 1 is the most populated group but its occurence in the beginning of the simulation indicates that it is a stable intermediate stage. Cluster 2 is the second highly populated group which fluctuates with the 7th cluster although the 7th stage does not last long. The central segment shows a classic β-turn property of backbone reversal by bringing the two terminals close to each other and thus, the peptide appears to be curved for most conformations. Only in the 3rd cluster, the peptide is mostly linear with a slight curve and almost perpendicular C-terminal alignment. The strict α-helix is replaced by a spiral-like conformation.

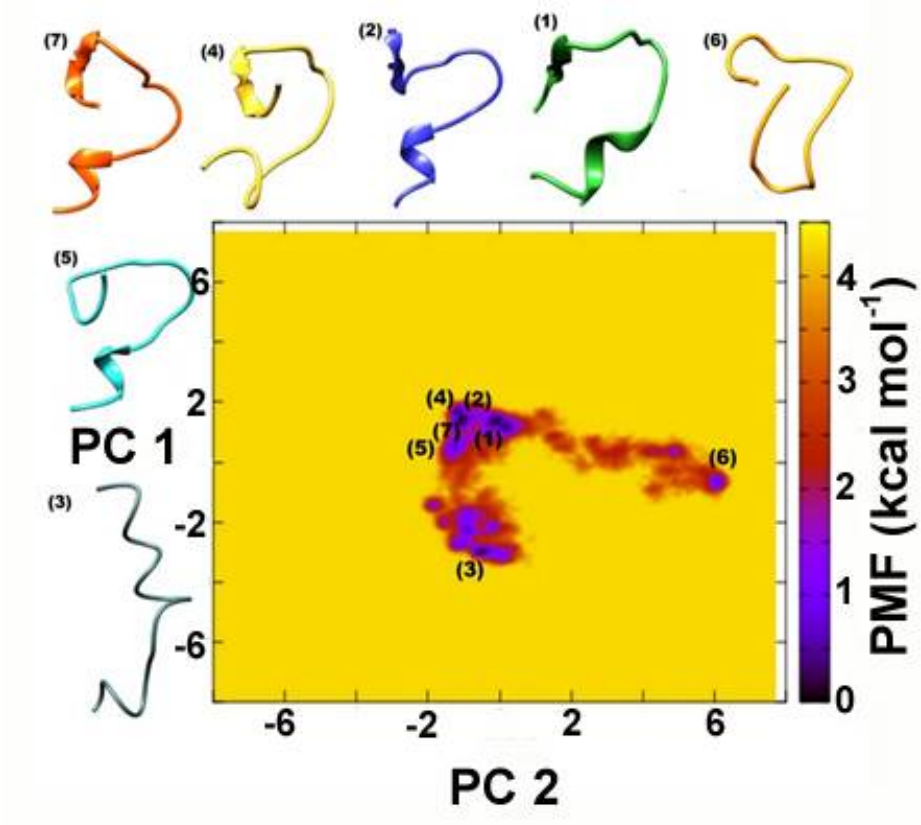


**Figure 10.** FEL showing various free energy basins and corresponding peptide conformations of TPN XIIc in implicit solvent.

Ferrera & Caflisch, (2000), have extensively shown implicit solvent models to be computationally efficient by accelerating the timescales involved in protein folding due to the absence of viscosity of solvent. However, the same authors reported that the folding times of these proteins were at lower bounds to the real folding time due to the absence of solute-solvent friction. Many questions are raised towards the accuracy of implicit solvation as they not only inadequately capture the folding kinetics but may also lead to a much faster collapse into a non-native conformation and a consecutive starkly different folding landscape. On the other hand, an explicit solution delays this collapse and may recover the interatomic contacts accurately before proceeding towards folding (Snow et al., 2005).

### 5.1.2. Explicit solvent simulations on the folded structure: comparison between water and methanol as solvents

We became curious to observe the folding behaviour in explicit solvent. TPN XIIc was simulated for a short 30 ns in explicit aqueous environment and in methanol. Methanol has been shown to promote comparable secondary structure folding in peptaibols as in phospholipid bilayers (Franklin et al., 1994; Miura, 2020). Ramachandran plots for each residue in explicit water (**Figure 11**) and methanol simulations (**Figure 12**) were studied and it was found that most residues are populated around the α- and δ-regions. All Aib residues, with the exception of Aib12, populate only the α/δ region. Aib12, on the other hand, lies in the δ' region. Div11 shows the same behaviour in all simulations. Pro13 lies almost exclusively in $P_{II}$ conformation in both explicit water and methanol solvents. This conformation is devoid of regular backbone hydrogen bonding and is separated from the β-region (classically both were accounted together). Ala10 was also found to scatter in this region as sampled from 20-30 ns in aqueous environment, which can be explained by the effect of Pro13.

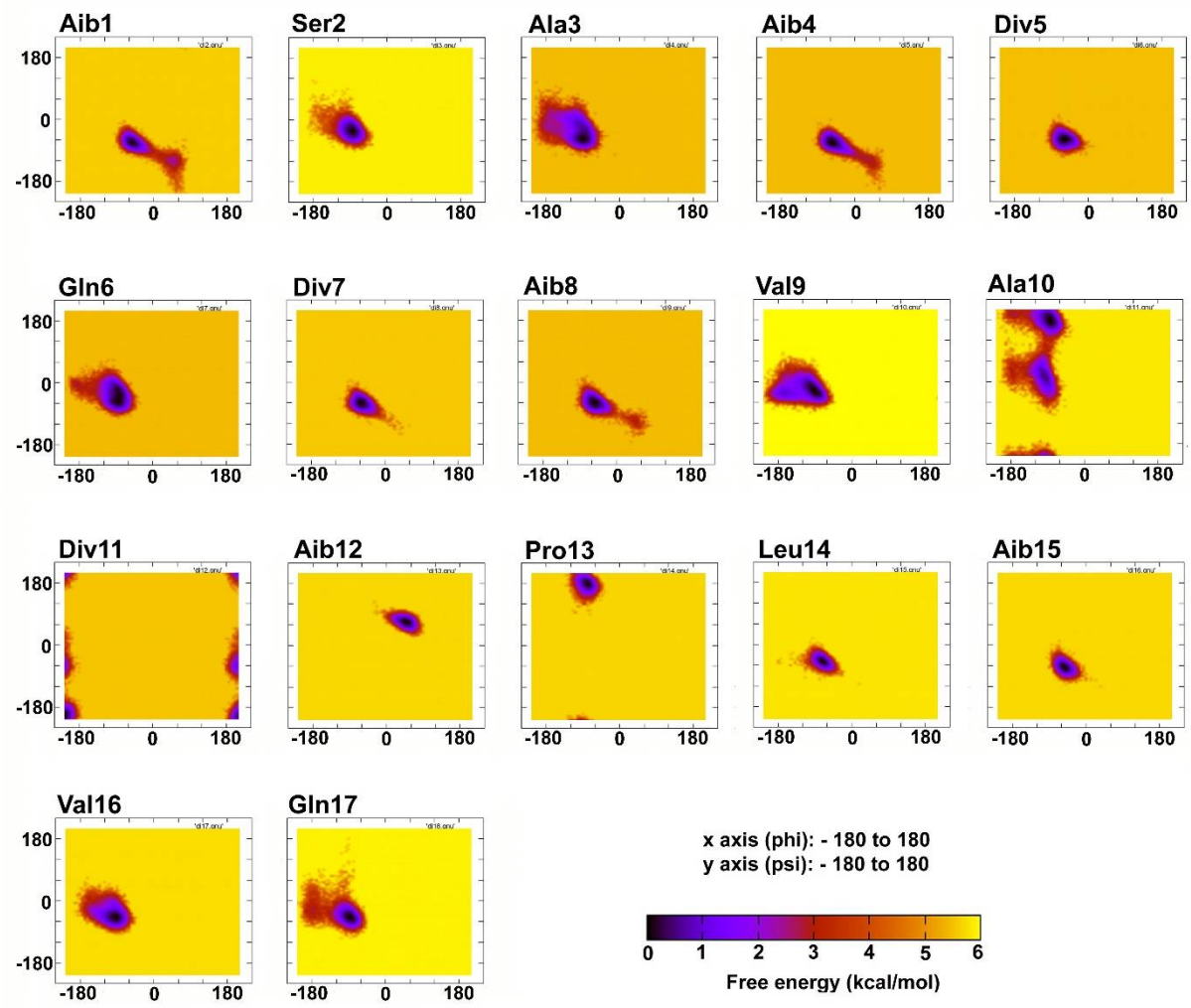


**Figure 11.** Free energy distribution of phi-psi torsion angles of each residue of TPN XIIc in explicit water solvent.

**Table 6** and **7** list the high occurrence of backbone H-bonds in TPN XIIc simulation in explicit water and methanol solvents, respectively. As it can be seen in these tables, the frequency of i+4→i H-bonds at the C-terminal is higher in water compared to methanol. For example, the Pol18→Leu14 (α-helix) bond occurs at a frequency of 54% in aqueous environment compared to 19% in methanol. By contrast, the Gln17→Leu14 (indicating δ region) bond occurs only 0.07% times in aqueous environment but at 37% times in methanol. Also, the Gln17→Pro13 bond occurs with 32% frequency in aqueous solvent against only 3% in methanol. It indicates that while aqueous solvent promotes α-helix stabilizing H-bonding pattern, methanol promotes the stability of δ region/$3_{10}$ helix (i+3 →i hydrogen patterns) at the C-terminal. Pol18→Aib15 and Pol18→Leu14 bonds occur with same frequency at 19% in methanol solvent indicating an equilibrium between $3_{10}$ and α-helix conformations.

**Table 6.** Backbone H-bonds of TPN XIIc in explicit water solvent along with their frequency of occurrence given by fraction, average distance and angle.

| Acceptor | Donor | Fraction | Average distance (Å) | Average angle (°) |
|---|---|---|---|---|
| Aib1 | Aib4 | 0.04 | 2.91 | 158.27 |
| Aib1 | Div5 | 0.03 | 2.91 | 163.47 |
| Ser2 | Gln6 | 0.04 | 2.90 | 159.80 |
| Ala3 | Gln6 | 0.25 | 2.87 | 153.69 |
| Ala3 | Div7 | 0.16 | 2.89 | 162.03 |
| Aib4 | Div7 | 0.09 | 2.90 | 159.56 |
| Aib4 | Aib8 | 0.02 | 2.91 | 165.17 |
| Gln6 | Ala10 | 0.19 | 2.87 | 160.21 |
| Div7 | Ala10 | 0.13 | 2.90 | 155.63 |
| Val9 | Div11 | 0.04 | 2.88 | 145.15 |
| Pro13 | Val16 | 0.28 | 2.87 | 148.57 |
| Pro13 | Gln17 | 0.32 | 2.88 | 161.57 |
| Leu14 | Gln17 | 0.07 | 2.89 | 151.99 |
| Leu14 | Pol18 | 0.54 | 2.87 | 158.52 |

The Div11→Val9 bond was discussed earlier for implicit solvent simulation of TPN XIIc for 100 ns. This bond occured at a relatively high frequency (14% out of 100,000 frames) and seemed to favour γ'-turn conformation, which now reduced to 4% and 1% in explicit aqueous and methanol solvent environments, respectively. Ramachandran plots for Ala10 (**Figure 11**) in aqueous solvent show highly populated scatter in γ' and δ regions before concentrating in the $P_{II}$ region towards the end of simulation, while in methanol, Ala10 shows almost negligible γ'-turn scatter (**Figure 12**). This indicates that γ'-turn conformation is not stable in explicit solvent environment, especially in methanol. The loss of the Div11→Val9 bond seems to reduce the hairpin-like bending of the backbone. Most of the high frequency H-bonds indicate the formation of δ turns (classically, $3_{10}$ helix regions).

The TPN XIIc structure can be discussed as three different segments, namely, N-terminal AcAib1-Aib8, central region Val9-Pro13, and the C-terminal Leu14-Pheol18. The N-terminal segment mostly shows formation of continuous β-turns. This shapes the peptide into a loose helix-like (spiral) conformation. The central region unwinds the helix at the Div11-Aib12-Pro13 site with Pro13 characteristically populating the $P_{II}$ conformation in both explicit water and methanol solvent-based simulations.

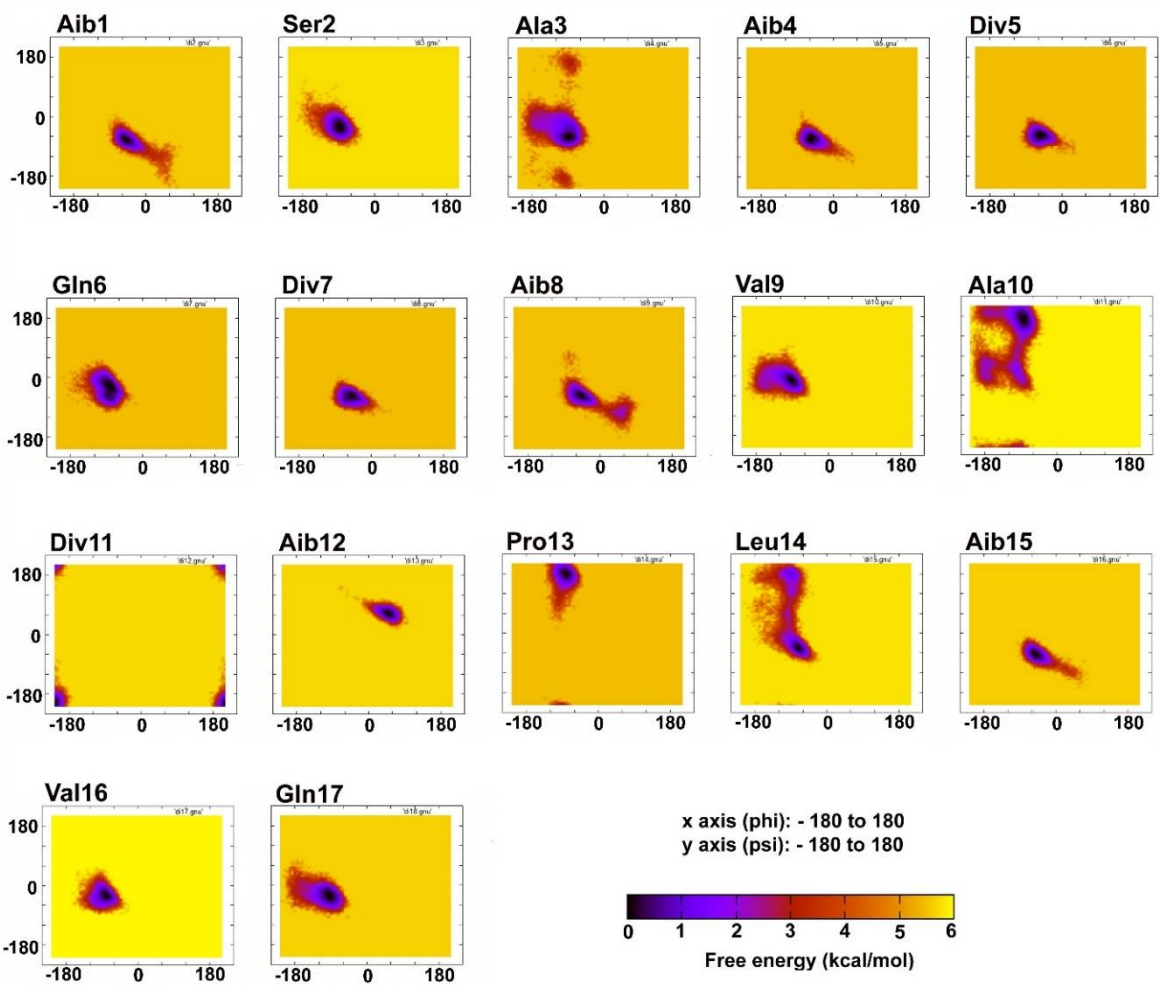


**Figure 12.** Free energy distribution of phi-psi torsion angles of each residue of TPN XIIc in explicit methanol solvent.

An interesting observation occurs in the C-terminal segment, where Leu14 remains concentrated mostly in the α/δ regions in aqueous environment but shows a high fluctuation while traversing the Ramachandran space from α- to $P_{II}$ in methanol environment after 20 ns. Contrary to this, Ala10 shows $P_{II}$ scatter in aqueous environment but not in methanol solvent. According to Brown & Zondlo, (2012), leucine shows the highest $P_{II}$ propensity after proline which is followed by alanine and other linear side-chain residues. The loss of $P_{II}$ conformation of alanine from water to alcohol (trifluroethanol) was also noted previously by Chen et al. (2004). In case of short alanine peptides (GGAGG) $P_{II}$ conformation is predominant in water at 20°C, which disappears when the solvent is changed to trifluroethanol and shows preference for γ'-turn (C7 eq) conformations (Liu et al., 2004). An explanation can be that $P_{II}$ structure is stabilized by hydrogen bonds between water and backbone amide nitrogen and carbonyl oxygen (Han et al., 1998). Another explanation is that $P_{II}$ conformation poses least effect on disruption of H-bonds present in liquid water (Pappu &

Rose, 2002; Drozdov et al., 2004). Comparing the angle averages of Div5, Div7 and all Aib residues except Aib12 by hydrogen bonding patterns, the formation of β-bend ribbon spirals at the N-terminal segment (Aib1-Ala10) is clearly evident (**Figure S2** and **S3**, see Appendix). The β-bend ribbons are formed when at least two hydrogen bonded β-turns are observed with i+3 →i and i+5 →i+2 bonding. Aib4 →Aib1, Gln6 →Ala3, Aib8 →Div5, Ala10 →Div7 H-bonds occur with high frequency, thereby continuing the β-bend ribbon spiral. In aqueous environment, however, Gln6→Ala3 and Aib8→Div5 bonds are not formed with high frequency, which indicates that the β-turns are not bonded by sufficient strength to maintain the continuity of ribbon spirals (**Table 7**). But in methanol solvent these bonds occur at high frequency providing greater stability to the spiral N-terminal conformation. Kinoshita et al. (2000) demonstrated that if the conformational energy of a particular secondary structure is lower for a peptide, the alcohol environment will promote its formation compared to water. It can be concluded that indeed, alcohol environment promotes the stability of any secondary structure that the constituent amino acid residues show natural propensity for.

**Table 7.** Backbone H-bonds of TPN XIIc in methanol solvent along with their frequency of occurrence given by fraction, average distance and angle.

| Acceptor | Donor | Fraction | Average distance (Å) | Average angle (°) |
|---|---|---|---|---|
| Aib1 | Div5 | 0.05 | 2.91 | 163.13 |
| Aib1 | Aib4 | 0.03 | 2.90 | 159.49 |
| Ala3 | Gln6 | 0.30 | 2.88 | 157.41 |
| Ser2 | Gln6 | 0.07 | 2.89 | 160.92 |
| Aib4 | Div7 | 0.17 | 2.90 | 159.23 |
| Ala3 | Div7 | 0.11 | 2.89 | 161.30 |
| Div5 | Aib8 | 0.03 | 2.91 | 160.61 |
| Gln6 | Val9 | 0.23 | 2.89 | 160.78 |
| Gln6 | Ala10 | 0.03 | 2.87 | 156.03 |
| Div7 | Ala10 | 0.32 | 2.89 | 159.39 |
| Val9 | Div11 | 0.01 | 2.87 | 145.05 |
| Pro13 | Val16 | 0.19 | 2.87 | 152.13 |
| Pro13 | Aib15 | 0.03 | 2.86 | 148.41 |
| Pro13 | Gln17 | 0.03 | 2.90 | 158.58 |
| Leu14 | Gln17 | 0.37 | 2.88 | 156.24 |
| Leu14 | Pol18 | 0.19 | 2.87 | 157.90 |
| Aib15 | Pol18 | 0.19 | 2.90 | 154.40 |

#### *5.1.2.1. Folding dynamics of TPN XIIc in explict water and methanol solvents*

The FEL landscape of TPN XIIc simulations in explicit solvents of water and methanol revealed interesting differences in peptide dynamics (**Figure 13**). Both simulations were carried out for 30 ns using folded peptide structure from previous simulation. In the aqueous environment, the peptide exists in two distinct conformations with an energy barrier of ~2-4 kcal $mol^{-1}$ between them. Cluster 1 is more populated and includes 70% of all lowest energy conformations and can be accepted as the native form. The folding dynamics follow the 2→1 path along the simulation time. Conformation 2 representative is highly curved, almost like a backbone reversal while conformation 1 is relatively linear. Similarly, upon comparison in the methanol environment, three low energy basins are formed with Cluster 1 being highly populated (84% of the total) with lowest energy conformations. The representative structure of Cluster 1 is reported as the native form in methanol solvent. It fluctuates with Cluster 2 where the energy barrier between them is only ~0.5-1.5 kcal $mol^{-1}$ and can be surpassed relatively easily. The folding dynamics follow the path, 3→1→2→1. The 3rd cluster is populated by conformations up to 5 ns (only 8% of total conformations) in the simulation time. The representative conformation for cluster 1 shows helix formation at the C-terminal and only a slight curve in the backbone. The coil-like helix can be visualized as β-bend ribbon spirals. It can be deduced that methanol environment promotes the formation of helix-like structures and reduces backbone curving in comparison to water solvent. These results are in accordance with many previous studies which report α-helical forming properties of alcohol solvents (Nelson et al., 1986; Hamada et al., 1995; Kinoshita et al., 2000; Yu et al., 2016).

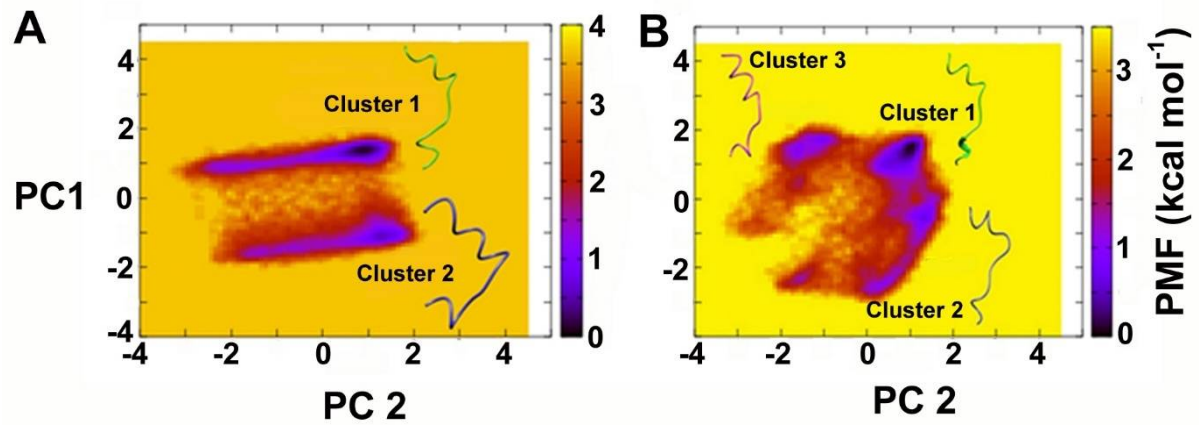


**Figure 13.** FEL showing various energy minima basins and corresponding peptide conformations of TPN XIIc in explicit water **(A)** and methanol **(B)** solvents.

## 5.2. Accelerated molecular dynamics simulations

The progressively diverse results obtained with increasing simulation time made it obvious that such short time scales were insufficient to carry out folding simulations. The ability of proteins or peptides to adopt multiple conformational states is essential to their function and elucidating the details of such conformational diversity under physiological conditions has been a major challenge (ElGamacy et al., 2019). The complete conformational landscape of a peptide is called a phase space and it must be elucidated to identify multiple intermediate or functional conformations. However, peptides (or proteins) may get stuck in a single energy minimum. To overcome the problem of obtaining insufficient conformational landscape of a peptide folding, we turned to accelerated molecular dynamics. One important factor in choosing aMD over other enhanced sampling methods was that it does not require a prior knowledge of the reaction coordinate to guide the simulation in a particular direction. The aMD simulations required GPU machines which were available through the NIIF clusters of Hungary and those available at University of Szeged and University of Debrecen were utilized. All simulations were carried out using *pmemd.cuda* implementation of Amber14, also available at the cluster. A run can be setup for a maximum of 7 days, therefore, checkpoints must be created from which a simulation may be started again in case it does not get completed.

### 5.2.1. TPN XIIc: peptide folding dynamics in aqueous and hydrophobic environment mimic

In this section, we show application of accelerated MD on a novel peptaibol, TPN XIIc, that has also been studied using classical MD techniques in the previous sections (refer section **5.1.1.2., 5.1.1.3.** and **5.1.2.**). We carried out four independent simulations of unfolded TPN XIIc in water and chloroform solvents each. The first three simulations were carried out for 500 ns while the fourth was carried out for 1 μs (1000 ns) using different boost parameters (**Table 8** and **9**). The dynamics of TPN XIIc was studied in detail using principal component analysis (PCA) on internal/dihedral angle coordinates and effective sampling was assessed using Cartesian coordinate-based PCA. The KLD method is used to measure the extent of overlap between a probability distribution (Kullback & Leibler, 1951), and has recently been adapted to discuss adequate sampling and convergence for protein MD simulations.

**Table 8.** Summary of coefficient a1, a2 and b1, b2 applied to consecutive simulations of TPN XIIc and the resulting average boost energy.

| | Water Simulation | | | Chloroform Simulation | | |
|---|---|---|---|---|---|---|
| | a1, a2 | b1, b2 | Avg. boost (kcal $mol^{-1}$) | a1, a2 | b1, b2 | Avg. boost (kcal $mol^{-1}$) |
| | 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 |
| Last 1 µs simulation | 4.5 | 0.16 | 13 | 6.0 | | 27 |

**Table 9.** Summary of various accelerated molecular dynamics parameters of TPN XIIc simulation.

| Simulations | Time (ns) | Boost Option | Vavg_dihed (kcal $mol^{-1}$) | Vavg_total (kcal $mol^{-1}$) |
|---|---|---|---|---|
| **In water** | 2500 (500 × 3 + 1000 ns) | iamd = 3 | 210 | −25429 |
| **In chloroform** | 2500 (500 × 3 + 1000 ns) | iamd = 2 | 206 | −7535 |

#### *5.2.1.1. Secondary structural populations of TPN XIIc in aqueous and chloroform solvents obtained from combined trajectory*

In explicit water: The reweighted phi-psi plots have been constructed for each individual residue of TPN XIIc. The colour scale denotes potential-of-mean-force or PMF (kcal $mol^{-1}$) to characterize energetically favoured conformations that were observed during simulations. The darkest violet regions denote the lowest energy minimum (**Figure 14**). In water solvent, the energy minima for non-standard residues, Aib1, Aib4, Div5, Aib8, Div11, and Aib12 lie in both right- and left-handed α-regions. This fluctuation is attributed to the achiral nature of Aib residue and the propensity of D-residues towards left-handed conformations. It can be noted that chirality of isovaline does not highly impact the screw sense of whole helical structure, possibly because of the limited difference in the length of the two side chains (De Zotti et al., 2012). 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 β-bends and $3_{10}$ helices (Formaggio et al., 1995).

A study by Shenkarev et al., (2013) compared structural and functional characteristics of Aam-I with Zrv-IIb. Both molecules acquire very similar structural topologies in membrane-mimicking environments, although significantly differ at the N-termini (1-8 residues) due to

high population 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 high 'motional' propensity of the former which may be caused by conformational exchange (characterized by positive and negative φ/ψ torsion angles) of Aib, Div and Gly residues. They hypothesized that such conformational exchange may be responsible for its high solubility in water 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 this peptide less bioactive. In the case of TPN XIIc, it is clear that the right-handed helical sense is energetically more favourable than the left-handed, although both right- and left-handed helical states are visited by 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.

A comparison with the phi-psi plots obtained through all-atom classical MD simulations of TPN XIIc discussed under **Section 5.1.2.**, it is clear that a higher number of possible low energy conformations are obtained when using aMD with a more accurate representation of free energy landscaape. The major differences can be observed for Ala10 and Div11 where the former showed energy minimum in the $P_{II}$ conformation and the latter in the ε region during classical MD. But for aMD, both residues show the largest energy minimum in the α-regions, thereby, bringing a continuous helical sense to the bakbone. It is evident that aMD simulations have brought forward all metastable and low energy states that remained hidden during classical MD simulations.

The energy minima for almost all standard residues like Ser2, Ala3, Gln6, Leu14, Val16, and Gln17 lie in the right-handed α-region, while the other secondary structural states could only be reached at a difference of 2 kcal $mol^{-1}$. One such state is the bridge between α- and β-regions, known as the δ-region which represents β-turn formation in the peptide chain and continuous formation of β-turns gives rise to β-bend ribbon spirals. Most standard residues of

TPN XIIc show considerable population in this region even if at a higher energy scale (~4 kcal $mol^{-1}$).

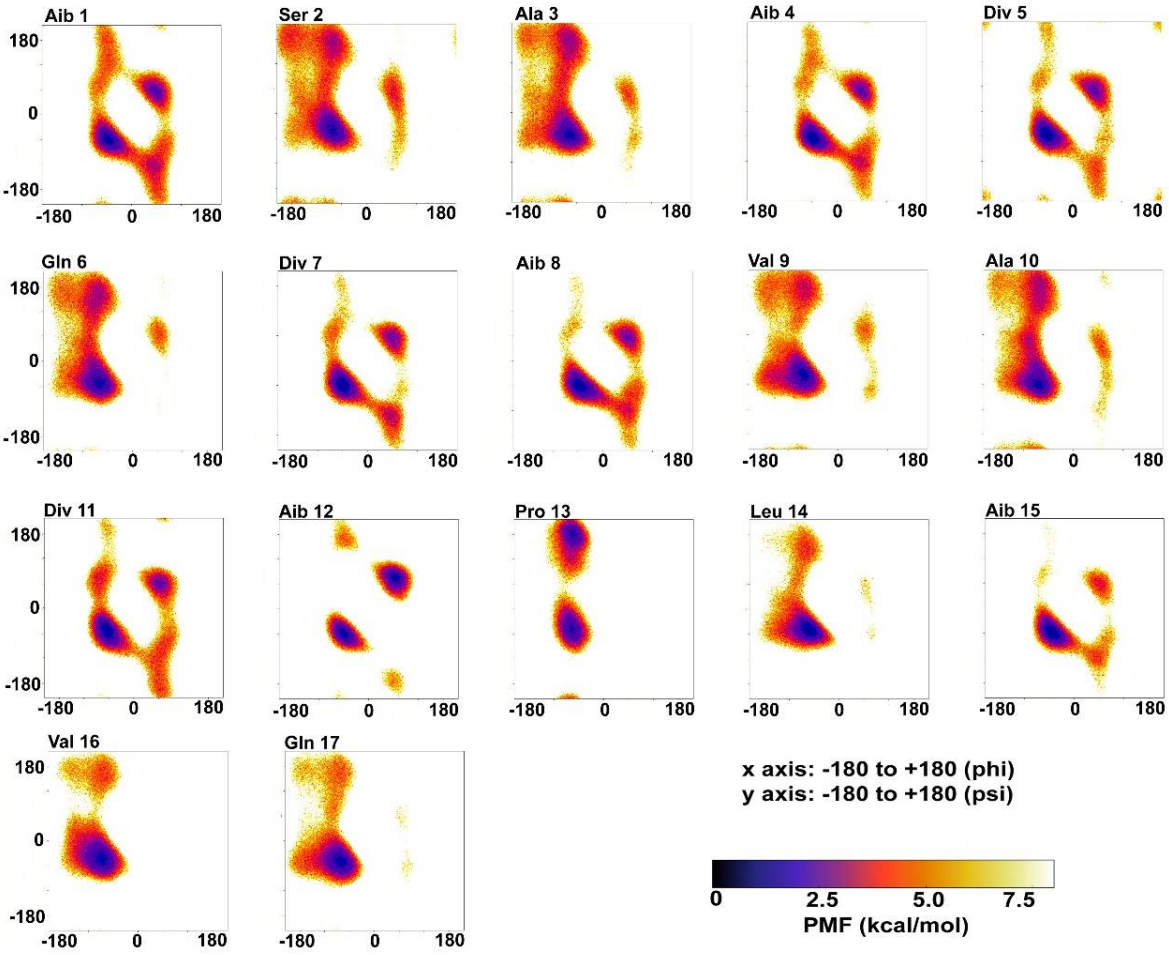


**Figure 14.** Reweighted PMF phi-psi dihedral angle plots for each TPN XIIc residue during explicit water aMD simulation. The x and y axes range from -180 to +180. The darkest violet regions indicate toward minimum energy secondary structural regions favored by each residue during the simulation.

On the other hand, Pro13 traverses poly-proline II region which is expected according to Brown & Zondlo (2012) as proline shows the highest $P_{II}$ propensity. In the water environment, TPN XIIc peptide shows fluctuating propensity for i+3→i and i+4→i type H-bonds (**Table 10**). Gln6→Ala3 (19%), Val9→Gln6 (15%), Ala10→Div7 (11%), Aib12→Val9 (4%), and Val16→Pro13 (19%) are examples of i+3→i type H-bonds, while Div7→Ala3 (14%), Aib12→Aib8 (12%), Ala10→Gln6 (14%), Div11→Div7 (14%), Gln17→Pro13 (21%) and Pheol18→Leu14 (44%) are examples of i+4→i type H-bonds. On the Ramachandran plot, α- 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 favoured instead of a strict helical conformation except at the C-terminal.

**Table 10.** Backbone H-bonds of TPN XIIc in explicit water solvent along with their frequency of occurrence given by fraction, average distance and angle.

| Acceptor | Donor | Fraction | Average distance (Å) | Average angle (˚) |
|---|---|---|---|---|
| Aib1 | Div5 | 0.05 | 2.89 | 162.38 |
| Ser2 | Gln6 | 0.08 | 2.88 | 160.03 |
| Ala3 | Gln6 | 0.19 | 2.87 | 154.95 |
| Ala3 | Div7 | 0.14 | 2.88 | 161.59 |
| Aib4 | Div7 | 0.02 | 2.90 | 155.38 |
| Aib4 | Aib8 | 0.09 | 2.89 | 161.68 |
| Div5 | Aib8 | 0.02 | 2.88 | 154.44 |
| Div5 | Val9 | 0.02 | 2.89 | 161.79 |
| Gln6 | Val9 | 0.15 | 2.87 | 154.92 |
| Gln6 | Aib8 | 0.05 | 2.78 | 149.02 |
| Aib8 | Aib12 | 0.12 | 2.89 | 159.16 |
| Gln6 | Ala10 | 0.14 | 2.87 | 160.11 |
| Div7 | Val9 | 0.06 | 2.79 | 147.89 |
| Div7 | Ala10 | 0.11 | 2.88 | 152.30 |
| Div7 | Div11 | 0.14 | 2.89 | 161.04 |
| Val9 | Aib12 | 0.04 | 2.88 | 159.08 |
| Ala10 | Aib12 | 0.04 | 2.80 | 148.52 |
| Ala10 | Leu14 | 0.03 | 2.88 | 159.89 |
| Div11 | Leu14 | 0.12 | 2.86 | 151.96 |
| Div11 | Aib15 | 0.18 | 2.88 | 162.72 |
| Pro13 | Val16 | 0.19 | 2.87 | 151.15 |
| Pro13 | Gln17 | 0.21 | 2.87 | 159.68 |
| Leu14 | Gln17 | 0.10 | 2.88 | 152.47 |
| Leu14 | Pol18 | 0.44 | 2.86 | 160.43 |

In explicit chloroform (hydrophobic environment): On the contrary, reweighted phi-psi plots obtained from chloroform simulation indicate towards a considerably different folding pattern (**Figure 15**). The energy minima for Aib1, Aib4, Div5, Div7, Aib8, and Aib15 lie only in the right-handed α-region unlike in the water solvent, where both right- and left-handed states are visited. A rather unusual behaviour is exhibited by Ala10 which favours the classic γ-turn formation, while the consecutive Div11 residue favours inverse γ-region. This would probably bend the backbone in a slight ‘S’-shape or an outward kink.

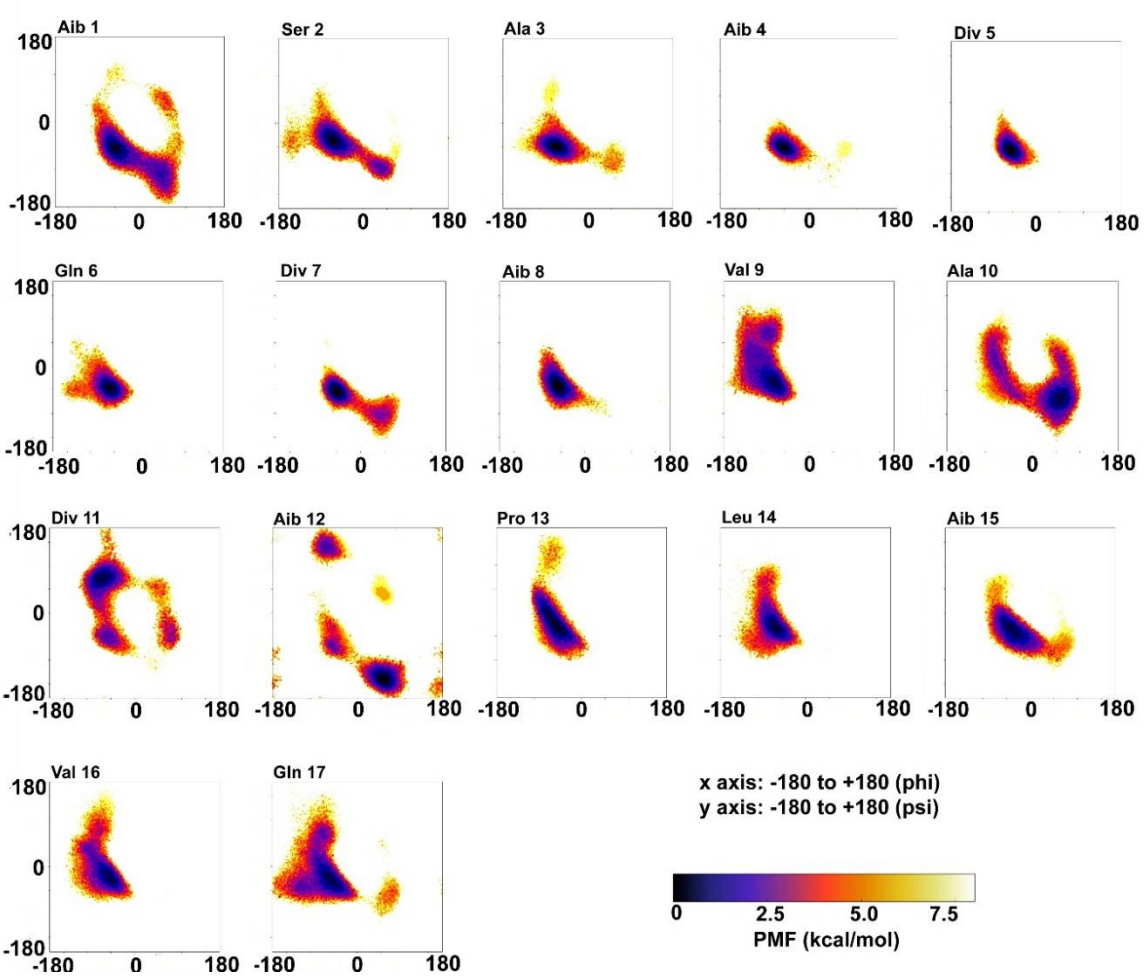


**Figure 15.** Reweighted PMF phi-psi dihedral angle plots for each TPN XIIc residue during explicit chloroform aMD simulation. The x and y axes range from -180 to +180. The darkest violet regions indicate toward minimum energy secondary structural regions favoured by each residue during the simulation.

The corresponding H-bond Aib12→Ala10, which occurs in 44% of the simulation time, also indicates towards a strong backbone reversal at this position due to formation of a γ-turn. The average H-bonds formed during this simulation have been summarized in **Table 11**. This observation can be attributed to the non-polar nature of chloroform environment that would cause the polar glutamine and serine residues to adopt a conformation that avoids the solvent. A recent study on the behaviour of alanine dipeptide in explicit chloroform and water solvents by Rubio-Martinez et al. (2017) described that $C_7^{eq}$ conformation appears as a low energy minimum only in chloroform. Standard residues like Ser2, Ala3, Gln6, Val9, Leu14, Val16, and Gln17 show energy minima strictly in the right-handed α-region.

**Table 11.** 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 (Å) | Average angle (°) |
|---|---|---|---|---|
| Aib1 | Div5 | 0.15 | 2.90 | 162.68 |
| Ser2 | Aib4 | 0.09 | 2.82 | 148.54 |
| Ser2 | Gln6 | 0.24 | 2.88 | 160.23 |
| Ala3 | Div5 | 0.10 | 2.80 | 149.61 |
| Ala3 | Div7 | 0.23 | 2.90 | 162.90 |
| Ala3 | Gln6 | 0.07 | 2.88 | 152.31 |
| Aib4 | Aib8 | 0.09 | 2.90 | 163.41 |
| Gln6 | Val9 | 0.10 | 2.89 | 154.32 |
| Gln6 | Div11 | 0.07 | 2.90 | 152.28 |
| Gln6 | Ala10 | 0.13 | 2.87 | 156.93 |
| Div7 | Val9 | 0.04 | 2.91 | 162.20 |
| Div7 | Ala10 | 0.24 | 2.88 | 152.99 |
| Aib8 | Ala10 | 0.02 | 2.88 | 146.67 |
| Val9 | Div11 | 0.18 | 2.83 | 147.37 |
| Val9 | Aib12 | 0.01 | 2.88 | 161.66 |
| Ala10 | Aib12 | 0.44 | 2.79 | 149.19 |
| Div11 | Leu14 | 0.13 | 2.87 | 155.91 |
| Div11 | Aib15 | 0.08 | 2.88 | 163.24 |
| Aib12 | Leu14 | 0.20 | 2.85 | 151.67 |
| Aib12 | Aib15 | 0.08 | 2.89 | 160.15 |
| Pro13 | Aib15 | 0.13 | 2.84 | 148.97 |
| Pro13 | Val16 | 0.19 | 2.88 | 158.57 |
| Aib15 | Pol18 | 0.20 | 2.88 | 158.80 |
| Leu14 | Val16 | 0.17 | 2.81 | 147.09 |
| Leu14 | Gln17 | 0.20 | 2.88 | 159.58 |
| Leu14 | Pol18 | 0.14 | 2.87 | 161.57 |
| Val16 | Pol18 | 0.08 | 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 γ-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 γ-turn-inducing bonds like Val16→ Leu14 (17% occurrence) in the chloroform solvent.

#### *5.2.1.2. Clustering based on FEL: vision through PCA*

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 analysed by PCA to reduce the dimensionality of data and to visualize the free-energy landscapes, revealing the intermediate states and their path to achieve the final folded state (Maisuradze et al., 2009). 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 (Sittel et al., 2014), therefore dihedral-angle based PCA (dPCA) (Altis et al., 2007) was employed to only include internal motions (defined by $\phi$, $\psi$ dihedral angles) for the peptide folding process. The FEL based on internal motions (projected along first two principal components PC1 and PC2 using $\mu(q_1, q_2) = -k_BT \ln P(q_1, q_2)$ provides accurate results of the minimum energy wells and barriers between them (as described through scree plots in **Figure S4** for simulation of TPN XIIc in water and **Figure S5** in chloroform, see Appendix), especially for systems undergoing large changes (Mu et al., 2005). All dihedral PCA-based FEL plots have been reweighted for calculation of PMF and accurate description of the energy minimum.

The simulations were carried out for 500 ns × 3 using different starting structures along with another 1000 ns long simulation. All simulations have been combined (2.5 μs) 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 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 16**) show the lowest energy conformation clusters which denote the low energy states of this peptide. As it 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.

In explicit water: The full trajectory was clustered into 10 major representative groups (**Figure 16A**). 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 $mol^{-1}$. The representative structures show major differences in the C-terminal folding, which is a loose spiral in the Clusters 1 and 5, but a highly folded helix in Cluster 2. 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 16B** 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.

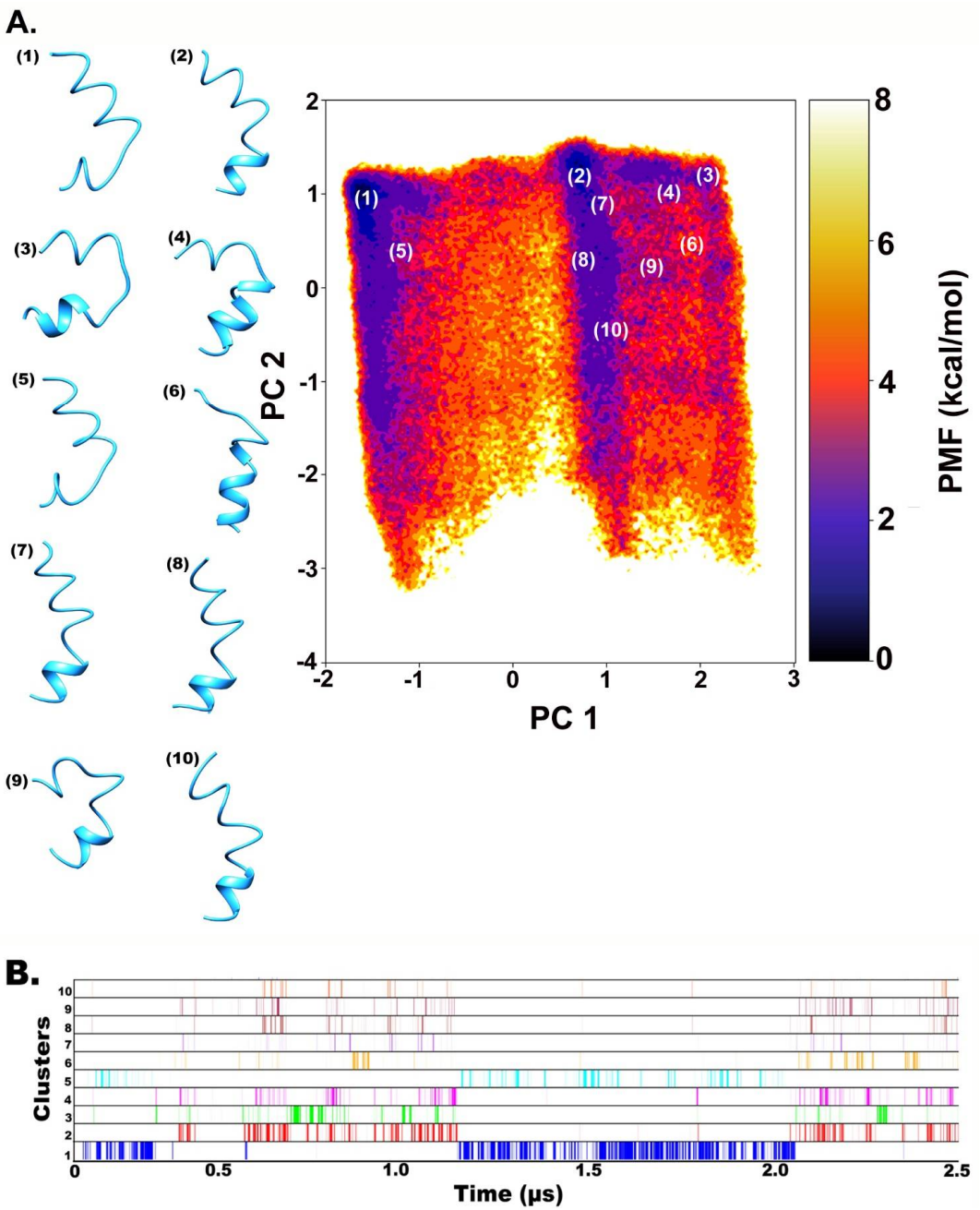


**Figure 16. (A**) Reweighted FEL of the first two PCs calculated from dihedral angles, phi-psi, for better clustering based on internal motions for TPN XIIc simulation in explicit water. **(B)** Diagrammatic representation of cluster distribution along the simulation trajectory in water.

This result suggests that the formation of an α-helical C-terminal fold is an uncommon energy barrier-crossing event and would not have been effectively sampled using short-timescale classical MD approaches. The PMF (in kcal $mol^{-1}$) was calculated 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 favoured (**Figure 17**).

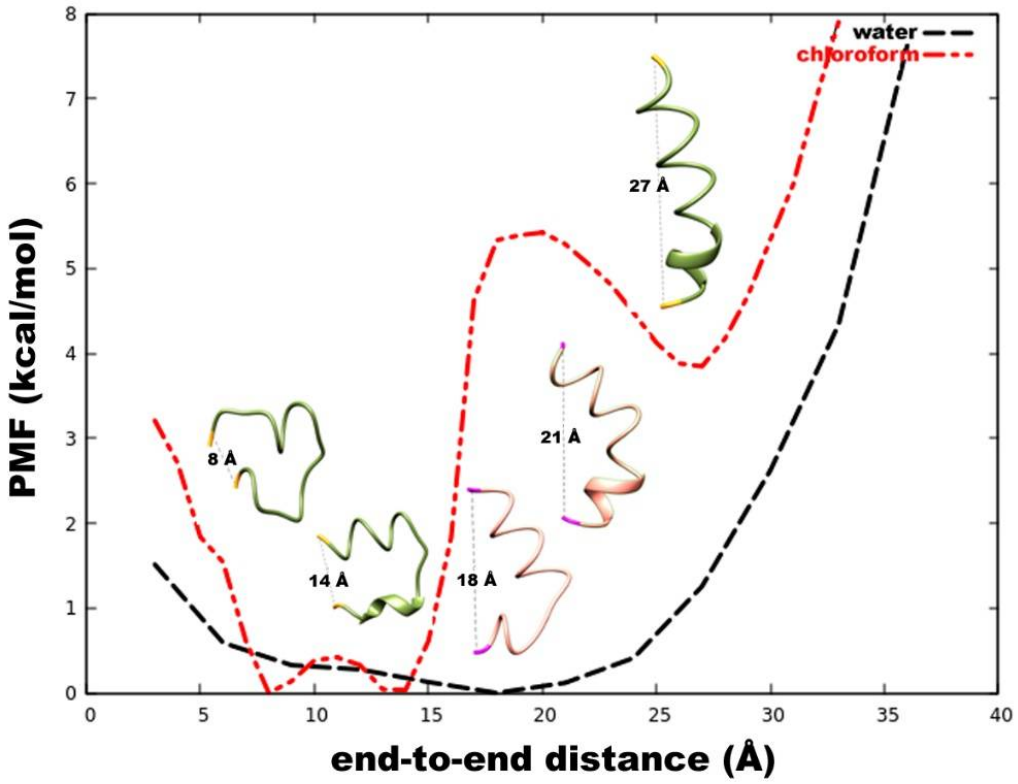


**Figure 17.** PMF values calculated in kcal $mol^{-1}$ as a function of end-to-end distance of TPN XIIc calculated for each step in the trajectory. The PMF values were calculated using Maclaurin series expansion method for reweighting end-to-end distance values. All throughout the study it should be noted that backbone curvature is a rather important structural characteristic of TPN XIIc. For water, the energy minima are obtained for structures that are slightly bent, i.e. end-to-end values are from ~18 to 21 Å as shown by two salmon-tinted peptides. For chloroform, the deepest energy minima are obtained for structures showing complete backbone reversal represented by the two olive-tinted structures, i.e. end-to-end value of 8 Å which denotes unfolded conformation and 14 Å which denotes folded conformation. The third structure with 27 Å is the straight backbone conformation that is obtained after crossing an energy barrier of ~ 5 kcal $mol^{-1}$ and lies at 4 kcal $mol^{-1}$. Such a rare event was not sampled in previous 500 ns long simulations and occurs only during long 1 μs simulation using aggressive boost parameters.

In explicit chloroform (hydrophobic environment): In the case of chloroform solvent (**Figure 18A**), eigth 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 which is observed as the region of deepest minimum denoted by Cluster 4 (13% occurrence throughout the combined trajectory). This is expected due to the presence of γ-turns that cause backbone reversal. It was determined that the Aib12→Ala10 bond results in an energetically stable backbone reversal by $i+1^{th}$ residue Div11 which populates the γ'-region (inverse γ-turn or $C_7^{eq}$ conformation) on the phi-psi distribution plot.

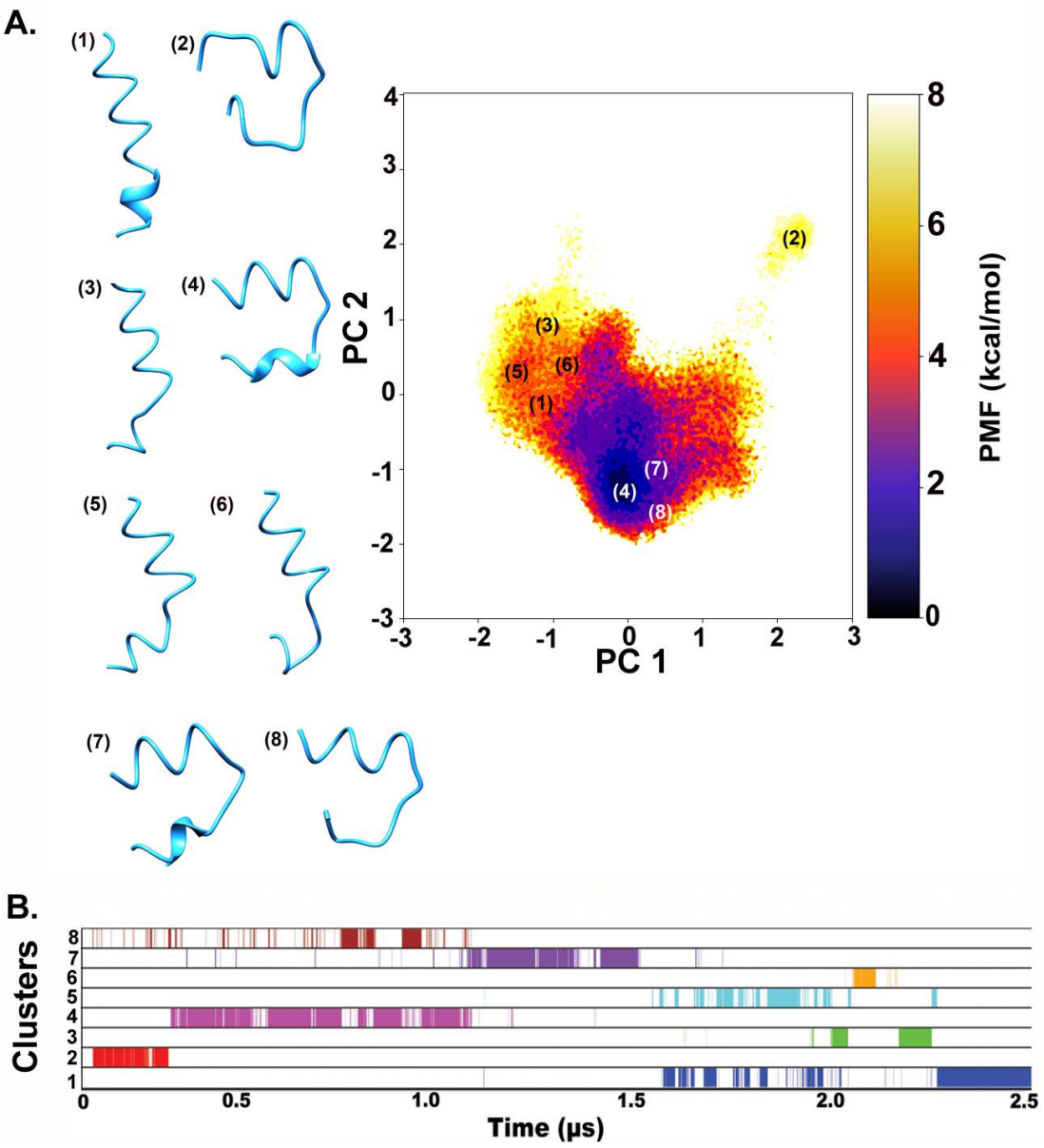


**Figure 18. (A)** Reweighted FEL of the first two PCs calculated from dihedral angles, phi-psi, for better clustering based on internal motions for TPN XIIc simulation in explicit chloroform. **(B)** Diagrammatic representation of cluster distribution along the simulation trajectory in chloroform.

An end-to-end distance of 14 Å (indicating backbone reversal) lies in an energy minimum while the unfolded conformation lies in another minimum (**Figure 17**). The last 1 μs 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 → Cluster 1) is ~4 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 kcal $mol^{-1}$ (**Figure 17**). There is a 6 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 18B**).

#### *5.2.1.3. Addressing the convergence*

A recurrent issue faced during folding simulations is the question of convergence, which in simple terms, asks how long a simulation must be run to reach the convergence of any observable quantity. 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. The early computational biologists tried to answer this by showing a convergence of RMSD calculated between a reference structure of the protein in question with its changing conformation during the course of simulation. Eventually, at some point the graph of RMSD would stabilize and the simulation was shown to be converged. This could hold true for a folded structure formed of strict secondary structural elements. But the same method is not relevant for a microsecond scale folding simulation of a highly dynamic peptide. A relatively new method was designed to compare, instead, whether two simulations for the same protein cover the same phase space eventually.

(a) KLD method

A quantitative measure of extent of overlap between any probability distribution is the KLD method which can indicate satisfactory sampling (Kullback & Leibler, 1951). 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 the conformational states that have been visited during the simulation. Two independent simulations started from different initial geeometries should eventually begin to sample the same conformations (Galindo-Murillo et al., 2015). 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 are regarded as converged.

In this section, we shall uncover the overall motions of the peptide to attain various metastable states. The combined trajectory (2.5 μs) was divided into 5 parts of equal time length, i.e. ~ 500 ns. This means that each individual trajectory and the last 1 μs long

simulation divided in two equal parts were considered. **Figure 19A** 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.

In explicit water: The essential PCs 1, 2 and 3 obtained from 5 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 5 independent trajectories.

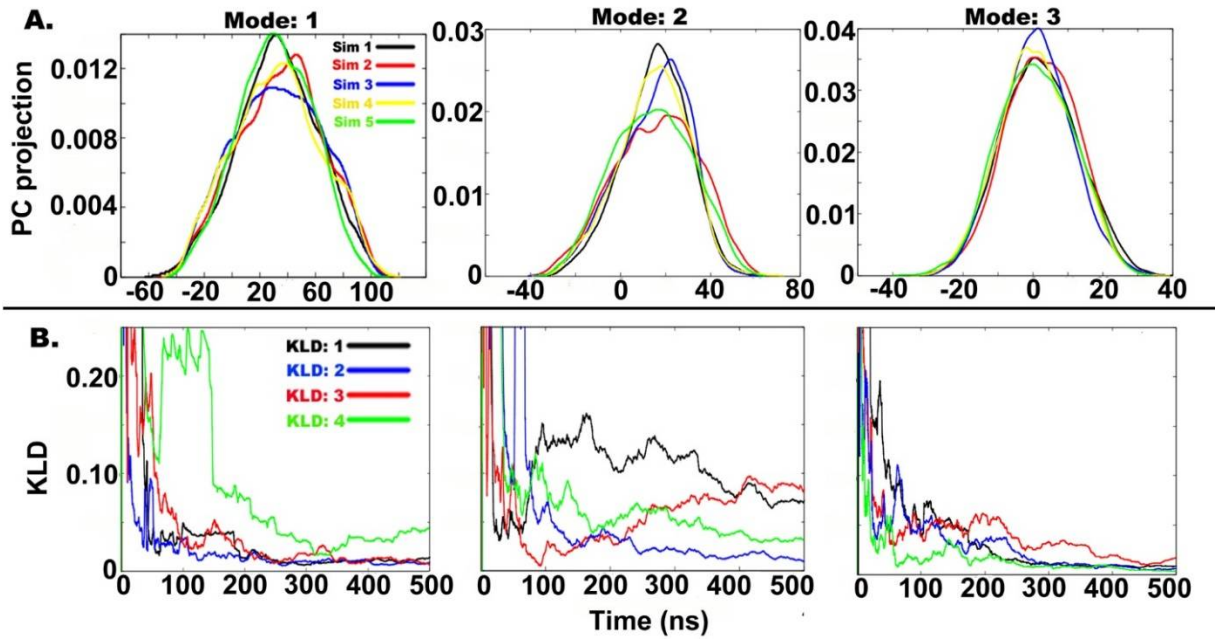


**Figure 19. (A)** Histograms of projection of principal components 1, 2 and 3 for all four simulations of TPN XIIc in water, where the last 1 μs long simulation is treated in 2 parts. Histograms were calculated using a Gaussian kernel density estimator. **(B)** A measure of overlap between histograms from independent simulations calculated using the KLD method. The slope values lying below 0.025 indicate convergence between two independent runs.

**Figure 19B** 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 & Sim2, "KLD:2" between Sim2 & Sim3, "KLD:3" between Sim3 & Sim4 and "KLD:4" between Sim4 & 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 μs 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.

In explicit chloroform: A similar analysis for chloroform solvent divulged some interesting results. **Figure 20A** 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 μs long simulation shows a drastically different mode of motion. However, such a stark difference is not observed for second PC (mode 2). The occurrence of a linear backbone structure in the last 1 μs simulation in contrast to a highly bent structure in previous simulations can be believed to be the reason of this difference. As the peptide undergoes significant motion around the Div11-Aib12-Pro13 hinge region to attain a straight backbone, a new state is achieved which is comparable to structures obtained from water simulation. **Figure 20B** 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 & Sim2 and Sim4 & Sim5, respectively. KLD:2 values for PC1 reach a value of 0.05 at around 150 ns before increasing again. This means that Sim2 & Sim3 initially sampled similar structures before diverging on separate conformational paths. KLD:3 values between Sim3 & Sim4 remain divergent, which is expected due to very different conformational spaces covered during the previous 500 ns long simulations and the latest 1 μs 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 μs simulation in contrast to a highly bent structure in previous simulations can be believed to be the reason of this difference. As the peptide undergoes significant motion around the Div11-Aib12-Pro13 hinge region to attain a straight backbone,

a new 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 two parts of the 1 μs long simulation shows that highly divergent internal motions have occurred in the peptide. In some instances, the KLD value rises again after hitting the cut off 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 μs 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. An important point to note is the application of only dihedral boost to the system which specifically targets the backbone torsional angles and can explain the structural divergence observed in TPN XIIc backbone linearity.

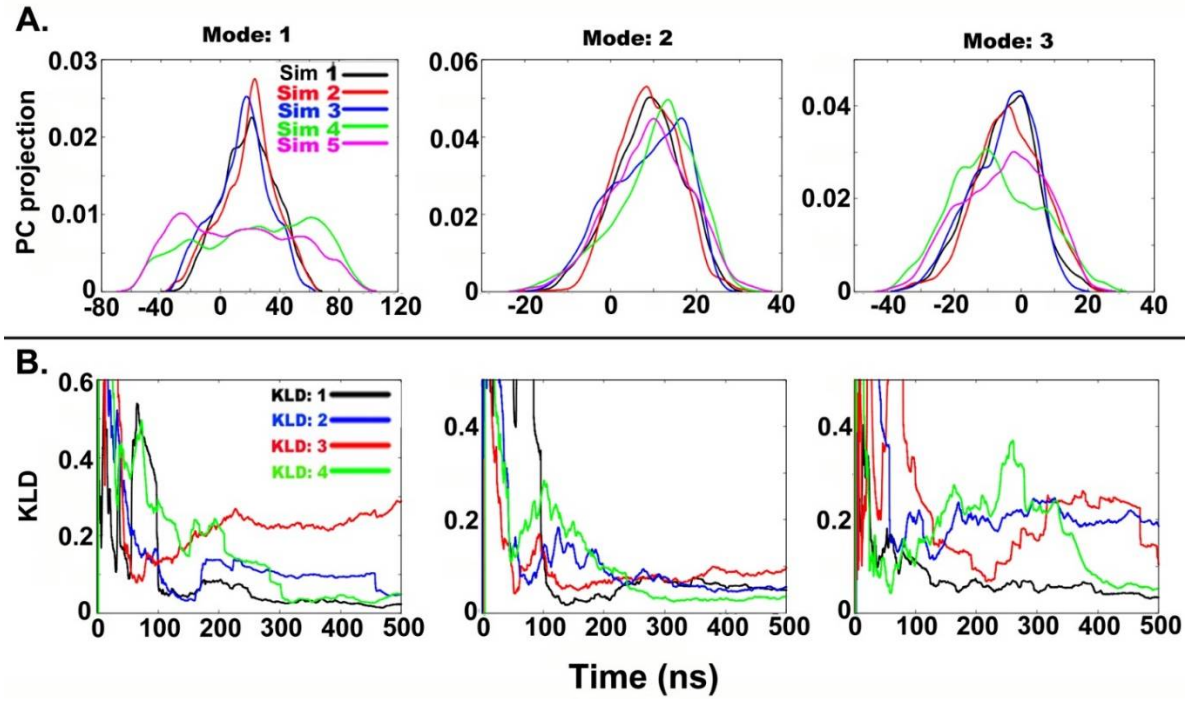


**Figure 20. (A)** Histograms of projection of principal component 1, 2 and 3 from all four simulations of TPN XIIc in chloroform where the last 1 μs long simulation is treated in 2 parts. Histograms were calculated using a Gaussian kernel density estimator. **(B)** A measure of overlap between histograms from independent simulations calculated using KLD method. The slope values lying below 0.025 indicate convergence between two independent runs.

(b) Good-Turing formalism for assessment of convergence

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 (Roe et al., 2014). Based on Good-Turing formalism (Shao et al., 2007; Koukos et al., 2014; Serafeim et al., 2016) applied on the RMSD values it could be deduced, that slightly higher structural convergence is achieved in chloroform than in water (**Figure 21**). In the case of water simulation, the most distinct structure we can expect to observe if we double the simulation time will differ by no more than approximately 3.9 Å (RMSD) from those already observed. To look at it another way, one out of every five (probability of unobserved species, P_unobs = 0.20) new structures encountered will differ by an RMSD of at least 3 Å.

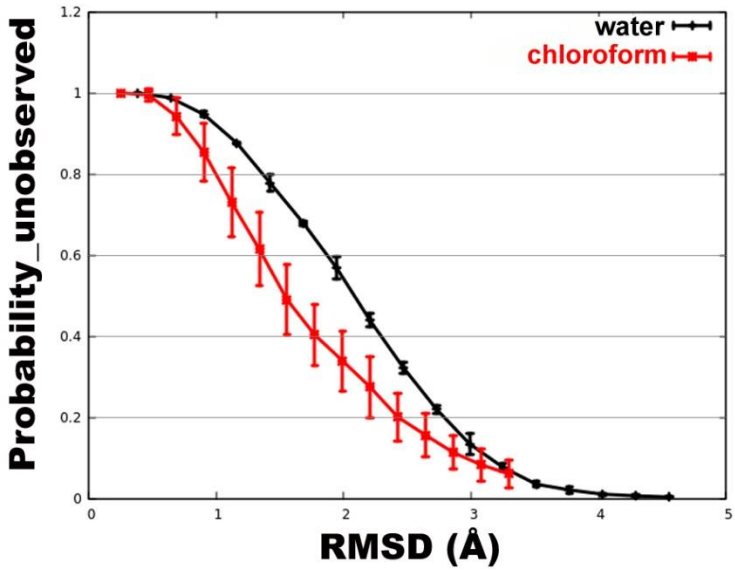


**Figure 21.** Estimation of extent of sampling of TPN XIIc using aMD based on Good-Turing formalism in water and chloroform. The probability of unobserved species reaches 0 at the maximum RMSD value that could be observed if the simulation is extended. Evidently, this value seems to be higher in water than in chloroform.

On the other hand, simulation in chloroform indicated towards slightly higher convergence as the maximum RMSD value of an unobserved species was calculated to be only 3.07 Å, i.e. one in eleven structures (P_unobs = 0.11) shall differ only by 2.8 Å from the conformations already observed. Keep in mind that this calculation includes all possible unfolded conformations and the high RMSD probably indicates towards different unfolded structures that have not been encountered. But such unfolded states are not required when we are studying the folding dynamics of a peptide. This method is an assessment of structural convergence of peptide dynamics based on the statistics obtained from previously sampled

conformations. It is a straightforward method using RMSD as the base value which is an easy but not always accurate measure of structural similarity. Nevertheless, such an analysis along with convergence of the principal component space indicates that all major low energy and metastable states of the TPN XIIc peptide in water and chloroform have been sampled using accelerated MD. This indicates that the folding of the TPN XIIc peptide is very dynamic in aqueous environment and probably can result in many unfolded or intermediate states, even though, all energetically stable states have already been achieved. To reiterate, higher convergence shown in chloroform can be attributed to restricted dynamics of TPN XIIc unable to escape an energy barrier. 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.

### 5.2.2. Alamethicin F30/3: a model peptaibol

#### *5.2.2.1. Comparison of four consecutive aMD simulations with Alm F30/3*

The unfolded conformation of Alm F30/3 was used as a starting point for three consecutive, ~900 ns long simulations (refer to **Table 12** for exact timescales) with different "boost" parameters.

**Table 12. Summary of various accelerated molecular dynamics parameters applied for Alm F30/3 simulation.**

| Simulations | Starting conformation | Time (ns) | a1,a2 total (kcal mol$^{-1}$) | b1,b2 dihedral (kcal mol$^{-1}$) | Avg boost (kcal mol−1) |
|---|---|---|---|---|---|
| **Sim 1** | Unfolded | 936 | 0.16 | 4 | 11.21 |
| **Sim 2** | Folded N-terminal | 950 | 0.20 | 4.5 | 11.10 |
| **Sim 3** | Folded with bent backbone | 897 | 0.20 | 4.5 | 29.85 |
| **Sim 4** | unfolded | 1000 | 0.20 | 4.5 | 24.09 |

The first simulation (Sim 1) was carried out with a1 = a2 = 0.16 and b1 = b2 = 4 which revealed successful folding of the N-terminal segment (Aib1-Leu12) but an incomplete folding of the C-terminus after dPCA-based clustering as shown through superimposition of the representative structure of the energy minimum and experimentally known structure (PDB ID: 1AMT) with backbone RMSD value of 5.02 Å in **Figure 22A**. The next simulation (Sim 2) was started from this point with slightly higher boost parameters of a1= a2 = 0.20 and b1= b2 = 4.5 to observe the time length of achieving complete peptide folding. This 950 ns long aMD simulation was clustered into three groups, out of which cluster 2 was closest to the experimental structure (**Figure 22B**) with RMSD of 1.87 Å. At this point, it was deemed a better choice not to increase the boost further as it may interfere with correct reweighting of

energy distribution. These boost parameters were deemed appropriate for fast folding. A third ~900 ns long simulation (Sim 3) was carried out at this point which achieved complete C-terminal folding. The clustering resulted in the most populated cluster whose representative structure is the closest to PDB conformation with a RMSD value of 1.51 Å (**Figure 22C**). After carrying out three consecutive aMD simulations with increasing boost parameters, we were also curious to observe the extent of folding observed using these boost parameters with a completely unfolded Alm configuration as the starting structure. A separate 1 μs long simulation (Sim 4) was carried out that resulted in highly folded structures with both bent and linear configurations. This simulation shows that ~1 μs long aMD carried out using GPUs using slightly aggressive boost parameters are sufficient for folding simulations of such short peptaibols. The representative structure of cluster 5 is closer to the experimental structure with an RMSD value of 1.8 Å between them (**Figure 22D**).

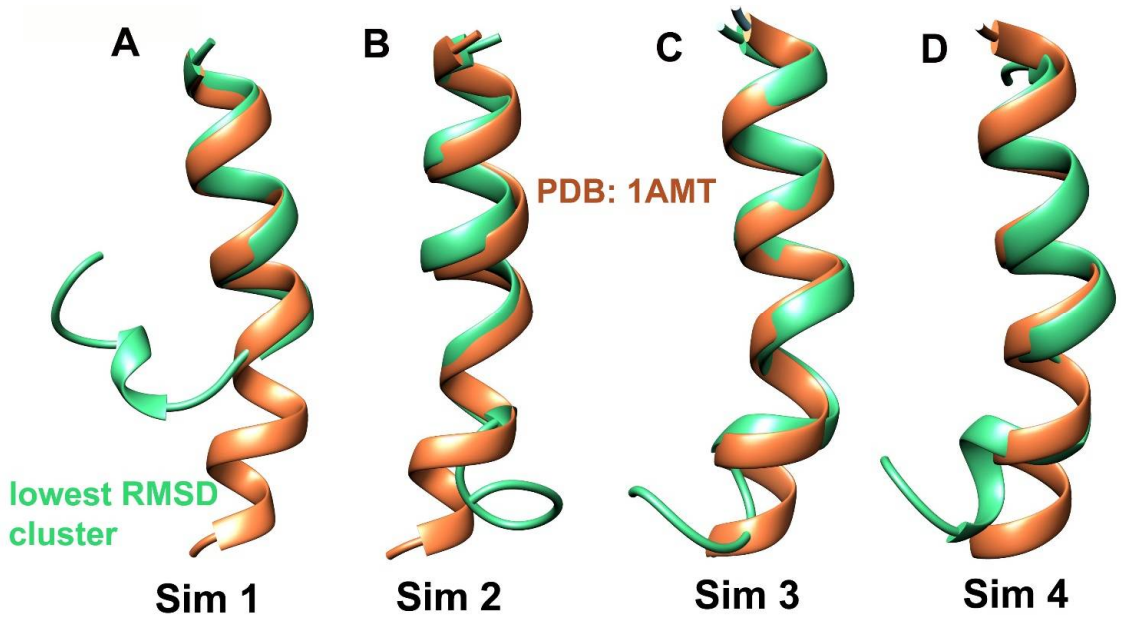


**Figure 22.** The representative structures of different clusters obtained for individual simulation chosen based on lowest RMSD with X-ray crystallographic Alm F30/3 structure available from PDB: 1AMT.

In conclusion, the combined trajectory of the first three simulations is comparable to the fourth ~1 μs long simulation carried out with aggressive boost parameters. The three trajectories were later combined for most of the analysis. The reweighted torsional (phi-psi) angle distribution (Ramachandran plot) for each residue calculated for all these three simulations have been provided and discussed in **Figure 23**. Complete Alm F30/3 folding was achieved within 1 μs, which indicates that this procedure is apt to elucidate short

peptaibol structure in a short period of time. Bucci et al. (2019) also reasoned that 1 μs long aMD simulations are sufficient for folding simulations of their modified tripeptide.

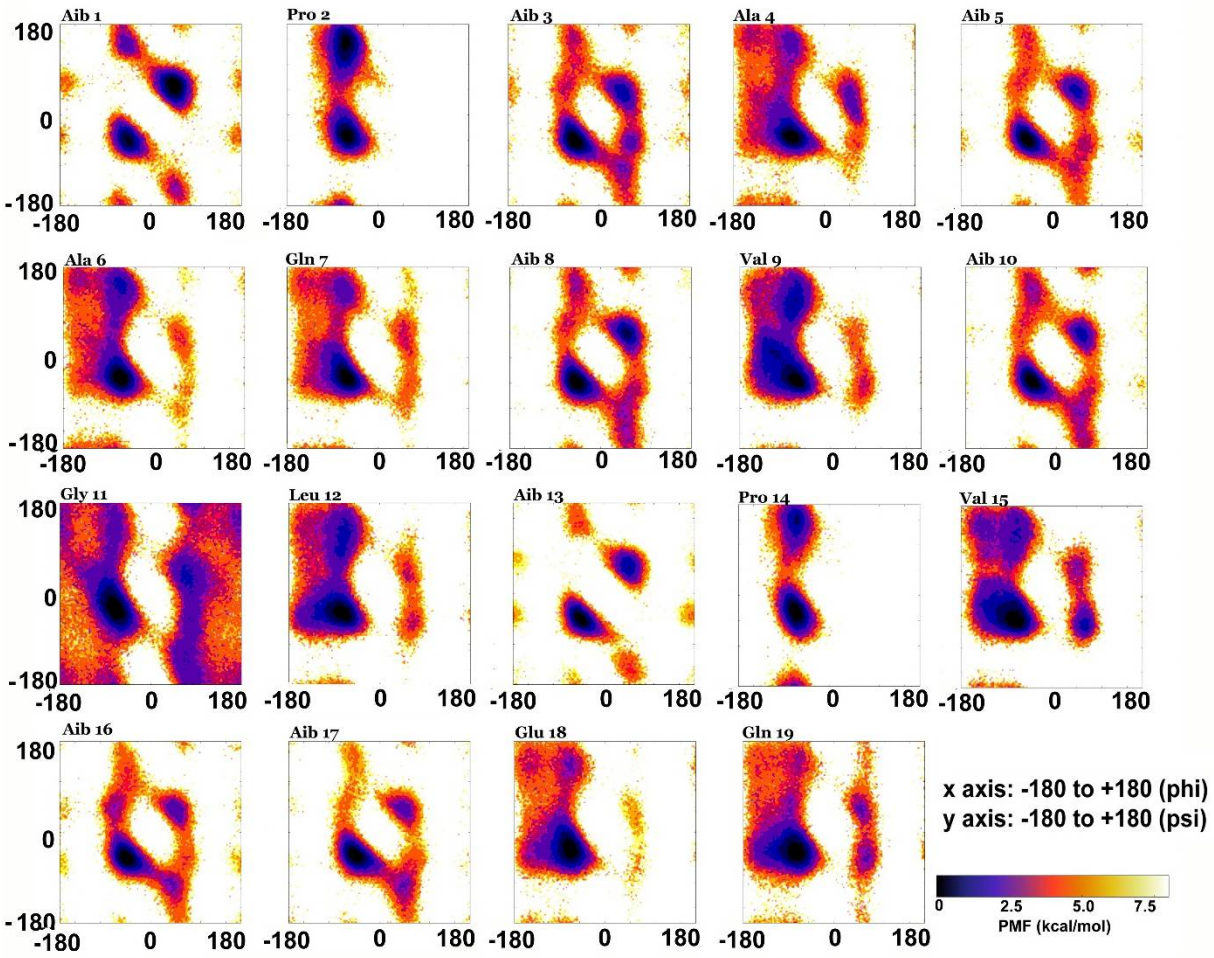


**Figure 23.** The reweighted phi-psi torsional distribution for each residue of Alm F30/3 averaged over the three 900 ns simulations. As can be observed, most residues show an energy minimum in the α-helical region. All Aib residues shift between right-handed and left-handed helical regions owing to their achiral nature. The Gly11 residue covers a huge area on the plot owing to the absence of a side-chain. The distribution was reweighted using the Maclaurin expansion method.

#### *5.2.2.2. Clustering and reweighted FEL*

The dCPA based FEL between PC1 and PC2 has been provided for the combined trajectory of the first three simulations clustered in 10 major representative groups (**Figure 24A**). As it can be observed on the reweighted maps, at least three distinct clusters can be identified in local energy minimum regions, i.e., the linear helical form (clusters 1, 5 and 9), the bent form (clusters 2, 4 and 6) and the incompletely folded (clusters 3, 7, 8 and 10) conformations. The inter-conversion between the bent and linear forms seems to be energetically allowed within 2 kcal $mol^{-1}$ and the jump between these two is achieved multiple times throughout the

trajectory. The C-terminus requires longer time than the N-terminus to completely sample the folded helical configuration.

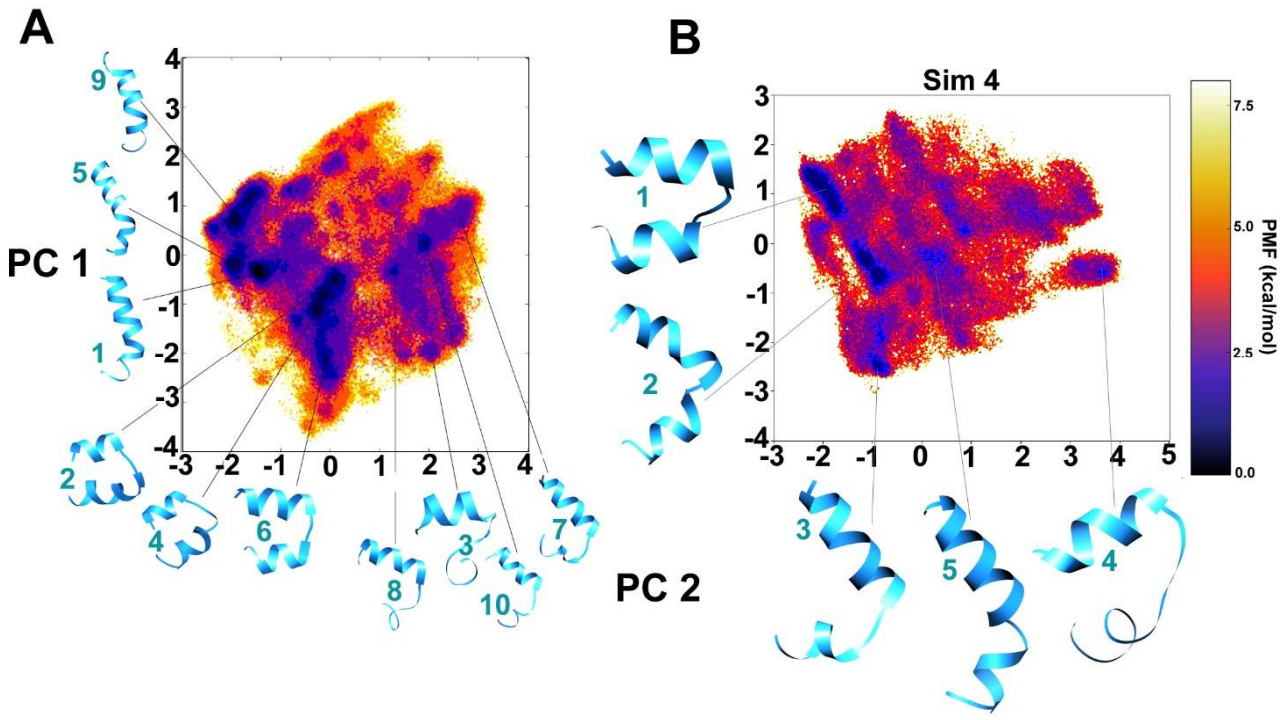


**Figure 24.** Reweighted PMF landscapes of the first two PCs calculated from dihedral angles, phi-psi, for better clustering based on internal motions. The deepest blue regions indicate energy minimum. The representative structures of each cluster obtained for Alm F30/3 are provided for **(A)** combined 3 simulations, **(B)** the 4$^{th}$ simulation. The cluster numbers are given in cyan.

Similarly, the fourth trajectory has been separately projected on to dihedral-based PCs in **Figure 24B** to avoid missing low energy conformations obtained during previous simulations. The bent conformations of clusters 1 and 2 appear interchangeably for about ~10% and ~8% of 1 μs long simulation trajectory, respectively, roughly from 300 to 800 ns and form the most populated conformational group. The conformation of cluster 5 is closest to the experimental structure. The dihedral angle based FEL plots of the combined trajectory from the first three simulations in comparison with the fourth 1 μs long trajectory shows that a similar conformational space could be covered in less time if slightly aggressive boost parameters are used.

#### *5.2.2.3. Addressing the convergence*

In order to observe structural convergence amongst the three independent simulations, KLD was calculated to measure the extent of overlap between probability distribution (Kullback &

Leibler, 1951). This probability distribution here is based on PCA of the Cartesian coordinates instead of dihedral angles. The trajectories are named as “Sim 1” to “Sim 4” and the extent of PC overlap signifies convergence between independent runs. The degree of PC overlap suggests that the independent simulations sampled similar conformational space, hence, convergence. It is based on the idea that any two simulations will eventually sample the same phase space even when started from different starting structures as a measure of true convergence. **Figure 25A** shows the histograms of projection of coordinates along the first three eigenvectors or modes (accounting for 53% overall motion). The first three PCs obtained from the combined trajectory account for 29%, 16% and 8% (in same order) of overall motions. In case of all PC histograms, the best overlap can be observed between Sim 1, Sim 3 and Sim 4 in comparison to Sim 2 projection which signifies that Sim 2 undergoes a slightly different folding pathway. A better overlap can be observed for all trajectories in case of modes 2 and 3. This divergence can be quantified by calculating KLD vs time (ns).

**Figure 25B** shows KLD between subsequent histograms as a function of time from four simulations for PCs 1, 2 and 3. KLD:1 is divergence between Sim 1 and Sim 2, KLD:2 between Sim 2 and Sim 3 and KLD:3 between Sim 3 and Sim 4. The rapidly decreasing slope of KLD vs time of any two trajectories indicates the reduction in divergence between sampled conformations. The KLD value of 0.025 was chosen as cutoff for convergence. As evident from **Figure 25B**, convergence for Mode 1 for all four trajectories, except Sim 2, is obtained only after 700 ns. KLD:2 (divergence between first PC of Sim 2 and Sim 3) shows the highest divergence for Mode 1 even though both simulations were started with the same boost parameters but different starting structures. This also indicates that Sim 2 undergoes a different path of Alm F30/3 folding.

KLD:3 (divergence between first PC of Sim 3 and Sim 4) reaches the threshold at ~400 ns. The two trajectories, the former with a semi-folded starting structure and the latter with completely unfolded structure, evolve quickly and sample similar configurations to achieve close to experimental Alm F30/3 structure. In the case of Mode 2 and Mode 3, the convergence between all three trajectories is reached at ~400 ns time scale. Using boost parameters a1 = a2 = 0.20 and b1 = b2 = 4.5 for ~1 μs proved to be appropriate to achieve the near-native Alm F30/3 conformation starting with the unfolded Alm F30/3 structure. A similar approach can be adopted to model other peptaibols of unknown structure. The convergence of all simulations was proven based on the KLD method which showed that Sim

2 follows the most divergent path of conformational folding and takes longer timescales to converge, while the other three show structural convergence within 400 ns.

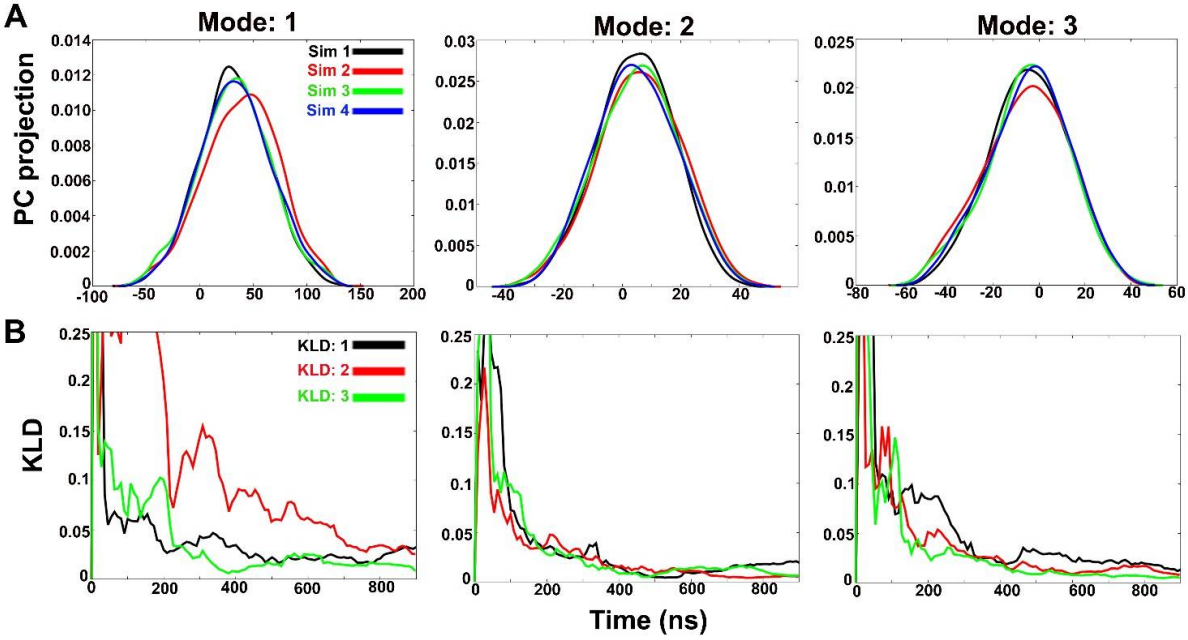


**Figure 25. (A)** Histograms of projection of principal components 1, 2 and 3 for all four simulations of Alm F30/3 in water. Histograms were calculated using a Gaussian kernel density estimator. **(B)** A measure of overlap between histograms from independent simulations calculated using KLD method. The slope values lying below 0.025 indicate convergence between two independent runs.

In the previous section we observed at least two very distinct Alm F30/3 conformations: linear and bent helices. We were curious to observe the phenomenon of helical bending obtained through aMD. Here we discuss whether there is a functional importance of such dynamic structural shift. This knowledge may direct us to understand how these peptides show membrane-perturbing properties. The end-to-end distance (in Å) between the first residue Aib1 (N-terminus) and Pol20 (C-terminus) was calculated for each frame of the combined trajectory. This data was used to calculate the PMF (in kcal $mol^{-1}$), which is simply the change in free energy as a function of any reaction coordinate. The PMF describes the energy minimum as the most stable state along that function. The end-to-end distance values of 9 and 10 Å designated by highly bent helical conformations (**Figure 26**) indicate their stability. The distance values of ~25 Å indicated by linear Alm backbone conformations are easily accessible with under ~1 kcal $mol^{-1}$ of energy boost.

Using paramagnetic enhancements of nuclear relaxation, North et al. (1994) demonstrated that the Alm backbone undergoes large structural fluctuations that result in shorter distances

between the C-terminus and various positions along the backbone. They linked this observation with the voltage-gating mechanism of the Alm channel. A previous study by Franklin et al. (1994) showed that simulated annealing with NMR restraints of Alm peptide bound to micelles yielded both bent and linear conformations, which prompted them to confirm their analysis by appending a spin label to one of the bent conformations and energy minimization using the steepest descent method. The same bent conformation was obtained as the energy minimum each time.

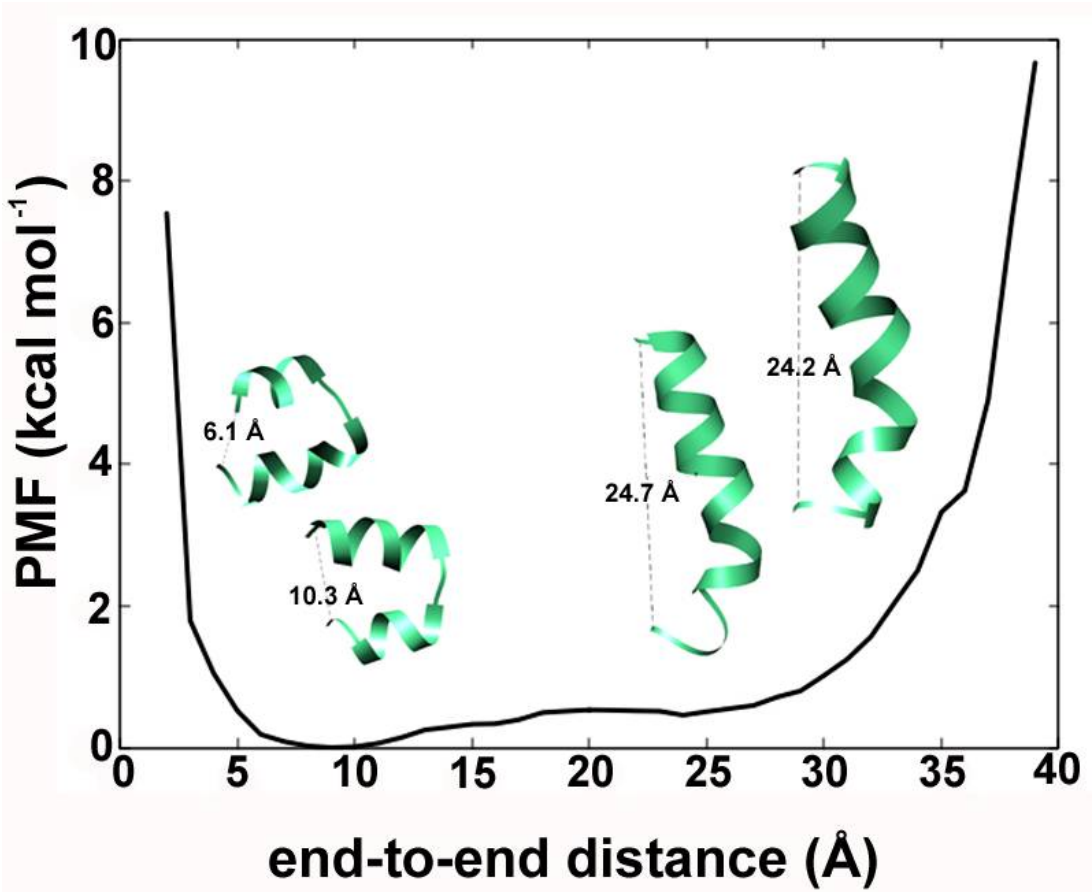


**Figure 26.** The reweighted PMF values in kcal $mol^{-1}$ as a function of end-to-end distance of Alm F30/3 calculated for each step in the combined trajectory. The energy minimum lies for a distance value of 10 Å which denotes a highly bent backbone.

Comparing these observations with previous studies, North et al. (1995) reasoned that Alm must be in a dynamic equilibrium of linear and bent conformations and that it may provide the "conformational switch" of voltage gating in the Alm channel. In other words, the bent/closed form of Alm bound to membrane may indicate the absence of transmembrane voltage, which – when applied – would allow conversion to the linear and amphipathic Alm conformation. Gibbs et al. 1997 reported on the phenomenon of helical bending around

residues Aib10-Aib13 of Alm observed during 1 ns long simulation in methanol. The structural states obtained had either Aib10 or Gly11 carbonyl group oriented away from the backbone and did not seem to greatly affect the adjacent helix structure. The functional role of the helical bend in channel formation was hypothesized. The results clearly show that the bent helix lies in the true energy minimum while the linear is easily accessible. Our analysis confirms the idea of a dynamic equilibrium between the two conformations and indicates functional relevance.

Moreover, Balashova et al. (2000) reported the crystal structure of a 16-mer peptide Zrv-IIb isolated from *Emericellopsis salmosynnemata*, that has been known to be active against bacteria while being non-toxic to eukaryotic cells (Argoudelis et al., 1974). It folds into a bent (40˚ angle) amphiphilic helix in both polar (methanol) and non-polar (chloroform) solvents. Chugh et al. (2002) described the crystal structure of an 18-residue long trichotoxin_A50E as a completely helical structure with a central bend arising at residues 10-13 where Gln residues at positions 6 and 17 align at the polar face of the transmembrane channel lumen. They proposed an octameric channel with 4-5 Å internal radius sufficient to transport ions, and a constricted ring formed by Gln residues. However, posterior a study by Duclohier et al. (2004) of the same group reported a hexameric channel that displayed a single conductance level in contrast to alamethicin that displays multiple conductance states in the experiments. Similarly, another transmembrane channel-forming 19-residue peptaibol, chrysospermin C was studied using NMR, which revealed well defined helices at the N- and C-termini with a ~38˚ angle bend at residues 10-12. Gessmann et al. (2012b) also reported a significantly bent $3_{10}$ right-handed helical conformation of the 14-mer trichovirin I-4A and proposed a possible transition into α-helical conformation during membrane integration. From these few experiments it is clear, that peptaibols generally fold into bent helical conformations and support the findings of this study.

An important observation comes from the fact that most peptaibols fold in such a way that one side, usually the concave one, has the hydrophobic residues lined up, while all hydrophilic residues are lined up on the convex face. The hydrophobicity of the concave side causes the extreme bending of backbone to avoid the aqueous environment in which it is placed. It is understandable as to why the highly bent, hairpin-like backbone falls in the true energy minima for most of the peptaibols instead of the linear backbone. The same phenomenon is observed universally in proteins where the hydrophobic residues make the

core of the globular structure and the hydrophilic residues are surfaced facing the solvent. The importance of Gln residues that line the hydrophilic lumen of transmembrane pore in channel stability and ion permeation has been highlighted by Asami et al, 2002. The Gln residues at $7^{th}$ and $18^{th}$ positions of Alm f30 and Alm f50 were replaced with Glu that resulted in reduced stability of channel structure upon ionization of Glu7 and established their importance in peptaibol structure. Highly bent backbones were also obtained during studies with other fungal peptaibols, for example, TRK-V, TPN XIIc, Par-B and H, and the newly discovered Brev-I and IV molecules.

The hairpin-like helical conformation, which is representative for Clusters 2, 4 and 6 (**Figure 24A**) can be explained due to the presence of the glycine residue at the R11 position. The N-terminal helical continuity in Alm F30/3 always breaks at the Gly11 position. Högel et al. (2018) systematically described the local helix bending observed at the glycine position that effectively perturbs the conformational flexibility in transmembrane helices. Glycine does not have a side-chain and can easily conform into many energetically stable Φ-Ψ torsional states as can be seen on the reweighted energy landscape of Gly11 in **Figure 23**. The importance of presence of proline was highlighted by Nagaoka et al., 1996 by substituting $Pro^{14} \rightarrow Aib^{14}$ which resulted in loss of the central bend and shorter life times and fewer substates during single-channel measurements. Moreover, Dathe et al, 1998, showed that the presence of Pro exactly at position 14 in alamethicin was crucial for its hemolytic activity, stimulation of catecholamine secretion and induction of metabolic activity in endothelial cells owing to changes in membrane affinity instead of conformational change. This highlights the importance of the presence of helix-breaking residues like glycine and proline in peptaibol sequences which provides an energetically feasible way to cause the backbone motion to opt stable conformations when placed in solvents of different hydrophobicity factors. The movement along the backbone from bent to linear conformation is the largest scale of motion which indicates that peptaibols like Alm are capable of adjusting their backbone bend in response to bilayer thickness or under the effect of transmembrane potential.

### 5.2.3. Re-evaluating TRK-V conformation using aMD

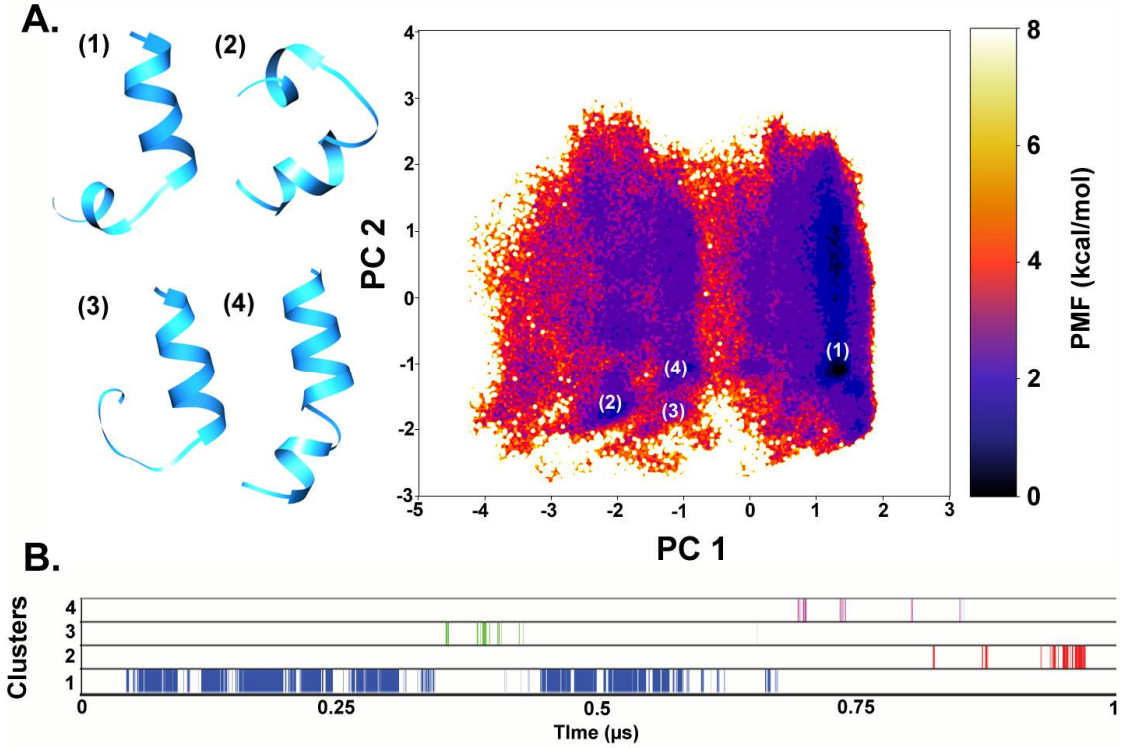


**Figure 27. (A)** Reweighted FEL of the first two principal components calculated from dihedral angles, phi-psi, for better clustering based on internal motions. The representative structures of TRK-V corresponding to various energy minima have also been provided, **(B)** Diagrammatic representation of cluster distribution along the simulation trajectory.

Using the newly optimized aMD parameters, the previously reported TRK-V molecule was elucidated again to obtain the complete canonical ensemble and compare with short classical MD conducted earlier (in **Section 5.1.1.1.**). **Figure 27A** describes the reweighted free-energy landscape of TRK-V as obtained from the two major components of dihedral PCA of 1 μs long aMD simulation. It is clear that the largest cluster lies at the energy minimum and the corresponding representative structure presumes a helical shape with the C-terminus showing hinge-like bend. The loss of helical fold before the C-terminus is responsible for this hinge-like motion and is a characteristic of the Aib-Pro bond found in all peptaibols. The next two largest clusters 2 and 3 correspond to highly bent and C-terminus loss-of-helix conformations, respectively, which probably indicates intermediate states. The 4$^{th}$ cluster with the smallest population lies at a separate region on the free-energy landscape and corresponds to the highly helical, slightly bent conformation. This structure is most likely to be obtained using experimental methods like X-ray crystallography. The 1$^{st}$ cluster appears to populate a separate region on the FEL map which can be accessed under 2 kcal mol$^{-1}$. Its representative

structure can be explained by the presence of a helix-destabilizing residue Ile16 at the C-terminus (Lyu et al., 1991). Their occurrence along the course of simulation can be visualized in **Figure 27B**.

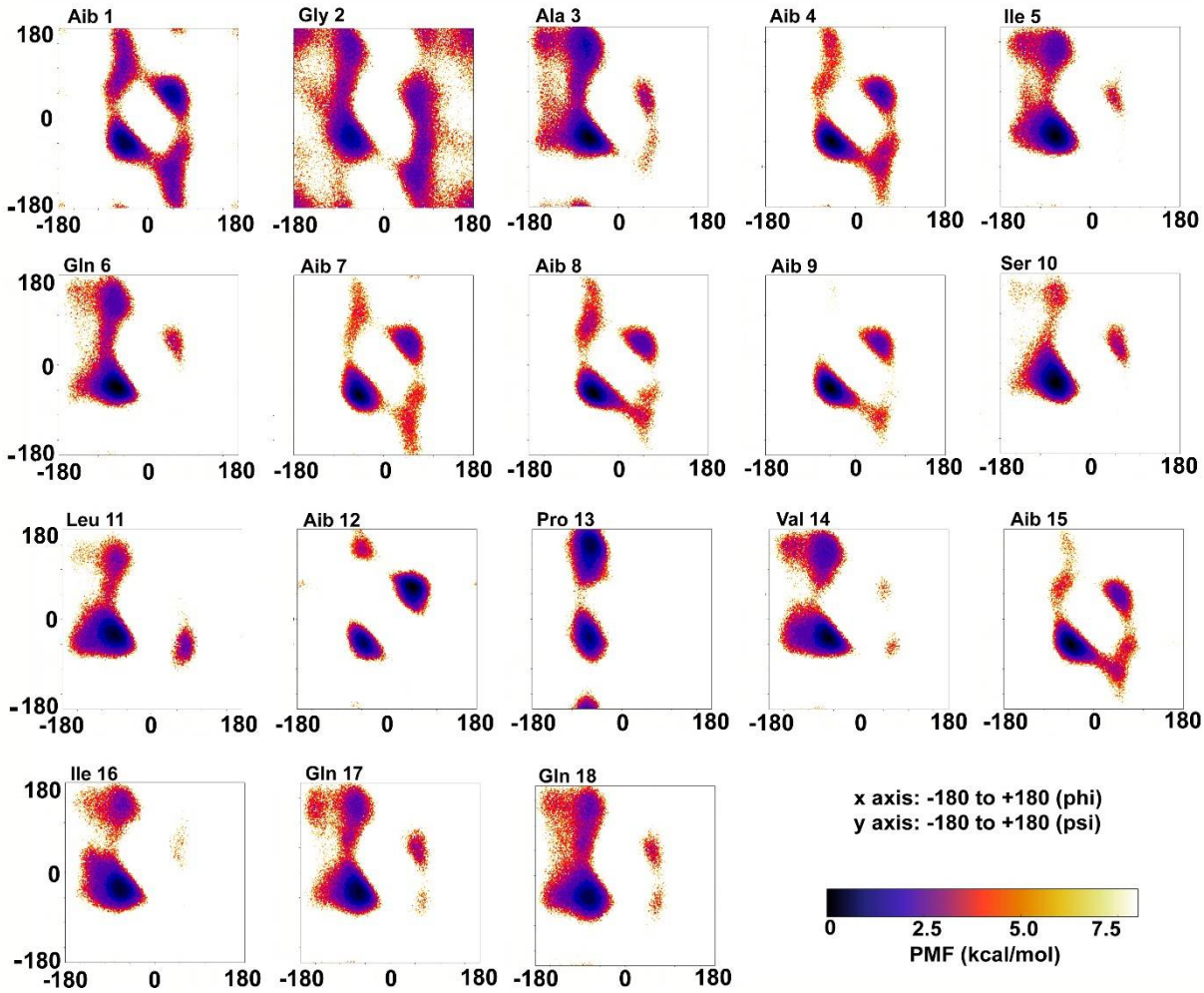


**Figure 28.** The reweighted phi-psi torsional distribution for each residue of TRK-V. As can be observed, most residues show an energy minimum in the α-helical region.

Similarly, upon comparing the reweighted phi-psi populations in **Figure 28** with the same plots from **Figure 7** (classical MD), it becomes clear, that although, most residues like Aib1-Ile5, Ser10, Leu11, Pro13, and Gln17-Gln18 lie in the same free energy minima, many others flanking the central region like Gln6-Aib9, Aib12, and Val14-Ile16 show shifts from the predominant left-handed helical regions to the right-handed helical region. It is understandable for all Aib residues as the probability of its lying in both left-handed and right-handed helical quadrants is the same due to its achiral nature. A significant shift was

observed in the free energy minima of Gln6 and Val14 residues from the left-handed to the right-handed helical region.

### 5.2.4. Paracelsins B and H and brevicelsins I and IV: comparative analysis

The first paracelsins A-D, isolated from *T. reesei*, were reported by Brückner & Graf (1983). The same group characterized their structure and activity (Brückner et al., 1984). Paracelsin was named after Paracelsus or *Theophrastus von Hohenheim,* a Swiss physician, alchemist and astrologer of the German Renaissance. Circular dichroism spectra obtained for paracelsins A-D were reported to be almost identical to that of alamethicin and suzukacillin. Unlike alamethicin, paracelsin and suzukacillin contain only one proline residue. Later, Pócsfalvi et al. (1997) reported sequences of the four new paracelsins F-I and discussed microheterogeneity. In this study, we also report plausible three-dimensional structures of paracelsins B (Par-B) and H (Par-H) for comparison with their 19-residue counterparts, the newly identified brevicelsins I (Brev-I) and IV (Brev-IV) (Marik et al. 2019). We reached an average of 150 ns/day of simulation efficiency with paracelsins and brevicelsins upon 1 μs long aMD simulation.

Our aim was to observe structural differences resulting from the loss of Ala at the R6 position, thereby, a comparison of the 19- and 20-residue long peptides. The length of the folded peptide can have direct effect on its activity against biological membranes in such a way that longer peptaibols are known to adopt transmembrane orientation while the shorter ones do not. For example, while alamethicin was shown to adopt a transmembrane configuration in POPC and 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) membranes, the shorter ampullosporin A, a 16-mer peptide, could only orient parallel to membrane surface of the same thickness mimicking natural membranes. It is in accordance with the fact that ampullosporin A shows much weaker antimicrobial effect (Ritzau et al., 1997) than alamethicin which may be explained by a more detergent-like effect of smaller peptaibols in membranes instead of transmembrane channel forming abilities (Salnikov et al., 2009). All peptides show a strong tendency to form right-handed helical structures with a slight bend at the Aib-Pro position (**Figure 29**). Cluster analysis of the simulation trajectories of all four peptaibols revealed different energetically stable conformations that occur during the folding, and the representative structures of the most populated cluster are provided for each peptaibol (except for Par-H where the structure closest to paracelsin-X has been reported). All peptides fold into an energetically favoured, highly bent helical conformation and a linear helical conformation.

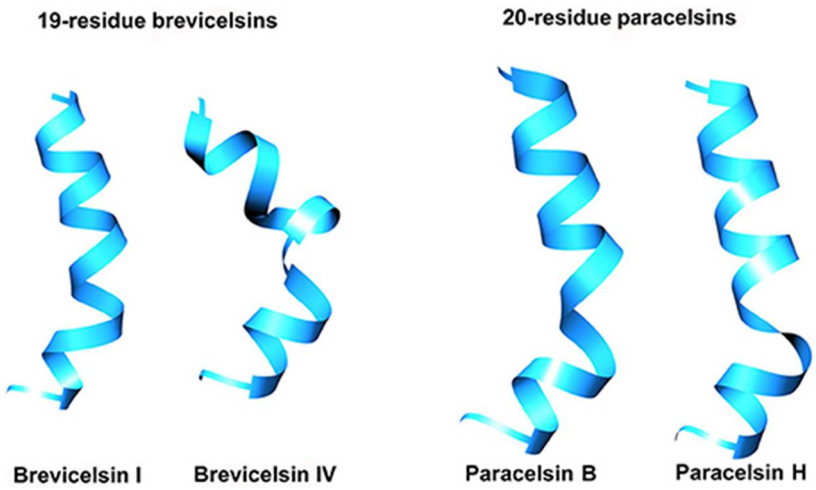


**Figure 29**. Representative structures of the most populated clusters for peptaibols, Brev-I and IV and Par-B and H (the second most populated for paracelsin H). The simulations were carried out for 1,000 ns in each case.

Based on the reweighted PMF values calculated for end-to-end distance (distance in Å from the N-terminus to the C-terminus), it can be speculated that a highly curved conformation for all peptaibols, except for Par-H, lies in the energy minimum and requires an energy "jump" of <1 kcal $mol^{-1}$ to attain the linear backbone conformation (**Figure 30A**).

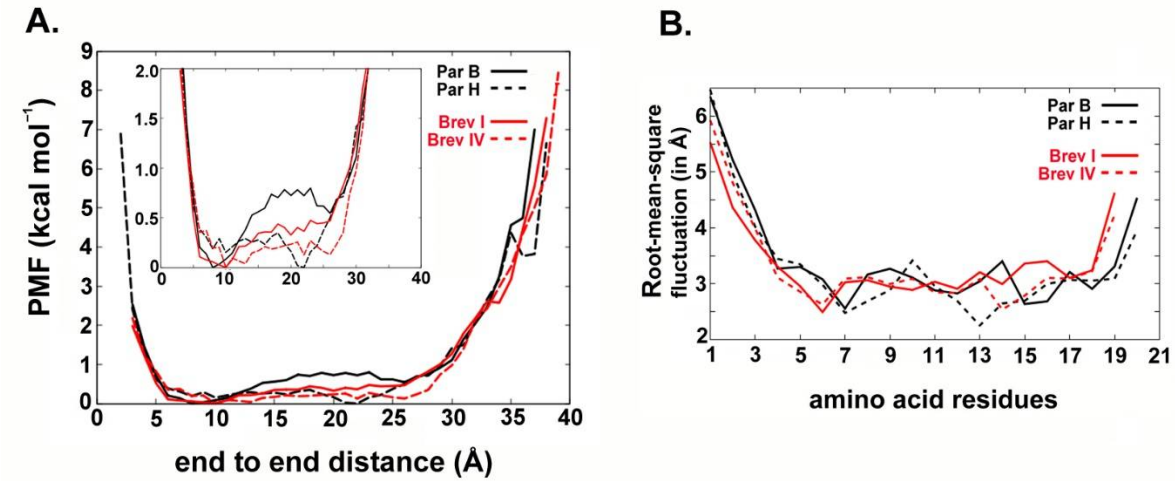


**Figure 30. (A)** The reweighted PMF (kcal $mol^{-1}$) values calculated for end-to-end distance (Å). The 20-residue paracelsins have been plotted in black, whereas 19-residue brevicelsins are shown in red. The zoomed version of the plot is provided in the inset. The end-to-end values where PMF is zero or close to zero are the energetically stable values and define the most favourable linear conformation. **(B)** The RMSD (Å) calculated for each residue for all sequences gives an idea of the average fluctuation undergone by the system. The 19-residue brevicelsins show higher fluctuation than their 20-residue paracelsin counterparts. Par, paracelsin; Brev, brevicelsin.

Overall, the end-to-end distance values from 5 to 27 Å, that lie close to the energy minima, show that all conformations starting from a hairpin-like helix structure to a straight backbone with just a slight bend are easily accessible. The PMF values increase rapidly beyond these two points for all four peptaibols, as shown in the inset image focusing only on PMF values up to 2 kcal $mol^{-1}$. However, the sequences Par-B and Brev-I, with an Aib residue in position R17, have higher PMF values for higher end-to-end distance values; the energy cost for attaining linearity of the helical backbone is slightly higher than in Par-H and Brev-IV, where a Val residue replaces Aib in the R17 position. The presence of Aib residue in position R17 (in Par-B and Brev-I) results in a highly dynamic folding process, which means that many conformations were visited during the trajectory, whereas Val in the same position (in Par-H and Brev-IV) led to fewer energetically stable conformers.

The root-mean-square-atomic fluctuation (RMSF) graph (**Figure 30B**) shows higher fluctuation of N- and C-terminal regions of all peptides in comparison to their central regions. There are no significant differences observed between the RMSF values of the 19-residue peptaibols, Brev-I and IV, in comparison to 20-residue peptaibols, Par-B and H, except that the sequences containing more Aib residues show a slight elevation in atomic fluctuation at the corresponding sequence position. For example, at R16 for Brev-I and R17 for Par-B, also, the R6 Aib in Par-B and H shows higher average atomic fluctuation than the R6 Gln of Brev-I and IV. This observation establishes the fluctuating and dynamic nature of the Aib residue in peptaibol sequences which can be explained by its tendency to oscillate between right- and left-handed helical forms.

The reweighted dihedral based PCA clusters and their representative structures were also obtained as discussed below. Amongst paracelsins, Par-B shows more dynamic behaviour than Par-H in terms of proximity of different clusters on the FEL. Most highly populated clusters, apart from 1, 2 and 3, obtained for Par-B lie on the same energy plane and are easily energetically accessible. The peptide shows dynamic interconversion between different states. The highest populated energy minimum state indicated by the representative structure 1 is closest to the native paracelsin-X structure from *T. reesei* known to us through X-ray diffraction methods (PDB ID: 4BY8). The structure is completely helical and linear. However, Clusters 2 and 3 with helical structure bent in a hairpin shape are close in population density. The $2^{nd}$ cluster appears only at the last 200 ns of simulation trajectory before reverting back to Cluster 1 in the end (**Figure 31**).

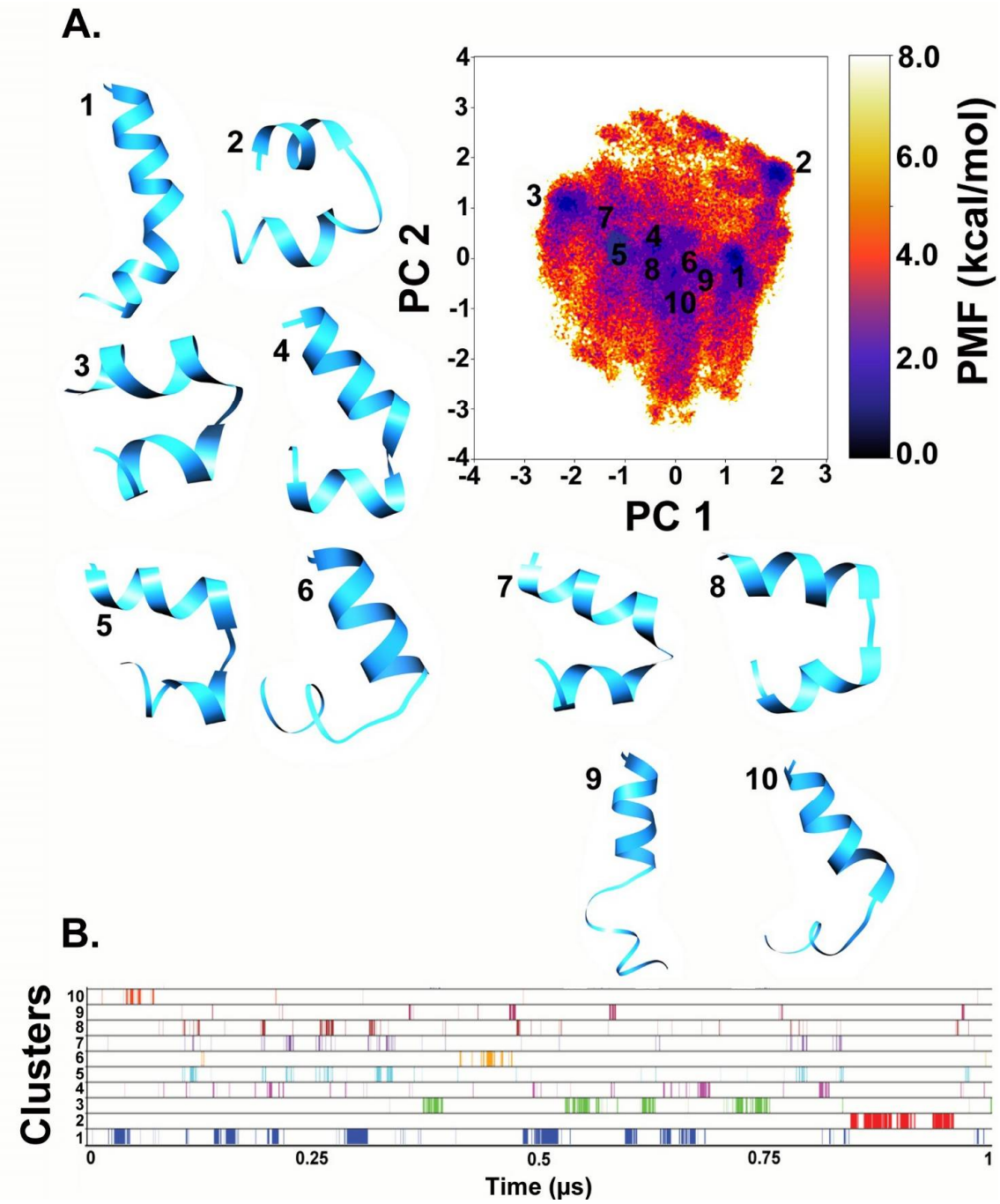


**Figure 31.** The representative structures of Par-B. **(A)** Reweighted FEL of the first two PCs calculated from dihedral angles, phi-psi, for better clustering based on internal motions. **(B)** Diagrammatic representation of cluster distribution along the simulation trajectory.

Par-H, on the other hand, can be observed to spread to physically separate regions on the FEL (**Figure 32**). All three main Clusters 1, 2 and 3 show a highly folded α-helix at the N-terminus, break at the Aib13-Pro14 bond and again α-helix at the C-terminus. The break introduces a bend in the backbone which forms the most populated energy minima cluster. The most populated Cluster 1 shows unstable helix formation at the C-terminus due to the

substitution of a Val17 instead of Aib17. Valine has a slight helix destabilizing nature owing to its branched side-chain when compared to Aib. The same effect was observed for TRK V as discussed in **Section 5.2.3.** due to the presence of isoleucine residue. The linear, fully helical structure represented by Cluster 3 and 8 is obtained later between 700-800 ns followed by Clusters 4, 5 and 7. Here, the structure corresponding to Clusters 3 and 8 is closest to the native paracelsin-X.

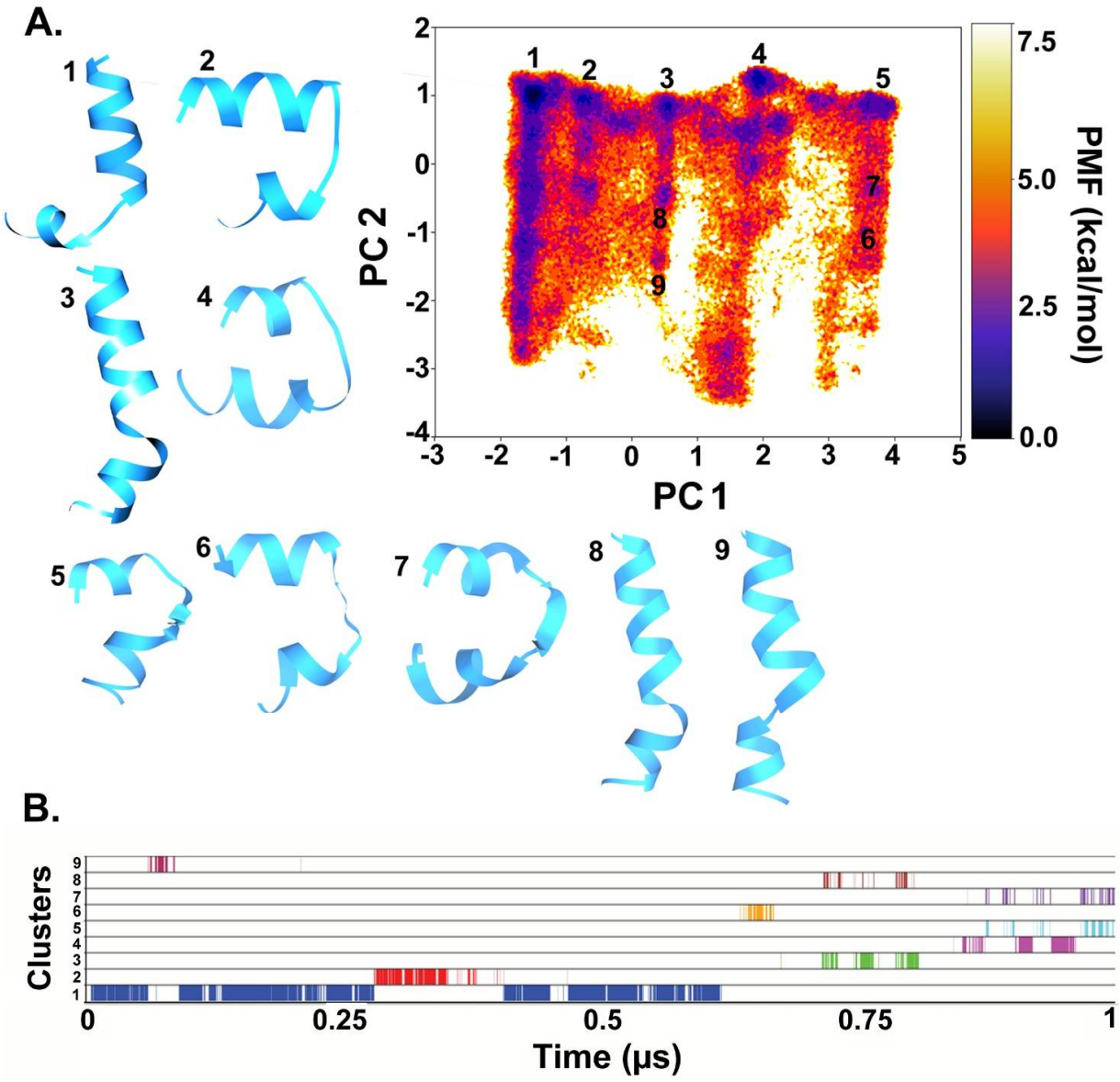


**Figure 32.** The representative structures of Par-H. **(A)** Reweighted FEL of the first two principal components calculated from dihedral angles, phi-psi, for better clustering based on internal motions. **(B)** Diagrammatic representation of cluster distribution along the simulation trajectory.

In comparison, their corresponding 19-residue peptaibols Brev-I and -IV are less dynamic in nature as observed from their FEL. Brevicelsins form clearly separate energy clusters and do not show multiple interconversions between states in comparison to behaviour of paracelsins. A highly folded, α-helical, linear structure is obtained as Cluster 1 for Brev-I (**Figure 33**). On the other hand, a similarly folded, α-helical but slightly bent structure is obtained as cluster 1 for Brev-IV (**Figure 34**). The loss of one Aib residue at R6 seems to reduce the spread of conformational phase space of brevicelsins. The same observation holds true in the case of the two paracelsins, where a valine replaces Aib at R17 in Par-H and introduces a clear segregation of clusters and reduced distribution of phase space. The energy minimum for Par-H lies at an end-to-end distance of 22 Å which is a bent conformation as can be seen in **Figure 30A**, whereas Brev-IV exhibits a slight fall at this point, even though its energy minimum also lies at 10 Å. Moreover, out of all clusters obtained in dPCA, the representative structure of the most populated cluster (Cluster 1) obtained during dihedral PCA of Brev-IV is a bent helical conformation with an end-to-end distance of 22 Å and the other one is a hairpin-like bent helical conformation represented by cluster 3 with an end-to-end distance of 10 Å. This favour for bent backbones shown by Par-H and Brev-IV can easily be attributed to the presence of a more hydrophobic valine at R17.

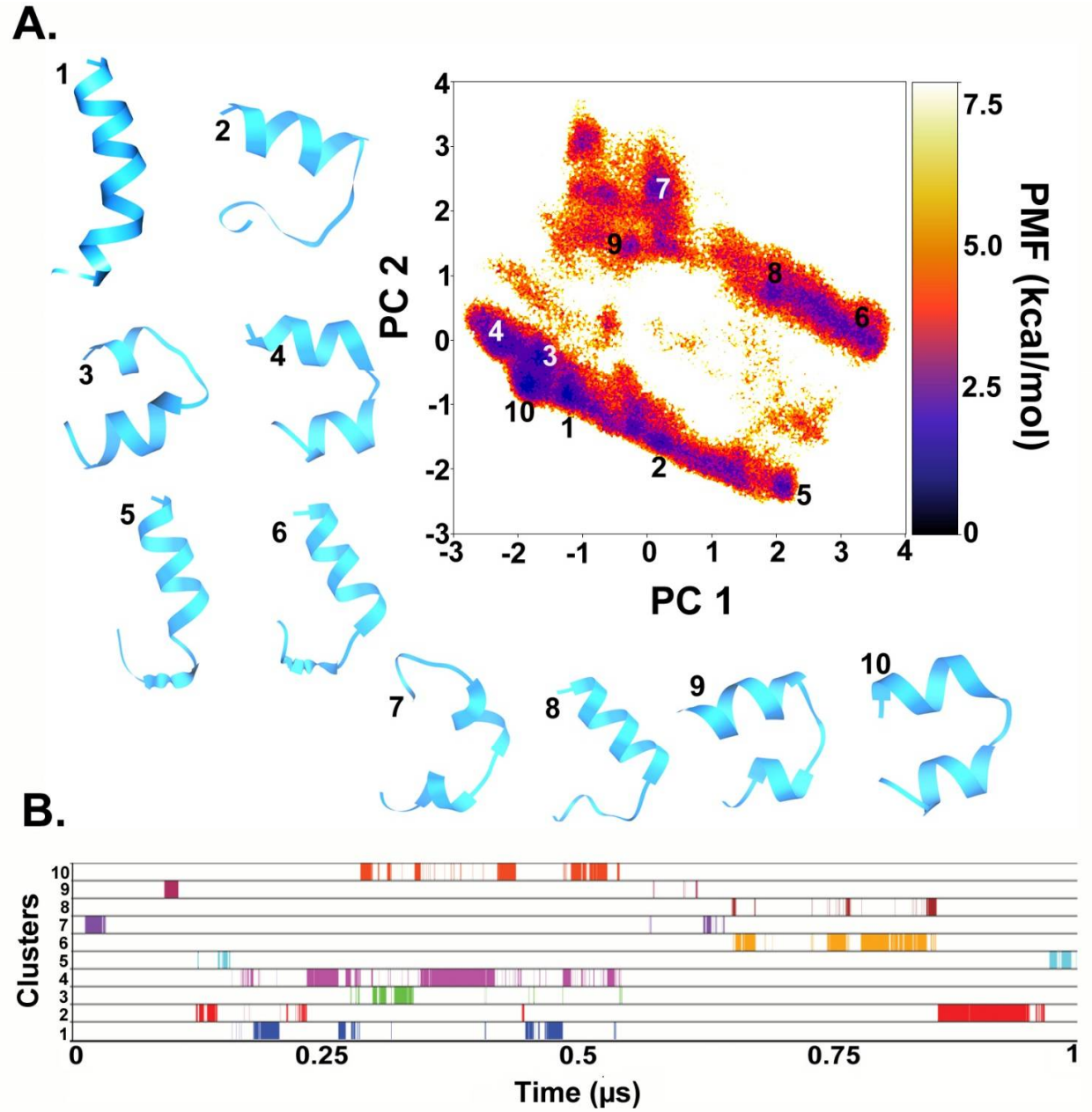


**Figure 33.** The representative structures of Brev-I. **(A)** Reweighted FEL of the first two principal components calculated from dihedral angles, phi-psi, for better clustering based on internal motions. **(B)** Diagrammatic representation of cluster distribution along the simulation trajectory.

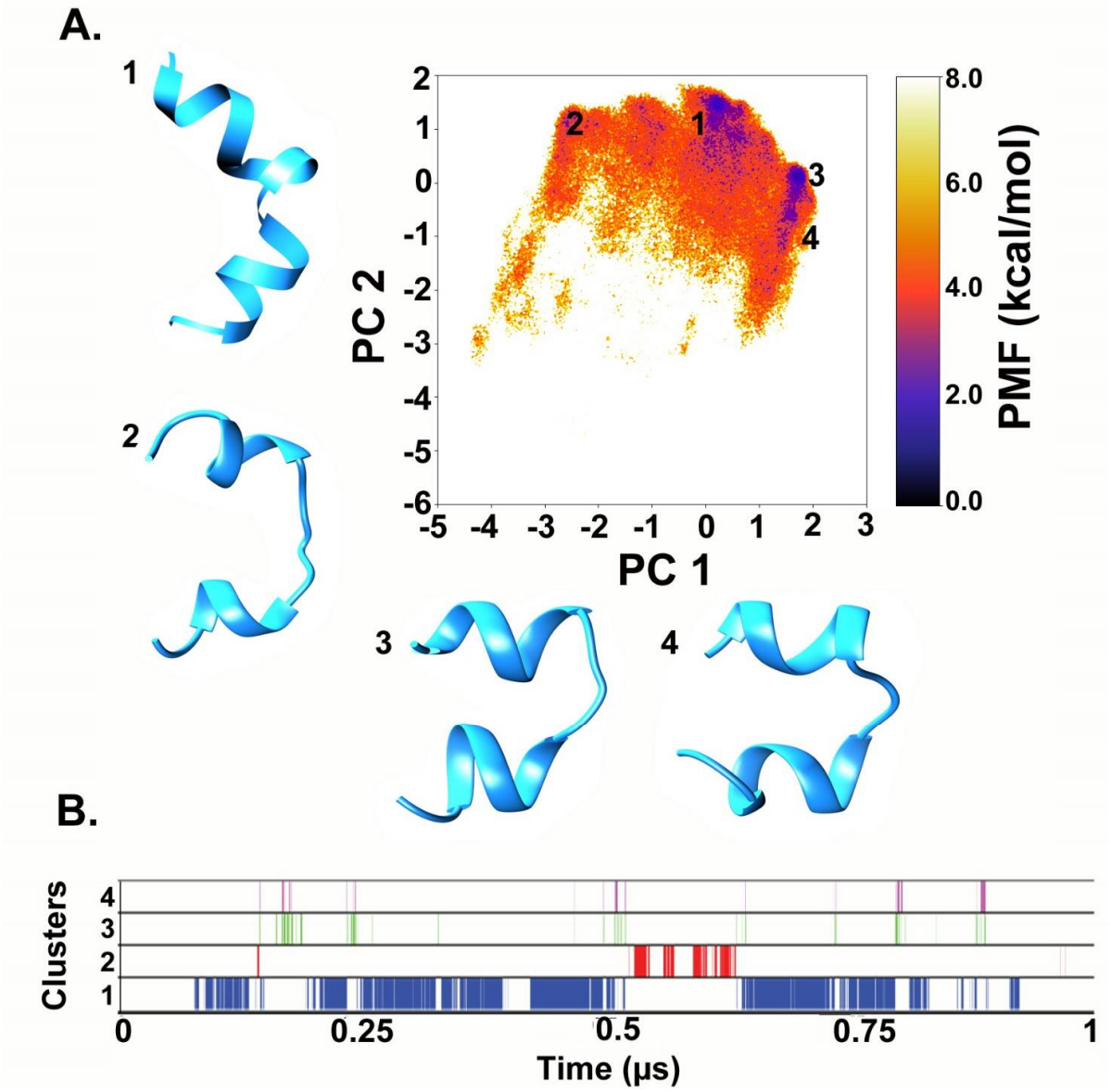


**Figure 34.** The representative structures of Brev-IV**. (A)** Reweighted FEL of the first two principal components calculated from dihedral angles, phi-psi, for better clustering based on internal motions. **(B)** Diagrammatic representation of cluster distribution along the simulation trajectory.

## 5.3. aMD simulations in basic lipid bilayer membranes: Alamethicin hexamer pore in a bacterial mimicking bilayer membrane

There have been attempts in the past to simulate the nature of ion channels, for example, Thogersen et al. (2008), carried out coarse-grained simulations of water penetration into alamethicin channel. Coarse-grained force fields are used to reduce the computational burden especially in the case of a system mimicking bilayer membrane but it comes at the cost of

losing the full atomic details by representing four atoms as a single entity (bead). In this manner, a single water bead diameter of 5.0 Å compared to 2.8 Å of an all-atom water molecule representation made it difficult to penetrate in the channel. To overcome this shortcoming, they resorted to use a short 50 ns simulation of all-atom representation of the system. This simulation resulted in the formation of a leaky bilayer where at least 95 % of water molecules within the bilayer are in direct contact with at least one alamethicin peptide. This experiment clearly outlines the importance and accuracy of all-atom simulations but the computational burden of carrying such a simulation must be solved. aMD has been tested to accurately depict ion channels formed by alamethicin hexamer.

It is a well-known fact that Alm shows multiple conductance states, which directly correlates with the number of peptide monomer units. A previous study performed to highlight the difference in channel conductance with quadromer, pentamer, hexamer, octamer and even nonamer Alm F30/3 pores highlighted the occluded state of 3-mer and 4-mer pores, while pores beyond 5-mer were comparable (Rahaman & Lazaridis, 2014). The hexamer Alm pore is the most widely accepted model of its channel formation. For this reason, we chose to simulate a hexamer Alm model in (DOPE:DOPG) 3:1 bilayer membrane (Lombardi et al., 2017). DOPE is a cationic or neutral lipid, whereas, DOPG is negatively charged. The electron density profiles for DOPE: DOPG lipids and water calculated across the bilayer normal (Z-direction) has been provided in **Figure 35A**. These were calculated by assuming an electron charge equal to the atomic number minus the atomic partial charge, located at the centre of each atom. A slight increase in water density from 0 to 10 Å in the bilayer signifies water displacement through the Alm pore. These results are in accordance with electron density profiles calculated for the same DOPE:DOPG system upon insertion of KcsA potassium channel (Schott-Verdugo & Gohlke, 2019). The lipid order parameters of the acyl chains were also determined, which can be directly compared with experimental $S_{CD}$ values. $S_{CD}$ is a measure of the relative orientation of the C-D bonds with respect to bilayer normal and can be calculated as $|S_{CD}| = 0.5 \langle 3\cos^2\theta - 1 \rangle$, where $\theta$ is the angle between bilayer normal and the vector joining $C_i$ to its deuterium atom, where $\langle\rangle$ means average of all lipid molecules. All contributions from conformational disorder, local tilting known as lipid wobble and collective motions constitute the $S_{CD}$ parameter and thus, can be a measure of membrane fluidity (Pastor et al., 1988; Pastor et al., 1991).

**Figure 35B** shows the average lipid acyl chain order parameters for mixed DOPE and DOPG. It could be compared to the plateau values of the two chains sn1 and sn2 taken from carbon number 4 to 6 for DOPE as 0.211 and 0.215, respectively (Venable et al., 2015). The plateau values for sn1 and sn2 in this case (for a DOPE:DOPG mixture) membrane is slightly lower, averaging at ~0.16 for both chains. This could be a result of DOPG mixing or the presence of membrane-perturbing peptide channel. It is evident that the membrane is more disordered than the pure DOPE membrane system. Moreover, another study conducted on *Pseudomonas aeruginosa* mimic membrane system, comprising DOPE and DOPG with a synthetic lipid, observed an average $S_{CD}$ value for pure DOPE inner membrane as 0.180 and for DOPG as 0.112 (Li et al., 2018). On the other hand, the average value for the membrane system used in this study is 0.10. This clearly shows that the membrane is highly disordered during the course of simulation due to the presence of the Alm F30/3 channel. The primary results of membrane-peptide simulations indicated that the Alm F30/3 hexamer channel increases membrane disorder, which eventually leads to leaking of water molecules and may lead to the disintegration of the bacterial cell. The lipid order parameters are in accordance with the results reported on the same system simulated using 5 replicas of 500 ns long classical MD by Schott-Verdugo & Gohlke, 2019.

The diffusion coefficient (DC) of water inside the Alm F30/3 pore was calculated using mean square displacement (MSD $Å^2$ $ps^{-1}$). It is the average distance that all water molecules travel from their starting position in XYZ direction. The MSD was reported along the z-direction. The speed of water movement can be estimated based on the rise in slope of MSD vs time plot. The diffusion constant is calculated by fitting a slope to the MSD vs time plot and multiplying it by 10.0/2 × N (where N is the number of dimensions). The diffusion constant was calculated to be $0.0311 \times 10^{-5}$ $cm^2\ s^{-1}$ and $0.0245 \times 10^{-5}$ $cm^2\ s^{-1}$ for the first and second simulation, respectively. This in comparison to the value of diffusion of water as a liquid at $0.23 \times 10^{-4}$ $cm^2\ s^{-1}$, is much lower which is expected as the behavior of water molecules in ion channel changes drastically in comparison to bulk water. Another important fact to note while calculating DC values from simulations is that most force fields, like the TIP3P water model used in this study, overestimate the diffusion coefficient even in bulk solution (Shinoda, 2016).

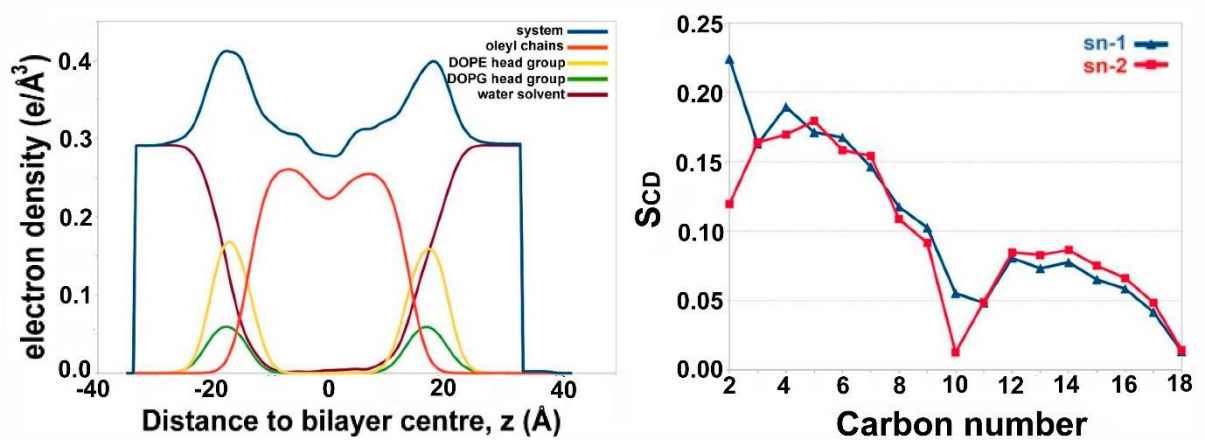


**Figure 35. (A)** The electron density profile for each constituent of membrane system and water calculated across the bilayer normal (Z-direction). A slight rise in the profile of water around 10 Å indicates presence of water in Alm F30/3 pore. Total electron density (in blue), the water density (brick red), the acyl tails of DOPE lipids (orange), the phosphoethanolamine lipid heads of DOPE (yellow) and phosphoglycerol heads of DOPG (green). **(B)** Lipid order parameters of the acyl chains, $|S_{CD}|$ can be compared with previous experimental values. The acyl chains show high disorder that may be a result of the presence of Alm F30/3 pore.

Sansom et al. (1997) showed that the dielectric constant of water in an α-helical bundle pore is significantly less than that of bulk water under an externally applied electric field which significantly affects the electrostatics of ion permeation through this channel. However, the applied electric field does not seem to be sufficient to induce the passage of $K^+$ from within the pore as observed from 300 ns of trajectory. It is apparent from DC values for the two trajectories that a higher external electric field value (0.07 V $nm^{-1}$) causes slightly higher MSD of water molecules across z direction (**Figure 36A**). The density of water molecules in the z direction is also shown in **Figure 36B** by histogramming all water O atoms on a grid with a spacing of 1 Å. The resultant file can be visualized in VMD software. The MSD analysis clearly indicated bulk movement of water molecules through the Alm F30/3 hexamer channel. **Figure 36C** is a graphical representation of number of water molecules present in a hypothetical shell. This shell was created around the peptide pore residues at a distance of 3.4 Å. Therefore, the number of water molecules present inside the pore are counted as a function of simulation time. It is clear, that from 200 to 230 water molecules are always present at any given time. A 14 μs long all-atom classical MD simulation of the Alm hexamer pore in DOPC membrane studied by Perrin Jr. & Pastor (2016) reported about 40 to 50 water molecules at any given time. The difference in this number can be attributed to multiple factors like the application of an external electric field in our case and the use of accelerated dynamics. They calculated the number of water molecules at a distance of 10 Å from the

bilayer centre forming a 20 Å region in total (to be in the pore), while our calculation includes the complete length of the Alm F30/3 pore (30 Å) and thus, the surface-bound regions and water molecules are also considered.

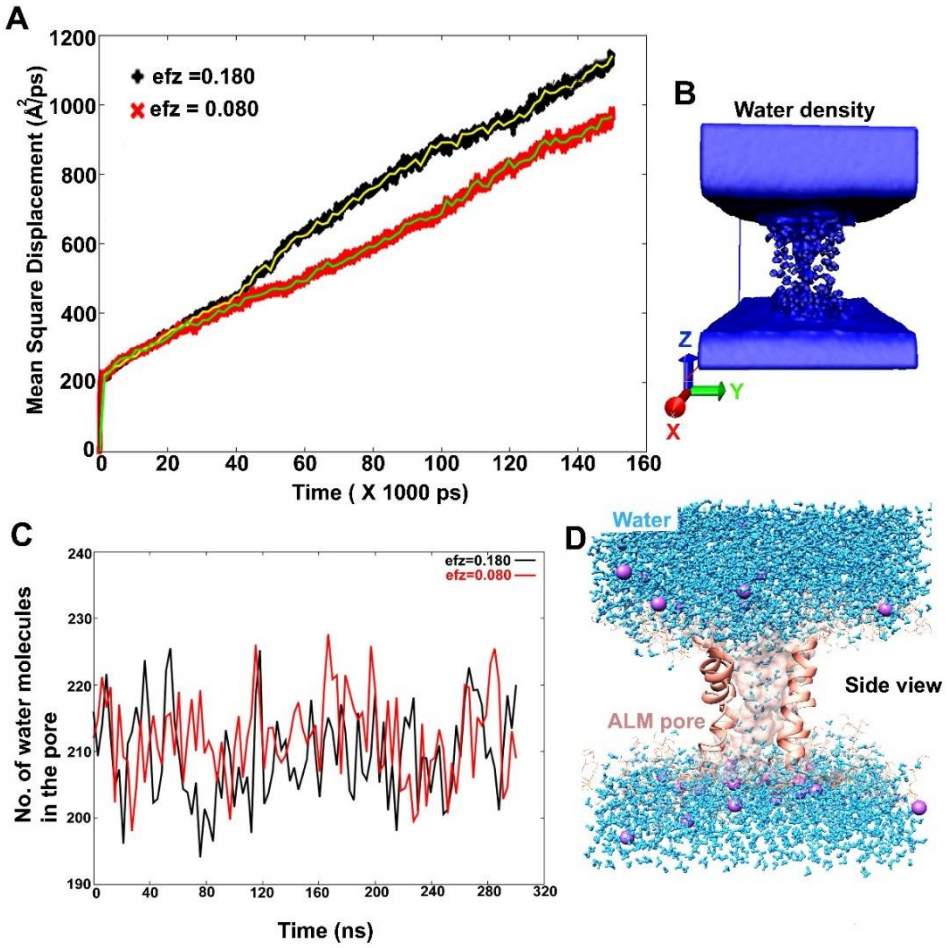


**Figure 36. (A)** The mean square displacement ($Å^2$ $ps^{-1}$) values of water calculated as a function of simulation time. Using the slope of MSD curve, the average diffusion coefficient was calculated to be $0.0311 \times 10^{-5}$ $cm^2$ $s^{-1}$ and $0.0245 \times 10^{-5}$ $cm^2$ $s^{-1}$ for Mem-sim1 and Mem-sim2, respectively. Mem-sim1 clearly shows higher displacement of water under the influence of a comparatively stronger external electric field (0.07 V $nm^{-1}$) defined by the efz value of 0.180 **(B)** A diagrammatic representation of water density around the pore. **(C)** The number of water molecules present in the Alm F30/3 pore at a given time. **(D)** A cartoon representation of water molecules passing through the Alm F30/3 pore. The front two monomers have been hidden for visual clarity.

There is not a clear difference between the number of water molecules present in the pore at a given time when compared under two different external electric fields, which means that a higher electric field does not affect the pore size so as to accommodate more water molecules, but may affect the speed of water to move through the pore as shown by varying DC values. A diagrammatic representation of the presence of water molecules in the channel

is shown in **Figure 36D**. The peptides present in the front view have been hidden for clear visualization. The size of hexamer pore must change dynamically to accommodate the influx of bulk water and thus, pore radius was calculated as function of simulation time.

The average pore radius (Å) as a function of z coordinate was calculated over each frame of the combined 600 ns trajectory (**Figure 37A**). It is clear, that the C-terminus (top of Alm F30/3 pore) undergoes stronger deviation than the N-terminus. The funnel-shaped top of the pore undergoes thickening and thinning of pore size continuously. The pore is thinnest at the center of the bilayer. Abbasi et al. (2018) visualized the formation of Alm pores in floating phospholipid membrane on gold electrodes, which confirmed the hexamer ion channel formation with the diameter of a pore calculated to be 2.3 ± 0.3 nm or 23 ± 3 Å. This value corresponds with the calculated radius from -15 to 10 Å of the z coordinate (transmembrane region). **Figure 37B** shows the top view of the pore surface surrounded by lipid heads. We were also curious to observe the changes in secondary structure of Alm F30/3 peptides in the hexamer as a function of simulation time. In lipid bilayers, the degree of helicity is dependent on the physical state of the lipid (Vogel, 1987), the lipid-peptide ratio (Cascio & Wallace, 1988) or the presence of transmembrane potentials (Brumfeld & Miller, 1990). Based on percent α-helicity calculated for each peptide monomer as shown in **Figure 37C**, it is apparent that helices 1 and 3 undergo major changes. Helices 1, 4 and 6 show an average 50% helicity while helix 2 shows the highest at 65% and helix 3 shows the lowest at 30%. It must be noted that all six peptides started with the same conformation and yet undergo vastly varied conformational changes. Perrin Jr. & Pastor (2016) also noted that the average % α-helicity drops to ~47%, which is in line with circular dichroism experiments of Alm channel within DOPC membranes. It could be a direct result of application of an external electric field. To correlate this change in percent α-helicity for each monomer, we calculated the angle (in degrees) between a vector passing through the center of mass of N-terminal residues and a vector passing through C-terminal residues (**Figure 37D**). This angle is mentioned as bend angle from here on. As expected, helix 3 that was observed to have the least percent α-helicity also has the least value of bend angle, followed by helix 5. It means that these monomers undergo extensive backbone bending during simulation, which is not feasible with a strict α-helix conformation. Therefore, these monomers unwind to achieve a more relaxed spiral conformation to accommodate pore transformation. Conversely, helices 1 and 4 show an average α-helicity around 50% and yet their bend angle values are amongst the highest. It is evident that helices 1 and 4 do not undergo drastic backbone bending but lose their helicity

indicating towards other factors affecting pore dynamics. One reason could be to accommodate the passage of bulk water molecules. Finally, the last observation regarding helix 2 showing highest α-helicity (65%) but a smaller bend angle of 80˚ indicates that a certain degree of backbone bend is possible without completely losing α-helicity. A highly bent conformation is sterically difficult to achieve with the presence of strict α-helix for the whole sequence. Therefore, few monomers show interconversion between α-helix and turns. On the other hand, few other monomers show that a certain degree of backbone bend is possible with strict α-helical conformation.

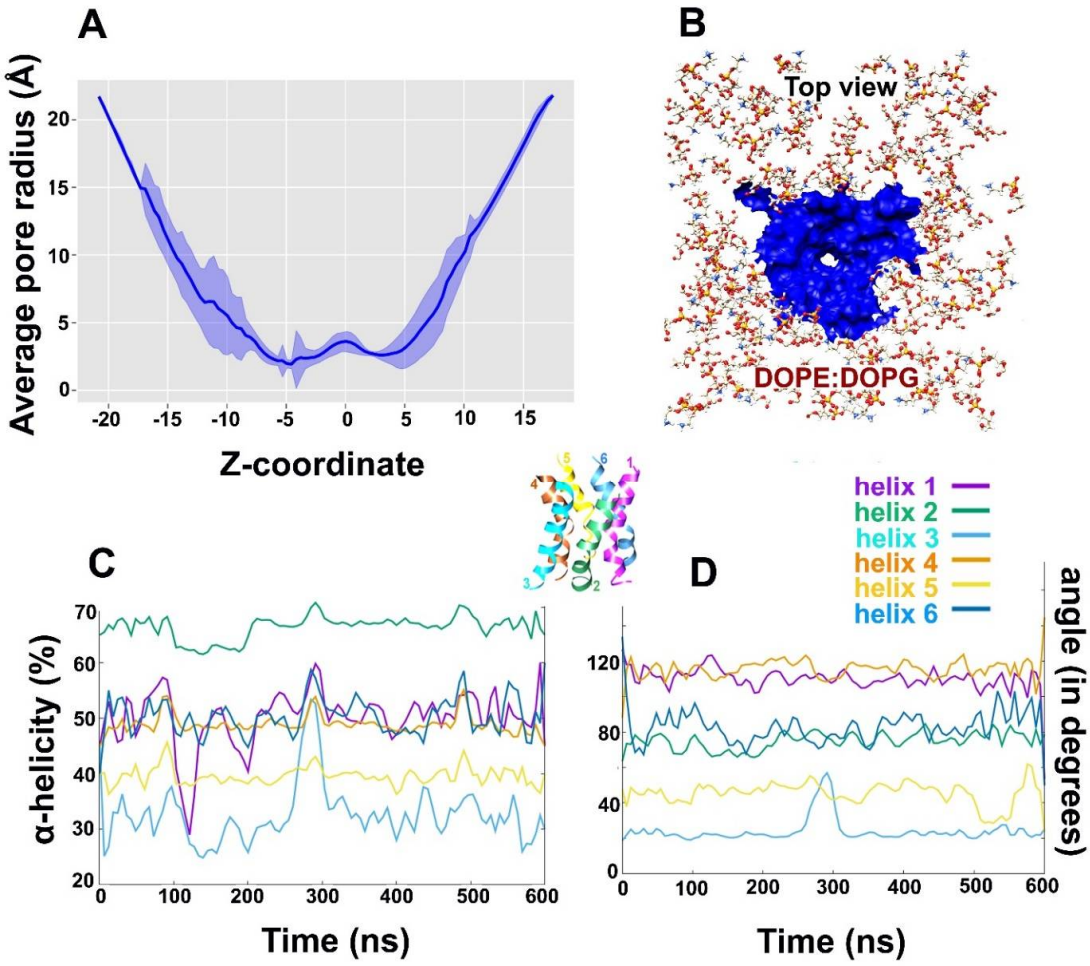


**Figure 37. (A)** The average radius of the Alm F30/3 pore (Å) calculated along transmembrane coordinate. The major fluctuation in the radius is shown by the C-terminus probably to accommodate incoming water flux. **(B)** Cartoon representation of the top view of the Alm F30/3 pore surface in blue. **(C)** Percent α-helicity calculated for each Alm F30/3 monomer from the pore with respect to simulation time (in ns). **(D)** The bend angle (calculated as dot product of two vectors passing through the N- and C-terminals of each monomer) as a function of simulation time. **Figures 37C** and **D** can be correlated assuming that the higher the bend angle, the lower will be its α-helicity. The helices are color-coded for visual interpretation.

# 6. SUMMARY

Present work was started with a simple goal to explore the three-dimensional structures of novel peptaibols using their sequences obtained from the genus *Trichoderma*. The understanding of their folding patterns and dynamics is the first step to describe their modes of action against host membranes. Earlier studies by our group employed 100 ns long simulations in implicit solvent for the lack of technical infrastructure. For statistical accuracy, 100 replica of each simulations were carried out by using a random starting conformation for each replica, a strategy based on the principle of Markov state models in which independent short simulations, with each trajectory describing a small portion of the protein's phase space, are aggregated in a statistical fashion. The results in the form of secondary structural populations and hydrogen bonds were reported as averages of 100 simulations for each peptaibol.

In this work, the initial studies carried out on TRK V and TPN XIIc involved a single 100 ns long simulation. The results obtained from TRK V were slightly surprising, as many residues like Gln6, Aib7, Aib8, Aib9, Aib12, Val14, Aib15, and Ile16 had a strong preference to form left-handed helix. It's not unexpected for the Aib residues, as their chirality allows them to adopt helical conformations in the left-handed and right-handed quadrants, but Gln and Val are associated with right-handed α-helix conformations. The remaining amino acid residues showed preference for the right-handed helix region and a few towards the β-space on the Ramachandran plot. This clash between opposing helical turns rendered the resulting three-dimensional structure as an unfolded spiral. The free energy distribution of RMSD and RoG values showed that a highly bent structure with the N- and C-termini in close proximity to each other obtained through most of the trajectory between 60-90 ns was the most energetically stable structure.

TPN XIIc, on the other hand, contains three Div residues that may have characteristic effects on the screw sense of the resulting helix, on bioactivity and even on the probability to be degraded by proteolytic enzymes. Therefore, an obvious additional interest arose to observe the folding behaviour of D-isovalines. The presence of Div residues brought in a preference for left-handed helical regions while affecting the neighboring Aib residues. The Div11-Aib12-Pro13 region resulted in unwinding of the spiral due to proline while the Val9-Ala10 region showed the backbone reversing γ'-region geometry. The result indicated the presence of β-turns for most of the sequences with a highly bent backbone. The comparative results in

water and methanol solvents showed that most residues are populated around the α- and δ-regions. Aib12, on the other hand, lies in the δ' region. Div11 showed the same behaviour in all simulations. Pro13 lies almost exclusively in $P_{II}$ conformation in both explicit solvents, water and methanol. The γ'-turn conformation is not stable in explicit solvent environment, especially in methanol. The loss of the Div11→Val9 γ'-turn bond seems to reduce the overall bending of the chain and bring linearity. Overall, we observed the formation of β-bend ribbon spirals at the N-terminal segment with higher stability in methanol. For the C-terminus, water solvent promotes the formation of α-helix while methanol promotes the formation of $3_{10}$-helix. The ever changing dynamic landscape of these peptaibols prompted to undertake enhanced sampling simulation techniques.

The newly identified TPN XIIc produced in the highest amount by the green mold *T. pleuroti* was our first target to test the enhanced simulation tehnique, viz. aMD. The first three simulations were set up to run for 500 ns consecutively, with varying boost parameters in explicit water solvent. The reweighted free energy plots of distribution of torsion angles for each residue were compared from each simulation and no change could be observed between individual simulations. In other words, each residue showed energy minimum in the exact same region of the Ramachandran plot for all three simulations and would result in a slightly bent but folded, spiral-like conformation. This observation led to running simulations in explicit chloroform solvent so as to mimic a hydrophobic environment provided by bilayer membranes. At this point, it was realized that accelerated MD algorithms have not been tested for non-aqueous solvents, thus the application of boost to the whole potential, i.e. the whole system including chloroform molecules, increases the energy of the system to unprecedented levels and thus, abruptly ends the simulation run. Three consecutive simulations for 500 ns each were set up but only with boost to dihedral energy. To compensate for lack of boost to total potential, a 1 μs long simulation was carried out in chloroform solvent with a higher dihedral boost. The dihedral PCA-based FEL plot revealed a new, highly linear conformation of TPN XIIc which had not been observed in the previous three runs in chloroform. Naturally, this led to a simulation run for 1 μs in water solvent which also resulted in a linear, spiral-shaped conformation represented as Cluster 2 and 8 which are accessible under ~2.5 kcal $mol^{-1}$ from each other. Similarly, the new highly linear conformation of TPN XIIc represented as cluster 1 in chloroform was not the energy minimum. The energy barrier between the two states (Cluster 4 → Cluster 1) is ~4 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. The earlier application of small boost energy and smaller simulation time (500 ns) must have resulted in the peptide being stuck in a single energy minimum dwell. The FEL plots have been obtained after combining all consecutive simulations of 2.5 μs. The integration of multiple aMD simulations with varying boost parameters to calculate the final free-energy distribution has been proven to be a good strategy to remove the large statistical noise while reweighting. The structure of TPN XIIc was found to be a continuous β-bend ribbon spiral with $\alpha/3_{10}$-helix at the C-terminal. The spiral shape instead of a regular $\alpha/3_{10}$-helix can be correlated with the presence of three Div residues.

Understanding the convergence through KLD method: In this theory, any two simulations, starting from different conformations, if run long enough must eventually traverse the same path so that their probability distributions overlap and the divergence between them must be the smallest. The divergence cutoff is chosen to be less than 0.025. In case of TPN XIIc, the 4 simulations were combined so as to obtain the same eigenvectors for each simulation (2.5 μs long trajectory) and then divided in 5 parts to calculate KLD values. The slopes for first four parts reduce below 0.025 within 300 ns which means that most part of the phase space is covered by all these four simulations. However, the convergence between the $4^{th}$ and $5^{th}$ part is not achieved, which can be attributed to the fact that Cluster 1 does not appear after 2 μs and the last 500 ns only present conformations in Cluster 2, 7 and 8. The KLD between parts 2 and 5 must be low as the same conformations appear during these trajectories. The KLD analysis in chloroform solvent simulation showed convergence amongst the first three simulations except the last simulation, where for the first time, a straight backbone had been observed and resulted in high values of KLD for parts 4 and 5. It became clear that a 1 μs long aMD simulation was required and found to be sufficient to map the phase space of a given peptaibol. This puts a question 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 which is clearly not true. Our rationale lies in the fact, that chloroform is a very restrictive solvent for peptide folding as it is clear from our results. The conformational landscape covered in the fourth simulation covers those metastable states which are separated by a high energy barrier and could not have been achieved during shorter aMD simulations. 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. But the combined trajectory of 2.5 μs 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 the previous methods of showing convergence, e.g. RMSD.

These observations prompted a systematic optimization of aMD simulations to be used for peptaibol structure elucidation, and Alm F30/3 was chosen for this purpose. The structure of alamethicin F30/3 has been experimentally studied and it is the most studied peptaibol. In order to optimize the process of peptaibol structure elucidation using aMD, we chose the unfolded Alm F30/3 conformation and set up three consecutive 1 μs long simulations. Due to technical disruptions, the data was obtained for ~900 ns for each simulation in explicit water solvent using small boost parameters. The idea to use any non-aqueous solvent has been dropped due to technical errors that may arise with it. The representative structures obtained from dPCA FEL plots of each simulation was superimposed with the PDB structure of alamethicin (PDB accession ID: 1amt) based on lowest RMSD values. By the third simulation, a linear, highly folded structure is obtained. It is necessary to keep in mind, that during these simulations the close-to-native conformation does appear, but not enough times to form a separate and significant energy cluster on the FEL map. It could be attributed to multiple other, more populated clusters being formed throughout the course of simulation. The reweighted phi-psi plots were observed for each residue and no significant change could be observed amongst individual simulations. The fourth simulation was carried out on Alm F30/3 for 1 μs using slighlty aggressive boost parameters and started from the original unfolded conformation which resulted in a close-to-native conformation of AlmF30/3. All evidence suggests that 1 μs long aMD simulations using GPUs are required and sufficient for folding simulations of such short peptaibols. The effect of boost applied on dihedral energy affects the folding speed of peptaibol simulations more than boost applied to total potential energy. A comparison between the reweighted dPCA FEL plots between the first 3 combined simulations and the last simulation suggests that similar conformational space could be covered with slightly aggressive boost parameters.

Consequently, the same method was used to elucidate the structures of four peptaibols, Par-B and H and Brev-I and IV, to ascertain the effect of the loss of Aib residue at the R6 position

on their folding dynamics and structure along with substitution of a valine in place of Aib at R17 position. Based on the reweighted PMF values calculated for end-to-end distance (distance in Å from the N-terminus to the C-terminus), it seems that the replacement of Aib to Val at the $17^{th}$ position in Par-H and Brev-IV lowers the energy barrier to attain linearity from a bent conformation. It's best to keep in mind at this point, that for all four peptides, the barrier between bent and linear conformations is very low in water solvent. The RMSF plots, on the other hand, did not highlight a significant difference between the atomic fluctuations of the two peptaibol groups except slightly higher fluctuation for Aib residue at its corresponding position R6 in paracesins and R16, R17 for Brev-I and Par-B, respectively.

TRK-V: The main goal of carrying out aMD simulations was to increase the sampling of the conformational space corresponding to the peptaibol folding dynamics which led to the re-evaluation of folding dynamics of TRK-V. We reported the structural characteristics of the TRK-V in **Section 5.1.1.1.** from a 100 ns long classical MD run in implicit solvent. While the phi-psi distribution of many residues in the central peptide region obtained in the classical MD run showed a high preference for left-handed helix region, and gradual unwinding of the spiral shape was observed. The same distribution from the aMD run shows that these residues, flanking the central region, show shifts from the predominant left-handed helical regions to the right-handed helical region. The comparison between the results obtained from a short classical MD in implicit solvent with 1 μs long accelerated MD in explicit water has reinforced the idea that longer time scales in explicit solution are definitely required as the results can be highly misleading in its absence.

The success of aMD in accurately predicting conformational ensemble of peptaibols without introducing unprecedented errors prompted us to carry out all-atom alamethicin channel simulations within bilayer membrane system. The presence of the peptide channel clearly introduced disorder to the membrane as shown by the lipid $S_{CD}$ parameter. All contributions from conformational disorder, local tilting known as lipid wobble and collective motions constitute the $S_{CD}$ parameter and hence, it is a good measure of membrane fluidity. The diffusion coefficient, which is the average distance that all water molecules travel while passing through the channel, was also calculated under two different values of applied external electric fields. Expectedly, the DC value is higher under a stronger electric field ($0.0311 \times 10^{-5}$ $cm^2$ $s^{-1}$ and $0.0245 \times 10^{-5}$ $cm^2$ $s^{-1}$ for the first and second simulation, respectively) but is much lower than DC of water in bulk. It is keeping in line with previous

findings, which suggest that water inside the pore differs in its dynamic properties from bulk water, the latter exhibiting decreased translational and rotational mobility (Breed et al., 1996). The calculation of average pore radius highlighted that the N- and C-termini of the pore undergo high fluctutaion and correlates with the hinge-like movement of peptaibol chains around central region. Few Alm F30/3 monomers in the hexamer channel also seemed to lose their α-helicity during the simulation and is supported by previous experiments (Perrin Jr. & Pastor, 2016).

# 7. ÖSSZEFOGLALÁS

Ez a PhD munka azzal a célkitűzéssel indult, hogy újonnan felfedezett peptaibolok háromdimenziós szerkezetét jellemezzük. A szerkezet-kialakulás módjának és dinamikájának a felderítése az első lépés a sejtmembránnal való kölcsönhatás módjának megértése felé. A peptaibolok szekvenciáját a *Trichoderma* genomjának ismeretében határoztuk meg, tekintettel a benne lévő, kódolt aminosavakkal azonos elemi összetételű nemkódolt aminosavak tömegspektrometriás azonosításának nehézségeire. Csoportunkban előzetesen 100 ns hosszú implicit vizes szimulációt alkalmaztunk a háromdimenziós szerkezet felderítésére. A statisztikai pontosság érdekében ezt 100 különböző, véletlen konformációval indulva ismételtük meg. Ez lényegében egy Markov-féle statisztikai modell volt, amiben a peptidek fázisterét rövid, egymástól független szimulációk egyesítésével jellemeztük. Ekkor a szerkezet háromdimenziós jellemzése lényegében a trajektóriumokból kapott másodlagos szerkezeti elemek és hidrogénhíd-kötések előfordulásának statisztikai gyakorisága volt.

PhD munkám során az előzetes vizsgálatok a TRK-V és TPN XIIc 100 ns-os szimulációival kezdődtek. Ezek alapján némileg meglepő volt, hogy a TRK-V-ben számos aminosav (Gln6, Aib7, Aib8, Aib9, Aib12, Val14, Aib15 és Ile16) erős hajlamot mutatott balmenetű hélix kialakítására. Ez nem meglepő az Aib esetében, mivel itt a kiralitás hiánya mindkét menetű hélix kialakulását segíti, azonban általánosságban a Gln és Val részvétele a jobbmentű α-hélixben jellemző. A többi felsorolt aminosav jelenléte szintén jobbmentű α-hélixekben, illetve a Ramachandran-térkép úgynevezett β-régiójában jellemző. Valószínűleg ez az ellentétes másodlagos szerkezetképző hajlam eredményezte a nem szabályos másodlagos szerkezeti elemként megjelenő spirális szerkezetet. A szabadenergia eloszlása az RMSD és forgási sugár függvényében azt mutatta, hogy az energetikailag legstabilabb szerkezet erősen hajlított, amiben az N- és C-terminális igen közel van egymáshoz. Ez a trajektórium 60-90 ns részén volt megfigyelhető.

A TPN XIIc esetében a benne lévő három Div aminosav jelenléte jelentős hatást gyakorolhat a csavarmenet irányára, bioaktivitására, sőt akár a proteolitikus enzimek lebontó hatására is. Sok antimikrobiális peptid nem alkalmazható terápiás céllal az emlőssejtekkel szembeni toxicitásuk, illetve a proteázokkal szembeni érzékenységük miatt. Utóbbi kiküszöbölésére a legsikeresebbnek tekinthető szerkezeti módosítások között felsorolhatók a C- és N-terminális módosítások, a ciklizáció, a D-aminosavak és nem-peptid jellegű gerincelemek beépítése és a multimerizáció (Falciani és mtsai., 2012; Kim és mtsai., 2014; Ong és mtsai., 2014; Khara és mtsai., 2014). Ezek alapján nyilvánvalónak tűnt a Div aminosav háromdimenziós szerkezet kialakulására gyakorolt hatásának vizsgálata. Több Div jelenléte megnöveli a balmenetes hélix kialakulására való hajlamot, egyúttal hatással van a szomszédos Aib részekre is. A Div11-Aib12-Pro13 régió a spirális szerkezet kitekeredését okozta, míg a Val9-Ala10 régió a γ' geometria megfordítását. Az eredmények β-kanyarok jelenlétét mutatták a szekvencia jelentős részében, erősen hajlott peptidgerinccel. A két oldószerben végzett szimulációkat összehasonlítva azt láthattuk, hogy a legtöbb aminosavrész a Ramachandran-térkép α- és δ-régiójában mutatkozott, míg az Aib12 a δ' régióban. A Div11 ugyanazt a viselkedést mutatta minden szimulációban. Pro13 szinte kizárólag a $P_{II}$ konformációban mutatkozott mindkét explicit oldószerben, vízben és metanolban. A γ' konformáció nem volt stabil explicit oldószerekben, különösképpen metanolban nem. A Div11→Val9 γ'-kanyar eltűnése csökkentette a peptidgerinc hajlását és egyenes gerincet eredményezett. Összességében béta-kanyarokból felépülő spirális szerkezetet észleltünk az N-terminális részben, mely metanolban stabilabbnak tűnt. A C-terminális régióban a víz α-helix képződését segítette, míg a metanol a $3_{10}$-helixét.

A hatékonyabb szimulációs módszerrekkel (accelerated MD, aMD) történő részletesebb vizsgálataink első célpontja az újonnan azonosított TPN XIIc volt, amit a *T. pleuroti* a legnagyobb mennyiségben termel. Három, egymást követő, explicit vízben folytatott 500 ns-os szimulációban eltérő boost paramétereket alkalmaztunk. A torziós szögek újrasúlyozott szabadenergia-eloszlását minden egyes alkotó aminosavra összehasonlítottuk a szimulációk között, ez alapján a szimulációk között nem volt észlelhető különbség, minden aminosavrész a Ramachandran-térkép ugyanazon régiójában mutatott energiaminimumot. A peptid szerkezete egy enyhén hajlott spirálként írható le. Következő lépésként a lipidmembrán hidrofób jellegét utánzó explicit kloroformban hajtottunk végre szimulációkat a vizes közeg hatásával történő összehasonlításhoz. Ekkor szembesültünk azzal, hogy az aMD-módszert nem vizsgálták nemvizes oldószerekre, ami váratlan nehézségeket okozott. A

boost teljes potenciálra való alkalmazása – beleértve a kloroform molekulákat is – olyan mértékben növelte meg a rendszer energiáját, ami lehetetlenné tette a szimulációt. Ezek alapján a következő szimulációkban a boost paramétereket csak a diéderes energiákra alkalmaztuk, azonban a teljes potenciális energia boost elmaradása miatt megnövelt értékkel. A diéderes PCA-n számolt FEL egy új, lineáris TPN XIIc konformációt is mutatott kloformban, ami az előző három futás során nem volt észlelhető. Összességében ez egy 1 μs hosszúságú szimulációt eredményezett vízben, aminek során a 2. és 8. klaszterben lineáris megjelenésű konformáció a jellemző, spirálszerű peptidgerinccel. A két klaszter között 2,5 kcal $mol^{-1}$ energikülönbség észlelhető. Hasonlóképpen, a TPN XIIc kloroformban észlelt, lineáris 1-es klasztere sem volt energiaminimumban. A két állapot közötti (4 → 1) energiagát ~ 4 kcal $mol^{-1}$ volt, vagyis a hajlított konformációból a lineárisba történő átmenet ritka eseménynek számít és hosszabb szimulációt igényel agresszív boost-paraméterekkel. A korábbi, kisebb boost-energiával végrehajtott rövidebb szimulációk során a peptid feltételezhetően egy kisebb energiájú szinten rekedt. Az egymás utáni szimulációk egyesítésével összesen 2,5 μs hosszúságú szimuláció eredményét láthatjuk a FEL ábrákon. A különböző boost-paraméterekkel végrehajtott aMD-szimulációk egyesítése újrasúlyozással jó stratégiának bizonyult a végső, realisztikus szabadenergia-eloszlás meghatározására és a nagymértékű statisztikus zaj kiszűrésére. A TPN XIIc szerkezete ezek alapján egy folytonos, spirálisan csavarodó β-redőként írható le, α- és $3_{10}$-hélix-szel a C-terminálisnál. A spirális alakzat megjelenése a hélixek helyett a három Div jelenlétével hozható összefüggésbe.

A szerkezeti konvergencia magyarázata a KLD-módszer alapján: A KLD elmélete szerint bármely két, különböző kezdeti feltétellel indított szimuláció – elegendően hosszú szimulációt alkalmazva – ugyanazt a konformációs teret járja be, így a valószínűség-eloszlásuk ugyanaz lesz, vagyis a kettő közötti eltérés (divergencia) a legkisebb. A divergencia értékét 0,025-ben állapították meg. A TPN XIIc esetében 4 szimulációt egyesítettünk, hogy azonos sajátvektorokat kapjunk minden egyes szimulációra (2,5 μs hosszúságú trajektórium), amelyek 5 részre bontásával számoltuk a KLD-értékeket. Az első négy részben az iránytangensek 300 ns-on belül csökkentek 0,025 alá, ami azt jelenti, hogy a fázistér legnagyobb részét lefedték. A 4. és 5. rész között azonban a konvergenciát nem sikerült elérni, ami annak tulajdonítható, hogy 2 μs elteltével az 1-es számú klaszter nem jelent meg, és az utolsó 500 ns csak a 2., 7. és 8. klasztereket tartalmazta. A 2. és 5. részek között a KLD szükségszerűen alacsony, mivel ezeken a trajektóriumokon belül ugyanazok a konformációk jelentek meg. A kloroformos szimulációk KLD-analízise az első három rész

konvergenciáját mutatta, ellentétben a negyedikkel, ahol először jelent meg egy egyenes gerincű konformáció, és magas KLD-értékeket mutatott a negyedik és ötödik részhez képest. Ezzel bebizonyosodott, hogy 1 μs hosszúságú szimulációra volt szükség az adott peptaibol fázisterének feltérképezéséhez.

Ezzel fölvetődik a kérdés, hogy a KLD-módszer vajon elégséges-e a konvergencia bizonyítására, mivel a negyedik szimuláció nélkül az első három teljes konvergenciát mutatott, ami egyértelműen nem igaz. A magyarázatunk erre az a tény, hogy a kloroform erősen gátolja a peptid-konformációk egymásba alakulását, ami az eredményeink alapján egyértelmű. A negyedik szimuláció konformációs tere nagyobb energiájú állapothoz tartozik, ami rövidebb szimulációk során nem tud létrejönni. Ez az 500 ns-os szimuláció elégtelenségére utal a magasabb energiájú állapotok eléréséhez kloroformban, ami a korlátozott boost következménye lehetett. Kloroformban a peptid látszólag egyetlen energiavölgyben ragadt, míg vízben nagymértékben dinamikus és viszonylag könnyen jutott túl különféle metastabil állapotokon. A kiinduló konformációktól függetlenül a rendszer egy bizonyos ponttól ugyanazt a konformációs teret kell bejárnia, amit kloroformban nem észleltünk. A 2,5 μs hosszúságú egyesített szimuláció azonban megfelelő mintavételt eredményezett. Ez egyúttal az alkalmazott statisztikai módszer hiányosságára is utal a konvergencia vizsgálatában, amennyiben a növelt hatékonyságú mintavételezést víztől különböző oldószerre alkalmazzák. Mindazonáltal a módszer megbízhatóbb a konvergencia vizsgálatára a korábbi módszerekhez (pl. RMSD) képest.

Ezen megfigyelések a gyorsított MD-szimulációk szisztematikus vizsgálatát kívánták a peptaibolok szerkezetének vizsgálatában, amihez az Alm F30/3 választottuk. Az alamethicin F30/3 volt a kísérletileg legkimerítőbb szerkezetvizsgálatoknak alávetett peptaibol. A peptaibolok aMD-módszerrel történő szerkezetvizsgálatának optimalizálásához az Alm F30/3 nyílt konformációjával három egymást követő 1 μs hosszúságú szimulációt indítottunk vízben. Technikai okok miatt mindhárom szimulációban 900 ns-os trajektóriumot tudtunk feldolgozni. Nemvizes oldószerben az előbbiekben részletezett okok miatt nem végeztünk szimulációt. A dPCA FEL-diagramok alapján választott reprezentatív szerkezeteket az alamethicin PDB-adatbázisban található szerkezetével (1amt) vetettük össze a legalacsonyabb RMSD-értékek figyelembevételével. A harmadik szimulációban egy lineáris gerincű, erősen hajlított konformációt találtunk. Meg kell azonban jegyezni, hogy ezen szimulációkkal a kísérletileg találthoz közeli (természetes) konformáció is megjelent, de

nem elegendő gyakorisággal ahhoz, hogy a FEL-diagramon szignifikánsan elkülönült energiaklaszterként jelenjen meg, ellentétben más, a szimuláció során megjelenő, nagyobb gyakoriságú klaszterekkel. A különböző szimulációkat összehasonlítva az újra súlyozott phi-psi diagramokon az egyes aminosavrészek között nem találtunk lényeges különbséget. A negyedik, 1 μs szimuláció során (az eredetileg alkalmazott nyújtott konformációból indulva) az Alm F30/3-nak a természeteshez közeli konformációja jelent meg. Ezek alapján kijelenthető, hogy a GPU-n végzett, 1 μs hosszúságú szimuláció elegendőnek tekinthető a hasonló rövid peptaibolok feltekeredésének vizsgálatára. Az első három és a negyedik szimulációt összehasonlítva az látható, hogy a konformációs tér enyhén agresszív boost-paraméterekkel bejárható.

Ezen eredmények birtokában ugyanezzel a módszerrel két peptaibolpár (paracelsin B és H, valamint brevicelsin I és IV,) összehasonlításával vizsgáltuk a tekeredés dinamikájának változását, egyrészt az Aib hiányával az R6 pozícióban, illetve valinnal helyettesítésével az R17 pozícióban. Az összehasonlítást az újrasúlyozott PMF-értékeknek a C- és N-terminális távolságának függésével követtük. Azt láthattuk, hogy az Aib17 Val-ra történő cseréje a Par-H és Brev-IV esetében csökkentette a hajlított szerkezetből a lineárisba történő átmenet energiagátját. Azt is figyelembe kell venni, hogy vízben mind a négy peptid esetében igen alacsony a hajlított és lineáris konformációk közötti átmenet energiagátja. Az RMSF görbék viszont nem mutattak ki szignifikáns különbséget a két peptaibol csoport atomi ingadozása között, kivéve a paracelsinek esetében az R6 pozícióban elhelyezkedő Aib aminosavat, valamint a Brev-I, ill. Par-B esetében az R16, ill. R17 pozíciókban elhelyezkedő szintén Aib aminosavakat, melyek kissé megnövekedett atomi fluktuációt mutattak.

TRK-V: az aMD-szimulációkkal itt a konformációs tér hatékonyabb feltérképezését tűztük ki, aminek célja a feltekeredés dinamikájának részletesebb vizsgálata volt. A TRK-V szerkezeti sajátosságait először a Section **5.1.1.1**-ben leírt módon, 100 ns, implicit vizes klasszikus MD alkalmazásával vizsgáltuk. Ebben a peptid középső részében található aminosavak erős hajlamot mutattak balmenetes hélix formálására, ami fokozatosan alakult át egy spirális alakzattá. Az aMD-szimulációkban a terminálisok felé eső aminosavak jobbmenetes hélix formálódását mutatták a kezdetben domináns balmenetes hélix átalakulásával. A két módszer összehasnlításával egyértelműen kijelenthető, hogy az implicit vizes, 100 ns klasszikus MD helyett hosszabb, 1 μs explicit vizes szimulációt kell alkalmazni, mert az implicit vizes szimulációk helytelen eredményre vezethetnek.

Alamethicin hexamer csatorna aMD-vizsgálata: az aMD-módszer eredményessége a peptaibolok konformációs tulajdonságainak vizsgálatában lehetővé tette, hogy az Alamethicin castornaképző tulajdonságait hasonló módon vizsgáljuk explicit vizes közegben. A csatorna megjelenése egyértelműen vezet a membrán szerkezetének és tulajdonságainak megváltozásához, ami a lipid $S_{CD}$ paraméterrel kimutatható. Az $S_{CD}$ paraméter megváltozása egy jó mérőszám a membrán fluiditásának követésére, aminek okozó tényezői között konformációs változások, lokális ferdeség (lipid wobble), kollektív mozgások találhatók. A víz membránon keresztül jutásának diffúziós együtthatóját (DC), ami a vízmolekulák által megtett átlagos út számításával kapható meg, két különböző külső elektromos tér hatására vizsgáltuk. Amint az várható volt, a membránban számolt DC nagyobb volt erősebb külső elektromos tér alkalmazásával ($0.0311 \times 10^{-5}$ $cm^2$ $s^{-1}$ és $0.0245 \times 10^{-5}$ $cm^2$ $s^{-1}$, az első és második szimulációval), és mindkét esetben lényegesen kisebb, mint a szabad víz esetében. Ez megfelel a korábbi tapasztalatoknak, miszerint a pórusban található víz dinamikai tulajdonságai különböznek a szabad vízétől, utóbbi nagyobb transzlációs és forgási szabadságot mutat (Breed és mtsai., 1996). A pórus átlagos átmérőjének vizsgálata azt mutatta, hogy a pórust kialakító peptaibol N- és C-terminálisa nagymértékben fluktuál, összhangban egy, a középső régió, mint csukló körüli mozgással. A szimuláció során a csatornát képző Alm F30/3 hexamer némely monomerében az α-helicitás eltűnt, összhangban korábbi kísérleti eredménnyekkel (Perrin Jr. & Pastor, 2016).

# 8. Conclusions

1. We showed that the use of enhanced sampling methods like aMD is crucial to accurately model the folding dynamics of peptaibols in comparison to classical MD techniques.
2. We show that 1 μs long aMD simulation on a GPU with slightly aggressive boost parameters results in the complete conformational ensemble of peptaibols.
3. The study of folding dynamics also highlighted the differences that can occur with a single residue substitution.
4. The presence of helix-breaking residues like glycine and proline was highlighted to allow the hinge-like backbone bending motion of peptaibols that may have functional relevance.
5. aMD can be successfully applied to model all-atom representations of bilayer membranes and their interaction with peptaibols without introducing grave errors.

# 9. ACKNOWLEDGMENTS

“Life is nothing but a movement of electrons!” - by Albert Szent-Györgyi, the Nobel Prize winner scientist from University of Szeged who discovered ascorbic acid probably describing computational biology in a nutshell.

I would like to begin with acknowledging the role of my mother, Dr. Padmawati Tyagi who instilled the love for science in me since early childhood. Her immediate presence and explanation of the simplest and the most complex of questions made it so much easier for me to remain on track to become a researcher. And my father, Basant Singh Tyagi, who despite not being a scientist, still taught me to question everything, even something as sensitive as faith in one’s own culture or religion. Their parenting made me confident and self-reliant that I was able to take up the challenge of the PhD life. My acknowledgment can not be complete without mentioning the contribution of my sister, Rohini Tyagi, in helping me overcome the hard days and to keep me grounded on the good days. A very special gratitude is reserved for my friend, mentor and husband, Tamás Marik, who walked with me through the hardest paths, who provided the scientific rigor and the company of an understanding friend, whatever the requirement at the time. A special gratitude to the bundle of happiness that is the Marik family for all their love towards me.

I am forever grateful to my supervisors, Dr. László Kredics who has been an inspiration in so many ways. His stories and lectures from “History of Microbiology”, *Trichoderma* and travels around the world made me inspired to seek the story behind science and scienists. And Dr. Ferenc Ötvös for helping me start from scratch when my PhD journey was at its lowest point. I am thankful for the patience, understanding and support that he provided to me during all these years. My special thanks to the head of department, Prof. Csaba Vágvölgyi, who has provided a very friendly and supportive environment to all of us. The guidance and familial support that I received by all my supervisors has been instrumental in my achievements. I will be forever thankful for everything they have taught me. I am also very thankful to Zoltán Násztor who taught, guided and helped me immensely during the first year.

The contribution of my colleagues and friends in the Department of Microbiology with whom I have laughed, shared and discussed my work can not be forgotten. I am thankful to Aruna Vigneshwari and Anuar Zhumakayev for endless talks and shared pizzas, Gábor Endre for multiple discussions on chemistry, to Dr. Lóránt Hatvani, Bettina Bóka-Bencsik, Mónika Vörös, Rita Büchner, Orsolya Kedves, Henrietta Allaga and Viktor Dávid Nagy for being so

kind to me and for all those flowers, cakes and mushrooms, to my deskmates Dr. Ottó Bencsik, Dr. Attila Szűcs, and Dr. Sándor Kocsubé for technical help and unwittingly teaching me the intricacies of the Hungarian language.

Lastly, the many friends I have been lucky to have, with whom I have celebrated and laughed, who have kept me going all through these years even before PhD and those who have been an inspiration, Nilopher Khan, Palak Aulakh, Ishan Bhatkoti, Ekta Gupta, Paras Gaur, Orsolya and David Li, Maitrayee Pathak and many others whom I met during my journey.

I would like to thank the Tempus Public Foundation, Hungary and Unversity Grants Commission, India to have selected me for the Stipendium Hungaricum scholarship. This work was supported by the Hungarian Government and the European Union within the frames of the Széchenyi 2020 Programme (GINOP-2.2.1-15-2016-00006 and GINOP-2.3.2-15-2016-00052).

# 10. REFERENCES


Abbasi, F., Alvarez-Malmagro, J., Su, Z., Leitch, J. J., & Lipkowski, J. (2018). Pore forming properties of alamethicin in negatively charged floating bilayer lipid membranes supported on gold electrodes. *Langmuir*, *34*(45), 13754-13765.

Altis, A., Nguyen, P. H., Hegger, R., & Stock, G. (2007). Dihedral angle principal component analysis of molecular dynamics simulations. *The Journal of Chemical Physics*, *126*(24), 244111.

Anders, R., Ohlenschläger, O., Soskic, V., Wenschuh, H., Heise, B., & Brown, L. R. (2000). The NMR solution structure of the ion channel peptaibol chrysospermin C bound to dodecylphosphocholine micelles. *European Journal of Biochemistry*, *267*(6), 1784-1794.

Argoudelis, A. D., Dietz, A., & Johnson, L. E. (1974). Zervamicins I and II, polypeptide antibiotics produced by Emericellopsis salmosynnemata. *The Journal of Antibiotics*, *27*(5), 321-328.

Asami, K., Okazaki, T., Nagai, Y., & Nagaoka, Y. (2002). Modifications of alamethicin ion channels by substitution of Glu-7 for Gln-7. *Biophysical journal*, *83*(1), 219-228.

Ayers, S., Ehrmann, B. M., Adcock, A. F., Kroll, D. J., Carcache de Blanco, E. J., Shen, Q., et al. (2012). Peptaibols from two unidentified fungi of the order Hypocreales with cytotoxic, antibiotic, and anthelmintic activities. *Journal of Peptide Science*, *18*(8), 500-510.

Ayton, G. S., & Voth, G. A. (2009). Systematic multiscale simulation of membrane protein systems. *Current Opinion in Structural Biology*, *19*(2), 138-144.

Balali-Mood, K., Bond, P. J., & Sansom, M. S. (2009). Interaction of monotopic membrane enzymes with a lipid bilayer: a coarse-grained MD simulation study. *Biochemistry*, *48*(10), 2135-2145.

Balashova, T. A., Shenkarev, Z. O., Tagaev, A. A., Ovchinnikova, T. V., Raap, J., & Arseniev, A. S. (2000). NMR structure of the channel-former zervamicin IIB in isotropic solvents. *FEBS Letters*, *466*(2-3), 333-336.

Baroncelli, R., Zapparata, A., Piaggeschi, G., Sarrocco, S., & Vannacci, G. (2016). Draft whole-genome sequence of *Trichoderma gamsii* T6085, a promising biocontrol agent of *Fusarium* head blight on wheat. *Genome Announc.*, *4*(1), e01747-15.

Benedetti, E., Bavoso, A., Di Blasio, B., Pavone, V., Pedone, C., Toniolo, C., & Bonora, G. M. (1982). Peptaibol antibiotics: a study on the helical structure of the 2-9 sequence of emerimicins III and IV. *Proceedings of the National Academy of Sciences*, *79*(24), 7951-7954.

Berendsen, H. J., Postma, J. P., van Gunsteren, W. F., & Hermans, J. (1981). Interaction models for water in relation to protein hydration. In Pullman B. (eds) *Intermolecular forces.* The Jerusalem Symposia on Quantum Chemistry and Biochemistry, 14, 331-342, Springer, Dordrecht.

Berg, A., Grigoriev, P. A., Degenkolb, T., Neuhof, T., Härtl, A., Schlegel, B., & Gräfe, U. (2003). Isolation, structure elucidation and biological activities of trichofumins A, B, C and D, new 11 and 13mer peptaibols from *Trichoderma* sp. HKI 0276. *Journal of peptide science: An Official Publication of the European Peptide Society*, *9*(11-12), 810-816.

Beutler, B. (2002). LPS in microbial pathogenesis: promise and fulfilment. *Journal of Endotoxin Research*, *8*(5), 329-335.

Bissett, J., Gams, W., Jaklitsch, W., & Samuels, G. J. (2015). Accepted *Trichoderma* names in the year 2015. *IMA Fungus*, *6*(2), 263-295.

Blin, K., Medema, M. H., Kottmann, R., Lee, S. Y., & Weber, T. (2016). The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Research*, *45*(D1):D555-D559.

Bozelli Jr, J. C., Sasahara, E. T., Pinto, M. R., Nakaie, C. R., & Schreier, S. (2012). Effect of head group and curvature on binding of the antimicrobial peptide tritrpticin to lipid membranes. *Chemistry and Physics of Lipids*, *165*(4), 365-373.

Breed, J., Sankararamakrishnan, R., Kerr, I. D., & Sansom, M. S. (1996). Molecular dynamics simulations of water within models of ion channels. *Biophysical Journal*, *70*(4), 1643-1661.

Brown, A. M., & Zondlo, N. J. (2012). 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*, *51*(25), 5041-5051.

Brückner, H., & Graf, H. (1983). Paracelsin, a peptide antibiotic containing α-aminoisobutyric acid, isolated from *Trichoderma reesei* Simmons Part A. *Experientia*, *39*(5), 528-530.

Brückner, H., Graf, H., & Bokel, M. (1984). Paracelsin; characterization by NMR spectroscopy and circular dichroism, and hemolytic properties of a peptaibol antibiotic from the cellulolytically active mold *Trichoderma reesei*. Part B. *Experientia*, *40*(11), 1189-1197.

Brumfeld, V., & Miller, I. R. (1990). Electric field dependence of alamethicin channels. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1024*(1), 49-53.

Bucci, R., Contini, A., Clerici, F., Pellegrino, S., & Gelmi, M. L. (2019). From glucose to enantiopure morpholino β-amino acid: a new tool for stabilizing γ-turns in peptides. *Organic Chemistry Frontiers*, *6*(7), 972-982.

Bucher, D., Grant, B. J., Markwick, P. R., & McCammon, J. A. (2011). Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics. *PLoS Computational Biology*, *7*(4), e1002034.

Bunkóczi, G., Schiell, M., Vértesy, L., & Sheldrick, G. M. (2003). Crystal structures of cephaibols. *Journal of Peptide Science: An Official Publication of the European Peptide Society*, *9*(11-12), 745-752.

Cascio, M., & Wallace, B. A. (1988). Conformation of alamethicin in phospholipid vesicles: implications for insertion models. *Proteins: Structure, Function, and Bioinformatics*, *4*(2), 89-98.

Case, D., Cerutti, D., Cheateham, T., Darden, T., Duke, R., Giese, T., et al. (2016). AMBER16 Package, *University of California, San Francisco, USA.*

Case, D., Ben-Shalom, I., Brozell, S., Cerutti, D., Cheatham III, T., Cruzeiro, V., et al. (2018) AMBER 18, *University of California, San Francisco, USA.*

Castrillo, M. L., Bich, G. A., Zapata, P. D., & Villalba, L. (2016). Biocontrol of *Leucoagaricus gongylophorus* of leaf-cutting ants with the mycoparasitic agent *Trichoderma koningiopsis*. *Mycosphere 7*(6): 810–819.

Chakraborty, B.N., Chakraborty, U., Sunar, K. (2020) Induced immunity developed by *Trichoderma* species in plants. In: Sharma A., Sharma P. (eds) *Trichoderma*. Rhizosphere Biology, Springer, Singapore.

Cheatham, T. E., & Kollman, P. A. (1997). Molecular dynamics simulations highlight the structural differences among DNA: DNA, RNA: RNA, and DNA: RNA hybrid duplexes. *Journal of the American Chemical Society*, *119*(21), 4805-4825.

Cheatham, T. I., Miller, J. L., Fox, T., Darden, T. A., & Kollman, P. A. (1995). Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *Journal of the American Chemical Society*, *117*(14), 4193-4194.

Chen, K., Liu, Z., & Kallenbach, N. R. (2004). The polyproline II conformation in short alanine peptides is noncooperative. *Proceedings of the National Academy of Sciences*, *101*(43), 15352-15357.

Chen, J. L., Sun, S. Z., Miao, C. P., Wu, K., Chen, Y. W., Xu, L. H. (2016). Endophytic *Trichoderma gamsii* YIM PH30019: a promising biocontrol agent with hyperosmolar, mycoparasitism, and antagonistic activities of induced volatile organic compounds on root-rot pathogenic fungi of *Panax notoginseng*. *Journal of Ginseng Research*, *40*(4), 315-324.

Chugh, J. K., Brückner, H., & Wallace, B. A. (2002). Model for a helical bundle channel based on the high-resolution crystal structure of trichotoxin_A50E. *Biochemistry, 41*(43), 12934-12941.

Cieplak, P., Cornell, W. D., Bayly, C., & Kollman, P. A. (1995). Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry*, *16*(11), 1357-1377.

Condamine, E., Rebuffat, S., Prigent, Y., Ségalas, I., Bodo, B., & Davoust, D. (1998). Three-dimensional structure of the ion-channel forming peptide trichorzianin TA VII bound to sodium dodecyl sulfate micelles. *Biopolymers: Original Research on Biomolecules*, *46*(2), 75-88.

Crisma, M., Barazza, A., Formaggio, F., Kaptein, B., Broxterman, Q. B., Kamphuis, J., & Toniolo, C. (2001). Peptaibolin: synthesis, 3D-structure, and membrane modifying properties of the natural antibiotic and selected analogues. *Tetrahedron*, *57*(14), 2813-2825.

Cutler SJ, Cutler HG (1999) *Biologically active natural products: pharmaceuticals. CRC Press*, New York, USA.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089-10092.

Dathe, M., Kaduk, C., Tachikawa, E., Melzig, M. F., Wenschuh, H., & Bienert, M. (1998). Proline at position 14 of alamethicin is essential for hemolytic activity, catecholamine secretion from chromaffin cells and enhanced metabolic activity in endothelial cells. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1370*(1), 175-183.

Degenkolb, T., & Brückner, H. (2008). Peptaibiomics: towards a myriad of bioactive peptides containing Cα-dialkylamino acids?. *Chemistry & Biodiversity*, *5*(9), 1817-1843.

Degenkolb, T., Dieckmann, R., Nielsen, K. F., Gräfenhan, T., Theis, C., Zafari, D., et al. (2008). The *Trichoderma brevicompactum* clade: a separate lineage with new species, new peptaibiotics, and mycotoxins. *Mycological Progress*, *7*(3), 177-219.

De Oliveira, C. A. F., Hamelberg, D., & McCammon, J. A. (2007). Estimating kinetic rates from accelerated molecular dynamics simulations: Alanine dipeptide in explicit solvent as a case study. *The Journal of Chemical Physics*, *127*(17), 11B605.

De Zotti, M., Biondi, B., Crisma, M., Hjørringgaard, C. U., Berg, A., Brückner, H., & Toniolo, C. (2012). Isovaline in naturally occurring peptides: A nondestructive methodology for configurational assignment. *Peptide Science*, *98*(1), 36-49.

Donohue, J. (1953). Hydrogen bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, *39*(6), 470.

Dowhan, W. (1997). Molecular basis for membrane phospholipid diversity: why are there so many lipids? *Annual Review of Biochemistry*, *66*(1), 199-232.

Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., & Shaw, D. E. (2012). Biomolecular simulation: a computational microscope for molecular biology. *Annual Review of Biophysics*, *41*, 429-452.

Drozdov, A. N., Grossfield, A., & Pappu, R. V. (2004). Role of solvent in determining conformational preferences of alanine dipeptide in water. *Journal of the American Chemical Society*, *126*(8), 2574-2581.

Du, L., Risinger, A. L., Mitchell, C. A., You, J., Stamps, B. W., Pan, N., et al. (2017). Unique amalgamation of primary and secondary structural elements transform peptaibols into potent bioactive cell-penetrating peptides. *Proceedings of the National Academy of Sciences*, *114*(43), E8957-E8966.

Duclohier, H., Alder, G. M., Bashford, C. L., Brückner, H., Chugh, J. K., & Wallace, B. A. (2004). Conductance studies on trichotoxin_A50E and implications for channel structure. *Biophysical Journal, 87*(3), 1705-1710.

Dunkley Jr, E. A., Clejan, S., Guffanti, A. A., & Krulwich, T. A. (1988). Large decreases in membrane phosphatidylethanolamine and diphosphatidylglycerol upon mutation to duramycin resistance do not change the protonophore resistance of *Bacillus subtilis*. *Biochimica et Biophysica Acta (BBA)-Biomembranes, 943*(1), 13-18.

Dupradeau, F. Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., et al. (2010). The REd. Tools: Advances in RESP and ESP charge derivation and force field library building. *Physical Chemistry Chemical Physics, 12*(28), 7821-7839.

Elad, Y., Chet, I., & Katan, J. (1980). *Trichoderma harzianum*: A biocontrol agent effective against *Sclerotium rolfsii* and *Rhizoctonia solani*. *Phytopathology, 70*(2), 119-121.

ElGamacy, M., Riss, M., Zhu, H., Truffault, V., & Coles, M. (2019). Mapping local conformational landscapes of proteins in solution. *Structure, 27*(5), 853-865.

Engelberth, J., Koch, T., Kühnemann, F., & Boland, W. (2000). Channel-forming peptaibols are potent elicitors of plant secondary metabolism and tendril coiling. *Angewandte Chemie International Edition, 39*(10), 1860-1862.

Epand, R. M., Rotem, S., Mor, A., Berno, B., & Epand, R. F. (2008). Bacterial membranes as predictors of antimicrobial potency. *Journal of the American Chemical Society, 130*(43), 14346-14352.

Epand, R. M., & Epand, R. F. (2009a). Domains in bacterial membranes and the action of antimicrobial agents. *Molecular BioSystems, 5*(6), 580-587.

Epand, R. F., Wang, G., Berno, B., & Epand, R. M. (2009b). Lipid segregation explains selective toxicity of a series of fragments derived from the human cathelicidin LL-37. *Antimicrobial Agents and Chemotherapy, 53*(9), 3705-3714.

Epand, R. M., & Epand, R. F. (2011). Bacterial membrane lipids in the action of antimicrobial agents. *Journal of Peptide Science, 17*(5), 298-305.

Epand, R. M., Walker, C., Epand, R. F., & Magarvey, N. A. (2016). Molecular mechanisms of membrane targeting antibiotics. *Biochimica et Biophysica Acta (BBA)-Biomembranes, 1858*(5), 980-987.

Escalona, Y., Garate, J. A., Araya-Secchi, R., Huynh, T., Zhou, R., & Perez-Acle, T. (2016). Exploring the membrane potential of simple dual-membrane systems as models for gap-junction channels. *Biophysical Journal, 110*(12), 2678-2688.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics, 103*(19), 8577-8593.

Falciani, C., Lozzi, L., Pollini, S., Luca, V., Carnicelli, V., Brunetti, J., et al. (2012). Isomerization of an antimicrobial peptide broadens antimicrobial spectrum to gram-positive bacterial pathogens. *PLoS One, 7*(10).

Ferrara, P., & Caflisch, A. (2000). Folding simulations of a three-stranded antiparallel β-sheet peptide. *Proceedings of the National Academy of Sciences, 97*(20), 10780-10785.

Formaggio, F., Crisma, M., Bonora, G. M., Pantano, M., Valle, G., Toniolo, C., et al. (1995). (R)-isovaline homo-peptides adopt the left-handed 3 (10)-helical structure. *Peptide Research, 8*(1), 6-15.

Fox, R. O., & Richards, F. M. (1982). A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5-Å resolution. *Nature, 300*(5890), 325-330.

Franklin, J. C., Ellena, J. F., Jayasinghe, S., Kelsh, L. P., & Cafiso, D. S. (1994). Structure of micelle-associated alamethicin from 1H NMR. Evidence for conformational heterogeneity in a voltage-gated peptide. *Biochemistry, 33*(13), 4036-4045.

Frenkel, D., & Smit, B. (2001). *Understanding molecular simulation: from algorithms to applications* (Vol. 1). *Academic Press, Elsevier*, Florida, USA.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). Gaussian, Inc., Wallingford CT. *Gaussian 09*.

Frisch, M. J. E. A., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2014). Gaussian~ 09 Revision D. 01.

Gajera, H. P., Bambharolia, R. P., Patel, S. V., Khatrani, T. J., & Goalkiya, B. A. (2012). Antagonism of *Trichoderma* spp. against *Macrophomina phaseolina*: evaluation of coiling and cell wall degrading enzymatic activities. *Journal of Plant Pathology & Microbiology, 3*(7).

Galbraith, T. P., Harris, R., Driscoll, P. C., & Wallace, B. A. (2003). Solution NMR studies of antiamoebin, a membrane channel-forming polypeptide. *Biophysical Journal, 84*(1), 185-194.

Galindo-Murillo, R., Roe, D. R., & Cheatham III, T. E. (2015). Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d (GCACGAACGAACGAACGC). *Biochimica et Biophysica Acta (BBA)-General Subjects, 1850*(5), 1041-1058.

Gessmann, R., Axford, D., Evans, G., Brückner, H., & Petratos, K. (2012a). The crystal structure of samarosporin I at atomic resolution. *Journal of Peptide Science, 18*(11), 678-684.

Gessmann, R., Axford, D., Owen, R. L., Brückner, H., & Petratos, K. (2012b). Four complete turns of a curved $3_{10}$-helix at atomic resolution: the crystal structure of the peptaibol trichovirin I-4A in a polar environment suggests a transition to α-helix for membrane function. *Acta Crystallographica Section D: Biological Crystallography, 68*(2), 109-116.

Gessmann, R., Axford, D., Brückner, H., Berg, A., & Petratos, K. (2017). A natural, single-residue substitution yields a less active peptaibiotic: the structure of bergofungin A at atomic

resolution. *Acta Crystallographica Section F: Structural Biology Communications*, *73*(2), 95-100.

Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., & Walker, R. C. (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *Journal of Chemical Theory and Computation*, *8*(5), 1542-1555.

Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seyler, S. L., et al. (2019). *MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations* (No. LA-UR-19-29136). Los Alamos National Lab.(LANL), Los Alamos, NM (USA).

Graf, J., Nguyen, P. H., Stock, G., & Schwalbe, H. (2007). Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. *Journal of the American Chemical Society*, *129*(5), 1179-1189.

Gruber, S., & Seidl-Seiboth, V. (2012). Self versus non-self: fungal cell wall degradation in *Trichoderma*. *Microbiology*, *158*(1), 26-34.

Grünberg, R., Nilges, M., & Leckner, J. (2006). Flexibility and conformational entropy in protein-protein binding. *Structure*, *14*(4), 683-693.

Guarro, J., Antolín-Ayala, M. I., Gené, J., Gutiérrez-Calzada, J., Nieves-Díez, C., & Ortoneda, M. (1999). Fatal case of *Trichoderma harzianum* infection in a renal transplant recipient. *Journal of Clinical Microbiology*, *37*(11), 3751-3755.

Haile, J. M. (1992). *Molecular dynamics simulation: elementary methods*. John Wiley & Sons, Inc..

Halgren, T. A., & Damm, W. (2001). Polarizable force fields. *Current Opinion in Structural Biology*, *11*(2), 236-242.

Hamada, D., Kuroda, Y., Tanaka, T., & Goto, Y. (1995). High helical propensity of the peptide fragments derived from β-lactoglobulin, a predominantly β-sheet protein. *Journal of molecular biology*, *254*(4), 737-746.

Hamelberg, D., Mongan, J., & McCammon, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics*, *120*(24), 11919-11929.

Hamelberg, D., Shen, T., & McCammon, J. A. (2005). Phosphorylation effects on cis/trans isomerization and the backbone conformation of serine− proline motifs: accelerated molecular dynamics analysis. *Journal of the American Chemical Society*, *127*(6), 1969-1974.

Hamelberg, D., de Oliveira, C. A. F., & McCammon, J. A. (2007). Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *The Journal of Chemical Physics*, *127*(15), 10B614.

Han, W. G., Jalkanen, K. J., Elstner, M., & Suhai, S. (1998). Theoretical study of aqueous N-Acetyl-l-alanine N ‘-Methylamide: structures and raman, VCD, and ROA Spectra. *The Journal of Physical Chemistry B*, *102*(14), 2587-2602.

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, *4*(1), 17.

Haris, P. I., & Chapman, D. (1988). Fourier transform infrared spectra of the polypeptide alamethicin and a possible structural similarity with bacteriorhodopsin. *Biochimica et Biophysica Acta - Biomembranes*, *943*(2), 375-380.

Harman, G. E., & Kubicek, C. P. (1998). *Trichoderma And Gliocladium, Volume 2: Enzymes, Biological Control and commercial applications. CRC Press.* London, UK.

Hatvani, L., Kocsubé, S., Manczinger, L., Antal, Z., Szekeres, A., Druzhinina, I. S., & Kredics, L. (2008). The green mould disease global threat to the cultivation of oyster mushroom (*Pleurotus ostreatus*): a review. In: *Science and cultivation of edible and medicinal fungi: Mushroom science XVII, proceedings of the 17th congress of the international society for mushroom science,* Ed: M. Gruening (Cape Town, RSA) Vol. 17, 485-495.

Hatvani, L., Manczinger, L., Vágvölgyi, C., & Kredics, L. (2013). 17 *Trichoderma* as a human pathogen. *Trichoderma: Biology and Applications*, 292-313.

Högel, P., Götz, A., Kuhne, F., Ebert, M., Stelzer, W., Rand, K. D., et al. (2018). Glycine perturbs local and global conformational flexibility of a transmembrane helix. *Biochemistry*, 57(8), 1326-1337.

Hollingsworth, S. A., & Karplus, P. A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular Concepts*, *1*(3-4), 271-283.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, *65*(3), 712-725.

Hu, K., Jiang, Y., Xie, Y., Liu, H., Liu, R., Zhao, Z., et al. (2015). Small-anion selective transmembrane “holes” induced by an antimicrobial peptide too short to span membranes. *The Journal of Physical Chemistry B*, *119*(27), 8553-8560.

Huelsmann, H., Heinze, S., Ritzau, M., Schlegel, B., & Graefe, U. (1998). Isolation and structure of peptaibolin, a new peptaibol from *Sepedonium* strains. *The Journal of Antibiotics*, *51*(11), 1055-1058.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, *14*(1), 33-38.

Hutchinson, E. G., & Thornton, J. M. (1994). A revised set of potentials for β-turn formation in proteins. *Protein Science*, *3*(12), 2207-2216.

Javeria, S., Kumar, A., Kharkwal, A. C., Varma, A., Srinivasa, N., & Sharma, P. (2020). Evaluation of rhizospheric *Trichoderma* species strains for producing cell wall-degrading and defense related enzymes in response to *Fusarium oxysporum* f. sp. *lentis*. *Indian Phytopathology*, 1-7.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, *79*(2), 926-935.

Kandula, D. R., Jones, E. E., Stewart, A., McLean, K. L., & Hampton, J. G. (2015). *Trichoderma* species for biocontrol of soil-borne plant pathogens of pasture species. *Biocontrol Science and Technology*, *25*(9), 1052-1069.

Kappel, K., Miao, Y., & McCammon, J. A. (2015). Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor. *Quarterly Reviews of Biophysics*, *48*(4), 479-487.

Karle, I. L., Flippen-Anderson, J. L., Agarwalla, S., & Balaram, P. (1991). Crystal structure of [Leu1] zervamicin, a membrane ion-channel peptide: implications for gating mechanisms. *Proceedings of the National Academy of Sciences*, *88*(12), 5307-5311.

Kawai, M., Omori, Y., Yamamura, H., Butsugan, Y., Taga, T., & Miwa, Y. (1993). Crystal structures of Boc-D-and L-Iva-L-Pro-OBZl: Unturned conformation of Aib-Pro sequence unaffected by replacement of Me with Et in Aib. *Biopolymers: Original Research on Biomolecules*, *33*(8), 1207-1212.

Khara, J. S., Priestman, M., Uhía, I., Hamilton, M. S., Krishnan, N., Wang, Y., et al. (2016). Unnatural amino acid analogues of membrane-active helical peptides with anti-mycobacterial activity and improved stability. *Journal of Antimicrobial Chemotherapy, 71*(8), 2181-2191.

Kim, H., Jang, J. H., Kim, S. C., & Cho, J. H. (2014). De novo generation of short antimicrobial peptides with enhanced stability and cell specificity. *Journal of Antimicrobial Chemotherapy, 69*(1), 121-132.

Kinoshita, M., Okamoto, Y., & Hirata, F. (2000). Peptide conformations in alcohol and water: Analyses by the reference interaction site model theory. *Journal of the American Chemical Society 122*(12), 2773-2779.

Koller, D., & Lohner, K. (2014). The role of spontaneous lipid curvature in the interaction of interfacially active peptides with membranes. *Biochimica et Biophysica Acta - Biomembranes*, *1838*(9), 2250-2259.

Koukos, P. I., & Glykos, N. M. (2014). On the application of Good-Turing statistics to quantify convergence of biomolecular simulations. *Journal of Chemical Information and Modeling*, *54*(1), 209-217.

Kredics, L., Garcia Jimenez, L., Naeimi, S., Czifra, D., Urbán, P., Manczinger, L., & Hatvani, L. (2010). A challenge to mushroom growers: the green mould disease of cultivated champignons. *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*, *1*, 295-305.

Kredics, L., Szekeres, A., Czifra, D., Vágvölgyi, C., & Leitgeb, B. (2013). Recent results in alamethicin research. *Chemistry & Biodiversity*, *10*(5), 744-771.

Kronen, M., Görls, H., Nguyen, H. H., Reißmann, S., Bohl, M., Sühnel, J., & Gräfe, U. (2003). Crystal structure and conformational analysis of ampullosporin A. *Journal of Peptide Science: An Official Publication of the European Peptide Society, 9*(11-12), 729-744.

Kubicek, C. P., Mikus, M., Schuster, A., Schmoll, M., & Seiboth, B. (2009). Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. *Biotechnology for Biofuels*, *2*(1), 19.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79-86.

Kushwaha, M., & Verma, A. K. (2014). Antagonistic activity of *Trichoderma* spp, (a bio-control agent) against isolated and identified plant pathogens. *International Journal of Chemical and Biological Sciences Research*, *1*(1), 1-6.

Lam, N. H., Ma, Z., & Ha, B. Y. (2014). Electrostatic modification of the lipopolysaccharide layer: competing effects of divalent cations and polycationic or polyanionic molecules. *Soft Matter*, *10*(38), 7528-7544.

Lane, T. J., Shukla, D., Beauchamp, K. A., & Pande, V. S. (2013). To milliseconds and beyond: challenges in the simulation of protein folding. *Current Opinion in Structural Biology*, *23*(1), 58-65.

Leitgeb, B., Szekeres, A., Manczinger, L., Vágvölgyi, C., & Kredics, L. (2007). The history of alamethicin: a review of the most extensively studied peptaibol. *Chemistry & Biodiversity*, *4*(6), 1027-1051.

Li, A., Schertzer, J. W., & Yong, X. (2018). Molecular dynamics modeling of *Pseudomonas aeruginosa* outer membranes. *Physical Chemistry Chemical Physics*, *20*(36), 23635-23648.

Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., & Shaw, D. E. (2012). Systematic validation of protein force fields against experimental data. *PloS one*, *7*(2), e32131.

Liu, Z., Chen, K., Ng, A., Shi, Z., Woody, R. W., & Kallenbach, N. R. (2004). Solvent dependence of $P_{II}$ conformation in model alanine peptides. *Journal of the American Chemical Society, 126*(46), 15141-15150.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). *Molecular Cell Biology. 4th edition*: *Section 15.4 Intracellular ion environment and membrane electric potential*. *WH Freeman and Company,* New York, USA.

Lombardi, L., Stellato, M. I., Oliva, R., Falanga, A., Galdiero, M., Petraccone, L., et al. (2017). Antimicrobial peptides at work: Interaction of myxinidin and its mutant WMR with lipid bilayers mimicking the *P. aeruginosa* and *E. coli* membranes. *Scientific Reports*, *7*, 44425.

Lyu, P. C., Sherman, J. C., Chen, A., & Kallenbach, N. R. (1991). Alpha-helix stabilization by natural and unnatural amino acids with alkyl side chains. *Proceedings of the National Academy of Sciences, 88*(12), 5317-5320.

Machaidze, G., Ziegler, A., & Seelig, J. (2002). Specific binding of Ro 09-0198 (cinnamycin) to phosphatidylethanolamine: a thermodynamic analysis. *Biochemistry, 41*(6), 1965-1971.

Mahalakshmi, R., & Balaram, P. (2006). Non-protein amino acids in the design of secondary structure scaffolds. In: *Protein design* Eds: Guerois, R. & Paz, M. L. *Humana Press*, New Jersey, USA, 71-94.

Mahlapuu, M., Håkansson, J., Ringstad, L., & Björn, C. (2016). Antimicrobial peptides: an emerging category of therapeutic agents. *Frontiers in Cellular and Infection Microbiology, 6*, 194.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation, 11*(8), 3696-3713.

Maisuradze, G. G., Liwo, A., & Scheraga, H. A. (2009). Principal component analysis for protein folding dynamics. *Journal of Molecular Biology, 385*(1), 312-329.

Marahiel, M. A., Stachelhaus, T., & Mootz, H. D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chemical Reviews, 97*(7), 2651-2674.

Marik, T., Urbán, P., Tyagi, C., Szekeres, A., Leitgeb, B., Vágvölgyi, M., et al. (2017). Diversity profile and dynamics of peptaibols produced by green mould *Trichoderma* species in interactions with their hosts *Agaricus bisporus* and *Pleurotus ostreatus*. *Chemistry & Biodiversity, 14*(6), e1700033.

Marik, T., Tyagi, C., Racić, G., Rakk, D., Szekeres, A., Vágvölgyi, C., & Kredics, L. (2018). New 19-residue peptaibols from *Trichoderma* clade Viride. *Microorganisms, 6*(3), 85.

Marik, T., Tyagi, C., Balázs, D., Urbán, P., Szepesi, Á., Bakacsy, L., & Druzhinina, I. S. (2019). Structural diversity and bioactivities of peptaibol compounds from the Longibrachiatum Clade of the filamentous fungal genus *Trichoderma*. *Frontiers in Microbiology, 10*, 1434

Markwick, P. R., Bouvignies, G., & Blackledge, M. (2007). Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy. *Journal of the American Chemical Society, 129*(15), 4724-4730.

Martin, D. R., & Williams, R. J. (1976). Chemical nature and sequence of alamethicin. *Biochemical Journal, 153*(2), 181.

Matsuzaki, K., Sugishita, K. I., Ishibe, N., Ueha, M., Nakata, S., Miyajima, K., & Epand, R. M. (1998). Relationship of membrane curvature to the formation of pores by magainin 2. *Biochemistry, 37*(34), 11856-11863.

McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded proteins. *Nature, 267*(5612), 585-590.

McClendon, C. L., Hua, L., Barreiro, G., & Jacobson, M. P. (2012). Comparing conformational ensembles using the Kullback–Leibler divergence expansion. *Journal of chemical theory and computation 8*(6), 2115-2126.

McMullin, D. R., Renaud, J. B., Barasubiye, T., Sumarah, M. W., & Miller, J. D. (2017). Metabolites of *Trichoderma* species isolated from damp building materials. *Canadian Journal of Microbiology*, *63*(7), 621-632.

Meyer, C. E., & Reusser, F. (1967). A polypeptide antibacterial agent isolated from *Trichoderma viride*. *Experientia*, *23*(2), 85-86.

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, *32*(10), 2319-2327.

Miller, B. R., & Gulick, A. M. (2016). Structural biology of nonribosomal peptide synthetases. In: *Nonribosomal peptide and polyketide biosynthesis* Ed: Evans, B. S. *Humana Press*, New York, USA, 3-29.

Miura, Y. (2020). NMR studies on the conformation, stability and dynamics of alamethicin in methanol. *European Biophysics Journal*, 1-12.

Mohamed-Benkada, M., François Pouchus, Y., Vérité, P., Pagniez, F., Caroff, N., & Ruiz, N. (2016). Identification and biological activities of long-chain peptaibols produced by a marine-derived strain of *Trichoderma longibrachiatum*. *Chemistry & Biodiversity*, *13*(5), 521-530.

Molnár-Gábor, E., Doczi, I., Hatvani, L., Vágvölgyi, C., & Kredics, L. (2013). Isolated sinusitis sphenoidalis caused by *Trichoderma longibrachiatum* in an immunocompetent patient with headache. *Journal of Medical Microbiology*, *62*(8), 1249-1252.

Mu, Y., Nguyen, P. H., & Stock, G. (2005). Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins: Structure, Function, and Bioinformatics, 58*(1), 45-52.

Mularski, A., Wilksch, J. J., Wang, H., Hossain, M. A., Wade, J. D., Separovic, F., et al. (2015). Atomic force microscopy reveals the mechanobiology of lytic peptide action on bacteria. *Langmuir*, *31*(22), 6164-6171.

Nagao, T., Mishima, D., Javkhlantugs, N., Wang, J., Ishioka, D., Yokota, K., et al. (2015). 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. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1848*(11), 2789-2798.

Nagaoka, Y., Iida, A., Kambara, T., Asami, K., Tachikawa, E., & Fujita, T. (1996). Role of proline residue in the channel-forming and catecholamine-releasing activities of the

peptaibol, trichosporin-B-VIa. *Biochimica et Biophysica Acta (BBA)-Biomembranes, 1283*(1), 31-36.

Nelson, J. W., & Kallenbach, N. R. (1986). Stabilization of the ribonuclease S-peptide α-helix by trifluoroethanol. *Proteins: Structure, Function, and Bioinformatics, 1*(3), 211-217.

Némethy, G., & Printz, M. P. (1972). The γ turn, a possible folded conformation of the polypeptide chain. Comparison with the β turn. *Macromolecules, 5*(6), 755-758.

North, C. L., Franklin, J. C., Bryant, R. G., & Cafiso, D. S. (1994). Molecular flexibility demonstrated by paramagnetic enhancements of nuclear relaxation. Application to alamethicin: a voltage-gated peptide channel. *Biophysical Journal, 67*(5), 1861-1866.

North, C. L., Barranger-Mathys, M., & Cafiso, D. S. (1995). Membrane orientation of the N-terminal segment of alamethicin determined by solid-state $^{15}$N NMR. *Biophysical Journal, 69*(6), 2392.

Nugent, T., & Jones, D. T. (2013). Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinformatics, 14*(1), 276.

Ong, Z. Y., Cheng, J., Huang, Y., Xu, K., Ji, Z., Fan, W., & Yang, Y. Y. (2014). Effect of stereochemistry, chain length and sequence pattern on antimicrobial properties of short synthetic β-sheet forming peptide amphiphiles. *Biomaterials, 35*(4), 1315-1325.

Olabiyi, T. I., & Ruocco, M. (2013). *In-vitro* Competition bio-assay experiment on the effect of *Trichoderma* species and some crop pathogenic fungi. *Journal of Biology, Agriculture and Healthcare, 3*(12): 115-120.

Ovchinnikova, T. V., Shenkarev, Z. O., Yakimenko, Z. A., Svishcheva, N. V., Tagaev, A. A., Skladnev, D. A., & Arseniev, A. S. (2003). Biosynthetic uniform 13C, 15N-labelling of zervamicin IIB. Complete $^{13}$C and $^{15}$N NMR assignment. *Journal of Peptide Science: An Official Publication of the European Peptide Society, 9*(11-12), 817-826.

Pappu, R. V., & Rose, G. D. (2002). A simple model for polyproline II structure in unfolded states of alanine-based peptides. *Protein Science, 11*(10), 2437-2455.

Paquet, E., & Viktor, H. L. (2015). Molecular dynamics, monte carlo simulations, and langevin dynamics: a computational review. *BioMed Research International, 2015*, 183918.

Pastor, R. W., Venable, R. M., & Karplus, M. (1991). Model for the structure of the lipid bilayer. *Proceedings of the National Academy of Sciences, 88*(3), 892-896.

Pastor, R. W., Venable, R. M., Karplus, M., & Szabo, A. (1988). A simulation-based model of NMR T 1 relaxation in lipid bilayer vesicles. *The Journal of Chemical Physics, 89*(2), 1128-1140.

Payne, J. W., Jakes, R., & Hartley, B. S. (1970). The primary structure of alamethicin. *Biochemical Journal, 117*(4), 757-766.

Perrin, B. S., Sodt, A. J., Cotten, M. L., & Pastor, R. W. (2015). The curvature induction of surface-bound antimicrobial peptides piscidin 1 and piscidin 3 varies with lipid chain length. *The Journal of Membrane Biology, 248*(3), 455-467.

Perrin Jr, B. S., & Pastor, R. W. (2016). Simulations of membrane-disrupting peptides I: alamethicin pore stability and spontaneous insertion. *Biophysical Journal, 111*(6), 1248-1257.

Petrenko, R., & Meller, J. (2001). Molecular dynamics. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester. *e LS.*

Petrenko, R, & Meller, J. (2010). Molecular Dynamics. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.

Pierce, B., Tong, W., & Weng, Z. (2005). M-ZDOCK: a grid-based approach for C n symmetric multimer docking. *Bioinformatics*, *21*(8), 1472-1478.

Pócsfalvi, G., Ritieni, A., Ferranti, P., Randazzo, G., Vékey, K., & Malorni, A. (1997). Microheterogeneity characterization of a paracelsin mixture from *Trichoderma reesei* using high-energy collision-induced dissociation tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, *11*(8), 922-930.

Pogliano, J., Pogliano, N., & Silverman, J. A. (2012). Daptomycin-mediated reorganization of membrane architecture causes mislocalization of essential cell division proteins. *Journal of Bacteriology, 194*(17), 4494-4504.

Quintas, A. (2013). What drives an amyloid protein precursor from an amyloidogenic to a native-like aggregation pathway?. *OA Biochemistry, 1*, 6.

Ramachandran, G. N. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology, 7*, 95-99.

Rahaman, A., & Lazaridis, T. (2014). A thermodynamic approach to alamethicin pore formation. *Biochimica et Biophysica Acta (BBA)-Biomembranes, 1838*(1), 98-105.

Reino, J. L., Guerrero, R. F., Hernández-Galán, R., & Collado, I. G. (2008). Secondary metabolites from species of the biocontrol agent *Trichoderma*. *Phytochemistry Reviews, 7*(1), 89-123.

Rinu, K., Sati, P., & Pandey, A. (2014). *Trichoderma gamsii* (NFCCI 2177): a newly isolated endophytic, psychrotolerant, plant growth promoting, and antagonistic fungal strain. *Journal of Basic Microbiology, 54*(5), 408-417.

Ritzau, M., Heinze, S., Dornberger, K., Berg, A., Fleck, W., Schlegel, B., et al. (1997). Ampullosporin, a new peptaibol-type antibiotic from sepedonium ampullospomm HKI-0053 with neuroleptic activity in mice. *The Journal of Antibiotics, 50*(9), 722-728.

Roe, D. R., Bergonzo, C., & Cheatham III, T. E. (2014). Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods. *The Journal of Physical Chemistry B, 118*(13), 3543-3552.

Roe, D. R., & Cheatham III, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, *9*(7), 3084-3095.

Rubio-Martinez, J., Tomas, M. S., & Perez, J. J. (2017). Effect of the solvent on the conformational behaviour of the alanine dipeptide deduced from MD simulations. *Journal of Molecular Graphics and Modelling*, *78*, 118-128.

Salnikov, E. S., Friedrich, H., Li, X., Bertani, P., Reissmann, S., Hertweck, C., et al. (2009). Structure and alignment of the membrane-associated peptaibols ampullosporin A and alamethicin by oriented $^{15}N$ and $^{31}P$ solid-state NMR spectroscopy. *Biophysical Journal*, *96*(1), 86-100.

Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., & Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *Journal of Chemical Theory and Computation*, *9*(9), 3878-3888.

Sansom, M. S., Smith, G. R., Adcock, C., & Biggin, P. C. (1997). The dielectric properties of water within model transbilayer pores. *Biophysical Journal*, *73*(5), 2404-2415.

Scheraga, H. A., Khalili, M., & Liwo, A. (2007). Protein-folding dynamics: overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.*, *58*, 57-83.

Schiell, M., Hofmann, J., Kurz, M., Schmidt, F. R., Vériest, L., Vogel, M., et al. (2001). Cephaibols, new peptaibol antibiotics with anthelmintic properties from *Acremonium tubakii* DSM 12774. *The Journal of Antibiotics*, *54*(3), 220-233.

Schott-Verdugo, S., & Gohlke, H. (2019). PACKMOL-Memgen: A simple-to-use, generalized workflow for membrane-protein–lipid-bilayer system building. *Journal of Chemical Information and Modeling*, *59*(6), 2522-2528.

Segarra, G., Casanova, E., Avilés, M., & Trillas, I. (2010). *Trichoderma asperellum* strain T34 controls *Fusarium* wilt disease in tomato plants in soilless culture through competition for iron. *Microbial ecology*, *59*(1), 141-149.

Serafeim, A. P., Salamanos, G., Patapati, K. K., & Glykos, N. M. (2016). Sensitivity of folding molecular dynamics simulations to even minor force field changes. *Journal of Chemical Information and Modeling*, *56*(10), 2035-2041.

Shao, J., Tanner, S. W., Thompson, N., & Cheatham, T. E. (2007). Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *Journal of Chemical Theory and Computation*, *3*(6), 2312-2334.

Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., & Eastwood, M. P. (2007). Anton, a special-purpose machine for molecular dynamics simulation *ACM SIGARCH Computer Architecture News 35*(2), 1-12.

Shenkarev, Z. O., Balashova, T. A., Efremov, R. G., Yakimenko, Z. A., Ovchinnikova, T. V., Raap, J., & Arseniev, A. S. (2002). Spatial structure of zervamicin IIB bound to DPC micelles: implications for voltage-gating. *Biophysical Journal*, *82*(2), 762-771.

Shenkarev, Z. O., Paramonov, A. S., Lyukmanova, E. N., Gizatullina, A. K., Zhuravleva, A. V., Tagaev, A. A., et al. (2013). Peptaibol Antiamoebin I: spatial structure, backbone dynamics, interaction with bicelles and lipid-protein nanodiscs, and pore formation in context of barrel-stave model. *Chemistry & Biodiversity, 10*(5), 838-863.

Shinoda, W. (2016). Permeability across lipid membranes. *Biochimica et Biophysica Acta (BBA)-Biomembranes, 1858*(10), 2254-2265.

Sittel, F., Jain, A., & Stock, G. (2014). Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates. *The Journal of Chemical Physics, 141*(1), 07B605_1.

Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A., & Sansom, M. S. (1996). HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *Journal of Molecular Graphics, 14*(6), 354-360.

Smit, B., & Frenkel, D. (1991). Vapor–liquid equilibria of the two-dimensional Lennard-Jones fluid (s). *The Journal of Chemical Physics, 94*(8), 5663-5668.

Snook, C. F., Woolley, G. A., Oliva, G., Pattabhi, V., Wood, S. P., Blundell, T. L., & Wallace, B. A. (1998). The structure and function of antiamoebin I, a proline-rich membrane-active polypeptide. *Structure, 6*(6), 783-792.

Snow, C. D., Sorin, E. J., Rhee, Y. M., & Pande, V. S. (2005). How well can simulation predict protein folding kinetics and thermodynamics? *Annual Reviews of Biophysics and Biomolecular Structure, 34*, 43-69.

Sörensen, M. R., & Voter, A. F. (2000). Temperature-accelerated dynamics for simulation of infrequent events. *The Journal of Chemical Physics, 112*(21), 9599-9606.

Stone, J. E., Hardy, D. J., Ufimtsev, I. S., & Schulten, K. (2010). GPU-accelerated molecular modeling coming of age. *Journal of Molecular Graphics and Modelling, 29*(2), 116-125.

Swegat, W., Schlitter, J., Krüger, P., & Wollmer, A. (2003). MD simulation of protein-ligand interaction: formation and dissociation of an insulin-phenol complex. *Biophysical Journal, 84*(3), 1493-1506.

Taylor, G., Nguyen, M. A., Koner, S., Freeman, E., Collier, C. P., & Sarles, S. A. (2019). Electrophysiological interrogation of asymmetric droplet interface bilayers reveals surface-bound alamethicin induces lipid flip-flop. *Biochimica et Biophysica Acta (BBA)-Biomembranes, 1861*(1), 335-343.

Thøgersen, L., Schiøtt, B., Vosegaard, T., Nielsen, N. C., et al. (2008). Peptide aggregation and pore formation in a lipid bilayer: a combined coarse-grained and all atom molecular dynamics study. *Biophysical Journal, 95*(9), 4337-4347.

Torrie, G. M., & Valleau, J. P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics, 23*(2), 187-199.

Tyagi, C., Bathke, J., Goyal, S., Fischer, M., Dahse, H. M., Chacko, S., et al. (2015). Targeting the intersubunit cavity of *Plasmodium falciparum* glutathione reductase by a novel natural inhibitor: Computational and experimental evidence. *The International Journal of Biochemistry & Cell Biology, 61*, 72-80.

Tyagi, C., Marik, T., Vágvölgyi, C., Kredics, L., & Ötvös, F. (2019). Accelerated molecular dynamics applied to the peptaibol folding problem. *International Journal of Molecular Sciences, 20*(17), 4268.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., et al. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry, 31*(4), 671-690.

Vanquelef, E., Simon, S., Marquant, G., Garcia, E., Klimerak, G., Delepine, J. C., et al. (2011). RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Research, 39*(suppl_2), W511-W517.

Venable, R. M., Brown, F. L., & Pastor, R. W. (2015). Mechanical properties of lipid bilayers from molecular dynamics simulation. *Chemistry and Physics of Lipids, 192*, 60-74.

Vogel, H. (1987). Comparison of the conformation and orientation of alamethicin and melittin in lipid membranes. *Biochemistry, 26*(14), 4562-4572.

Voievoda, N., Schulthess, T., Bechinger, B., & Seelig, J. (2015). Thermodynamic and biophysical analysis of the membrane-association of a histidine-rich peptide with efficient antimicrobial and transfection activities. *The Journal of Physical Chemistry B, 119*(30), 9678-9687.

Voter, A. F. (1997). Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Physical Review Letters, 78*(20), 3908.

Voter, A. F. (1998). Parallel replica method for dynamics of infrequent events. *Physical Review B, 57*(22), R13985.

Wei, X., Yang, F., & Straney, D. C. (2005). Multiple non-ribosomal peptide synthetase genes determine peptaibol synthesis in *Trichoderma virens*. *Canadian Journal of Microbiology, 51*(5), 423-429.

Wereszczynski, J., & McCammon, J. A. (2010). Using selectively applied accelerated molecular dynamics to enhance free energy calculations. *Journal of Chemical Theory and Computation, 6*(11), 3285-3292.

Whitmore, L., & Wallace, B. A. (2004). Analysis of peptaibol sequence composition: implications for *in vivo* synthesis and channel formation. *European Biophysics Journal, 33*(3), 233-237.

Yu, Y., Wang, J., Shao, Q., Shi, J., & Zhu, W. (2016). The effects of organic solvents on the folding pathway and associated thermodynamics of proteins: a microscopic view. *Scientific reports*, *6*, 19500.

Yun, B. S., Yoo, I. D., Kim, Y. H., Kim, Y. S., Lee, S. J., Kim, K. S., & Yeo, W. H. (2000). Peptaivirins A and B, two new antiviral peptaibols against TMV infection. *Tetrahedron Letters*, *41*(9), 1429-1431.

Zhang, S., Gan, Y., Xu, B., & Xue, Y. (2014). The parasitic and lethal effects of *Trichoderma longibrachiatum* against *Heterodera avenae*. *Biological Control*, *72*, 1-8.

Zhang, S., Gan, Y., & Xu, B. (2015). Biocontrol potential of a native species of *Trichoderma longibrachiatum* against *Meloidogyne incognita*. *Applied Soil Ecology*, *94*, 21-29.

Zhang, Y. B., & Zhuang, W. Y. (2018). New species of *Trichoderma* in the Harzianum, Longibrachiatum and Viride clades. *Phytotaxa*, *379*(2), 131-142.

Zhang, C., Yang, M., Ericsson, A. C. (2019). Antimicrobial peptides: potential application in liver cancer. *Frontiers in microbiology*, *10*, 1257.

Zocher, R., & Keller, U. (1996). Thiol template peptide synthesis systems in bacteria and fungi. In: *Advances in microbial physiology* Ed: Poole, R. K. *Academic Press, Elsevier*, Florida, USA, Vol. 38, 85-131.

# 11. APPENDIX

**Detailed description for residue parameterization**

The non-standard residues, aminoisobutyric acid (Aib) and D-isovaline (Div) were parameterized as part of one calculation along with 4 standard residues, alanine, serine, glycine and valine. The parameter file, .frcmod, is calculated collectively for all these residues and is provided below. To create their amber libraries in .off format, the respective structures of these residues were loaded using a leap script provided by R.E.D server. The reason for collective parameterization was to be able to compare the charges for standard residues with preexisting libraries of AmberTools16.

**Note**: The mol2 files contain information of Aib and Div as residue units with free ends that can be readily used in building a peptide sequence. They were parameterized as ACE-AIB/DIV-NME units.

**Using leap,**

```
loadAmberParams frcmod.known                    #parameter file for all 6 residues
AIB = loadmol3 m3-c1_f2.mol2                    #from R.E.D charge calculation results
DIV = loadmol3 m6-c1_f2.mol2


saveoff AIB AIB.off                             #saving amber libraries
saveoff DIV DIV.off
```

**The .mol2 files, for AIB**

```
@<TRIPOS>MOLECULE
F08
    13    12     1     0     1
SMALL
USER_CHARGES
@<TRIPOS>ATOM
   1 N1     1.179404   0.787225  -0.289608 N    1   F08 -0.4552  0.0000
****
   2 H7     1.401639   1.218298  -1.161329 H    1   F08  0.2870  0.0000
****
   3 C1    -0.246026   0.712307  -0.009072 CT   1   F08  0.1590  0.0000
****
   4 C2    -0.513722   0.943419   1.490893 CT   1   F08 -0.1445  0.0000
****
   5 C3    -0.874452   1.859278  -0.822371 CT   1   F08 -0.1445  0.0000
****
   6 H8    -0.683793   1.725172  -1.881967 HC   1   F08  0.0464  0.0000
****
   7 H9    -1.944751   1.917349  -0.681560 HC   1   F08  0.0464  0.0000
****
   8 H10   -0.438130   2.800958  -0.509348 HC   1   F08  0.0464  0.0000
```

```
****
   9 C6    -0.815324  -0.626254  -0.549026 C    1   F08  0.5633  0.0000
****
  10 O     -0.117607  -1.349894  -1.206062 O    1   F08 -0.5435  0.0000
****
  11 H1    -1.529198   1.266287   1.670775 HC   1   F08  0.0464  0.0000
****
  12 H2     0.148540   1.726800   1.836879 HC   1   F08  0.0464  0.0000
****
  13 H     -0.336510   0.060129   2.090571 HC   1   F08  0.0464  0.0000
****
@<TRIPOS>BOND
    1     1     2 1
    2     1     3 1
    3     3     4 1
    4     3     5 1
    5     3     9 1
    6     4    11 1
    7     4    12 1
    8     4    13 1
    9     5     6 1
   10     5     7 1
   11     5     8 1
   12     9    10 1
@<TRIPOS>SUBSTRUCTURE
      1 F08               1 ****               0 ****  ****
@<TRIPOS>HEADTAIL
N1 1
C6 1
@<TRIPOS>RESIDUECONNECT
1 N1 C6 0 0 0 0
```

**D-isovaline (DIV)**

```
@<TRIPOS>MOLECULE
F17
    16    15     1     0     1
SMALL
USER_CHARGES
@<TRIPOS>ATOM
   1 N1     1.069412   0.434947   0.415318 N    1   F17 -0.3647  0.0000
****
   2 H8     0.918993   1.367037   0.740563 H    1   F17  0.2358  0.0000
****
   3 C1    -0.156337  -0.267531   0.058905 CT   1   F17  0.0552  0.0000
****
   4 C2    -0.255076  -1.597290   0.828207 CT   1   F17 -0.2071  0.0000
****
   5 C3    -0.280048  -0.476971  -1.472639 CT   1   F17 -0.0427  0.0000
****
   6 C4    -0.175361   0.801565  -2.304979 CT   1   F17 -0.0917  0.0000
****
   7 H9    -0.971825   1.500344  -2.073312 HC   1   F17  0.0282  0.0000
****
   8 H10   -0.246482   0.559673  -3.360441 HC   1   F17  0.0282  0.0000
****
   9 H11    0.768238   1.306343  -2.139664 HC   1   F17  0.0282  0.0000
****
  10 C7    -1.286111   0.698418   0.513586 C    1   F17  0.6153  0.0000
```

```
****
  11 O     -1.017880   1.799366   0.916720 O    1   F17 -0.5641  0.0000
****
  12 H1     0.491864  -1.170168  -1.784452 HC   1   F17  0.0407  0.0000
****
  13 H2    -1.219741  -0.969252  -1.691480 HC   1   F17  0.0407  0.0000
****
  14 H3    -1.133983  -2.164910   0.562186 HC   1   F17  0.0660  0.0000
****
  15 H4    -0.269747  -1.411989   1.896661 HC   1   F17  0.0660  0.0000
****
  16 H      0.598634  -2.222187   0.612382 HC   1   F17  0.0660  0.0000
****
@<TRIPOS>BOND
    1     1     2 1
    2     1     3 1
    3     3     4 1
    4     3     5 1
    5     3    10 1
    6     4    14 1
    7     4    15 1
    8     4    16 1
    9     5     6 1
   10     5    12 1
   11     5    13 1
   12     6     7 1
   13     6     8 1
   14     6     9 1
   15    10    11 1
@<TRIPOS>SUBSTRUCTURE
      1 F17               1 ****               0 ****  ****
@<TRIPOS>HEADTAIL
N1 1
C7 1
@<TRIPOS>RESIDUECONNECT
1 N1 C7 0 0 0 0
```

**The parameter file used for both residues**

```
FRCMOD file generated by PyRED version SEP-2015 - q4md-forcefieldtools.org
MASS      mass              pol            Source
C         12.010            0.616          taken from parm10.dat
CT        12.010            0.878          taken from parm10.dat
CX        12.010            0.360          taken from parm10.dat
H          1.008            0.161          taken from parm10.dat
H1         1.008            0.135          taken from parm10.dat
HC         1.008            0.135          taken from parm10.dat
HO         1.008            0.135          taken from parm10.dat
N         14.010            0.530          taken from parm10.dat
O         16.000            0.434          taken from parm10.dat
OH        16.000            0.465          taken from parm10.dat

BOND   K(kcal.mol-1.ang-2)  Dist0(ang)     Source
C -CT     315.0             1.522          adapted from parm10.dat 317.0
C -CX     315.0             1.522          adapted from parm10.dat 317.0
C -N      490.0             1.335          taken from parm10.dat
C -O      570.0             1.229          taken from parm10.dat
CT-CT     310.0             1.526          taken from parm10.dat
CT-CX     310.0             1.526          taken from parm10.dat
CT-H1     340.0             1.090          taken from parm10.dat
CT-HC     340.0             1.090          taken from parm10.dat
```

```
CT-N      335.0             1.449          adapted from parm10.dat 337.0
CT-OH     320.0             1.410          taken from parm10.dat
CX-H1     340.0             1.090          taken from parm10.dat
CX-N      335.0             1.449          adapted from parm10.dat 337.0
H -N      435.0             1.010          adapted from parm10.dat 434.0
HO-OH     555.0             0.960          adapted from parm10.dat 553.0

ANGLE  K(kcal.mol-1.rad-2)  Theta0(deg)    Source
CT-C -N        70.0          116.60         taken from parm10.dat
CT-C -O        80.0          120.40         taken from parm10.dat
CX-C -N        70.0          116.60         taken from parm10.dat
CX-C -O        80.0          120.40         taken from parm10.dat
N -C -O        80.0          122.90         taken from parm10.dat
C -CT-CT       65.0          111.10         adapted from parm10.dat 63.0
C -CT-HC       50.0          109.50         taken from parm10.dat
C -CT-N        65.0          110.10         adapted from parm10.dat 63.0
CT-CT-CT       40.0          109.50         taken from parm10.dat
CT-CT-CX       40.0          109.50         taken from parm10.dat
CT-CT-HC       50.0          109.50         taken from parm10.dat
CT-CT-N        80.0          109.70         taken from parm10.dat
CX-CT-H1       50.0          109.50         taken from parm10.dat
CX-CT-HC       50.0          109.50         taken from parm10.dat
CX-CT-OH       50.0          109.50         taken from parm10.dat
H1-CT-H1       35.0          109.50         taken from parm10.dat
H1-CT-N        50.0          109.50         taken from parm10.dat
H1-CT-OH       50.0          109.50         taken from parm10.dat
HC-CT-HC       35.0          109.50         taken from parm10.dat
C -CX-CT       65.0          111.10         adapted from parm10.dat 63.0
C -CX-H1       50.0          109.50         taken from parm10.dat
C -CX-N        65.0          110.10         adapted from parm10.dat 63.0
CT-CX-H1       50.0          109.50         taken from parm10.dat
CT-CX-N        80.0          109.70         taken from parm10.dat
H1-CX-H1       35.0          109.50         taken from parm10.dat
H1-CX-N        50.0          109.50         taken from parm10.dat
C -N -CT       50.0          121.90         taken from parm10.dat
C -N -CX       50.0          121.90         taken from parm10.dat
C -N -H        50.0          120.00         taken from parm10.dat
CT-N -H        50.0          118.04         taken from parm10.dat
CX-N -H        50.0          118.04         taken from parm10.dat
CT-OH-HO       55.0          108.50         taken from parm10.dat

DIHEDRAL    Path V(kcal.mol-1.rad-1) Phase(deg.) Period Source
N -C -CT-CT   1    0.00000000e+00        0.0      -4.   taken from
parm10.dat
N -C -CT-CT   1    4.00000000e-01        0.0      -3.   taken from
parm10.dat
N -C -CT-CT   1    2.00000000e-01        0.0      -2.   taken from
parm10.dat
N -C -CT-CT   1    2.00000000e-01        0.0       1.   taken from
parm10.dat
N -C -CT-HC   1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-C-CT-X   0.0/6
N -C -CT-N    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-C-CT-X   0.0/6
O -C -CT-CT   1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-C-CT-X   0.0/6
O -C -CT-HC   1    8.00000000e-01        0.0      -1.   taken from
parm10.dat
O -C -CT-HC   1    0.00000000e+00        0.0      -2.   taken from
parm10.dat
O -C -CT-HC   1    8.00000000e-02      180.0       3.   taken from
```

```
parm10.dat
O -C -CT-N    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-C-CT-X   0.0/6
N -C -CX-CT   1    0.00000000e+00        0.0      -4.   taken from
parm10.dat
N -C -CX-CT   1    4.00000000e-01        0.0      -3.   taken from
parm10.dat
N -C -CX-CT   1    2.00000000e-01        0.0      -2.   taken from
parm10.dat
N -C -CX-CT   1    2.00000000e-01        0.0       1.   taken from
parm10.dat
N -C -CX-H1   1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-C-CX-X   0.0/6
N -C -CX-N    1    0.00000000e+00        0.0      -4.   taken from
parm10.dat
N -C -CX-N    1    5.50000000e-01      180.0      -3.   taken from
parm10.dat
N -C -CX-N    1    1.58000000e+00      180.0      -2.   taken from
parm10.dat
N -C -CX-N    1    4.50000000e-01      180.0       1.   taken from
parm10.dat
O -C -CX-CT   1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-C-CX-X   0.0/6
O -C -CX-H1   1    8.00000000e-01        0.0      -1.   taken from
parm10.dat
O -C -CX-H1   1    0.00000000e+00        0.0      -2.   taken from
parm10.dat
O -C -CX-H1   1    8.00000000e-02      180.0       3.   taken from
parm10.dat
O -C -CX-N    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-C-CX-X   0.0/6
CT-C -N -CT   1    2.50000000e+00      180.0       2.   adapted from
parm10.dat i.e X-C-N-X   10.0/4
CT-C -N -CX   1    2.50000000e+00      180.0       2.   adapted from
parm10.dat i.e X-C-N-X   10.0/4
CT-C -N -H    1    2.50000000e+00      180.0       2.   adapted from
parm10.dat i.e X-C-N-X   10.0/4
CX-C -N -CT   1    2.50000000e+00      180.0       2.   adapted from
parm10.dat i.e X-C-N-X   10.0/4
CX-C -N -H    1    2.50000000e+00      180.0       2.   adapted from
parm10.dat i.e X-C-N-X   10.0/4
O -C -N -CT   1    2.50000000e+00      180.0       2.   adapted from
parm10.dat i.e X-C-N-X   10.0/4
O -C -N -CX   1    2.50000000e+00      180.0       2.   adapted from
parm10.dat i.e X-C-N-X   10.0/4
O -C -N -H    1    2.50000000e+00      180.0      -2.   taken from
parm10.dat
O -C -N -H    1    2.00000000e+00        0.0       1.   taken from
parm10.dat
C -CT-CT-CT   1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CT-X   1.4/9
C -CT-CT-HC   1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CT-X   1.4/9
CT-CT-CT-CT   1    1.80000000e-01        0.0      -3.   taken from
parm10.dat
CT-CT-CT-CT   1    2.50000000e-01      180.0      -2.   taken from
parm10.dat
CT-CT-CT-CT   1    2.00000000e-01      180.0       1.   taken from
parm10.dat
CT-CT-CT-HC   1    1.60000000e-01        0.0       3.   taken from
parm10.dat
```

```
CT-CT-CT-N    1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CT-X   1.4/9
CX-CT-CT-HC   1    1.60000000e-01        0.0       3.   taken from
parm10.dat
HC-CT-CT-HC   1    1.50000000e-01        0.0       3.   taken from
parm10.dat
HC-CT-CT-N    1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CT-X   1.4/9
CT-CT-CX-C    1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
CT-CT-CX-H1   1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
CT-CT-CX-N    1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
H1-CT-CX-C    1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
H1-CT-CX-H1   1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
H1-CT-CX-N    1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
HC-CT-CX-C    1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
HC-CT-CX-H1   1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
HC-CT-CX-N    1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
OH-CT-CX-C    1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
OH-CT-CX-H1   1    0.00000000e+00        0.0      -3.   taken from
parm10.dat
OH-CT-CX-H1   1    2.50000000e-01        0.0       1.   taken from
parm10.dat
OH-CT-CX-N    1    1.55555556e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-CX-X   1.4/9
C -CT-N -C    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-CT-N-X   0.0/6
C -CT-N -H    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-CT-N-X   0.0/6
CT-CT-N -C    1    0.00000000e+00        0.0      -4.   taken from
parm10.dat
CT-CT-N -C    1    4.00000000e-01        0.0      -3.   taken from
parm10.dat
CT-CT-N -C    1    2.00000000e+00        0.0      -2.   taken from
parm10.dat
CT-CT-N -C    1    2.00000000e+00        0.0       1.   taken from
parm10.dat
CT-CT-N -H    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-CT-N-X   0.0/6
H1-CT-N -C    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-CT-N-X   0.0/6
H1-CT-N -H    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-CT-N-X   0.0/6
CX-CT-OH-HO   1    1.60000000e-01        0.0      -3.   taken from
parm10.dat
CX-CT-OH-HO   1    2.50000000e-01        0.0       1.   taken from
parm10.dat
H1-CT-OH-HO   1    1.66666667e-01        0.0       3.   adapted from
parm10.dat i.e X-CT-OH-X   0.5/3
C -CX-N -C    1    0.00000000e+00        0.0      -4.   taken from
parm10.dat
C -CX-N -C    1    4.20000000e-01        0.0      -3.   taken from
```

```
parm10.dat
C -CX-N -C    1    2.70000000e-01        0.0      -2.   taken from
parm10.dat
C -CX-N -C    1    0.00000000e+00        0.0       1.   taken from
parm10.dat
C -CX-N -H    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-CX-N-X   0.0/6
CT-CX-N -C    1    0.00000000e+00        0.0      -4.   taken from
parm10.dat
CT-CX-N -C    1    4.00000000e-01        0.0      -3.   taken from
parm10.dat
CT-CX-N -C    1    2.00000000e+00        0.0      -2.   taken from
parm10.dat
CT-CX-N -C    1    2.00000000e+00        0.0       1.   taken from
parm10.dat
CT-CX-N -H    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-CX-N-X   0.0/6
H1-CX-N -C    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-CX-N-X   0.0/6
H1-CX-N -H    1    0.00000000e+00        0.0       2.   adapted from
parm10.dat i.e X-CX-N-X   0.0/6

IMPROPER         V(kcal.mol-1.rad-1) Phase(deg.) Period Source
CT-N -C -O         1.05000000e+01      180.0       2.   adapted from
parm10.dat i.e X-X-C-O
CX-N -C -O         1.05000000e+01      180.0       2.   adapted from
parm10.dat i.e X-X-C-O
C -CT-N -H         1.10000000e+00      180.0       2.   taken from
parm10.dat
C -CX-N -H         1.10000000e+00      180.0       2.   taken from
parm10.dat

NONBON     R*(ang)   Eps(kcal.mol-1)       Source
C          1.9080    0.08600000            taken from parm10.dat
CT         1.9080    0.10940000            taken from parm10.dat
CX         1.9080    0.10940000            taken from parm10.dat
H          0.6000    0.01570000            taken from parm10.dat
H1         1.3870    0.01570000            taken from parm10.dat
HC         1.4870    0.01570000            taken from parm10.dat
HO         0.0000    0.00000000            taken from parm10.dat
N          1.8240    0.17000000            taken from parm10.dat
O          1.6612    0.21000000            taken from parm10.dat
OH         1.7210    0.21040000            taken from parm10.dat
```

**Pol/Pheol (Phenylalaninol) parameter file**

```
MASS
N  14.010        0.530              same as n
H  1.008         0.161              same as hn
CT 12.010        0.878              same as c3
H1 1.008         0.135              same as hc
HC 1.008         0.135              same as hc
CA 12.010        0.360              same as c2
HA 1.008         0.135              same as hc
OH 16.000        0.465              same as oh
HO 1.008         0.135              same as ho

BOND
N -H   410.20   1.009       same as hn-n
N -CT  330.60   1.460       same as c3-n
```

```
CT-H1  337.30   1.092       same as c3-hc
CT-CT  303.10   1.535       same as c3-c3
CT-HC  337.30   1.092       same as c3-hc
CT-CA  328.30   1.508       same as c2-c3
CA-CA  478.40   1.387       same as ca-ca
CA-HA  344.30   1.087       same as c2-hc
CT-OH  314.10   1.426       same as c3-oh
OH-HO  369.60   0.974       same as ho-oh

ANGLE
N -CT-H1   49.800     109.500   same as hc-c3-n
N -CT-CT   65.900     112.130   same as c3-c3-n
H -N -CT   46.000     116.780   same as c3-n -hn
CT-CT-HC   46.400     110.050   same as c3-c3-hc
CT-CT-CA   63.700     110.960   same as c2-c3-c3
CT-CT-H1   46.400     110.050   same as c3-c3-hc
CT-CT-OH   67.700     109.430   same as c3-c3-oh
CT-CT-CT   63.200     110.630   same as c3-c3-c3
CT-CA-CA   64.300     123.420   same as c2-c2-c3
HC-CT-HC   39.400     108.350   same as hc-c3-hc
HC-CT-CA   47.000     110.490   same as c2-c3-hc
CA-CA-HA   50.300     119.700   same as c2-c2-hc
CA-CA-CA   67.200     119.970   same as ca-ca-ca
CT-OH-HO   47.100     108.160   same as c3-oh-ho
H1-CT-H1   39.400     108.350   same as hc-c3-hc
H1-CT-OH   51.100     109.500   same as hc-c3-oh

DIHEDRAL
N -CT-CT-HC   1    0.156         0.000          3.000      same as X -c3-
c3-X
N -CT-CT-CA   1    0.156         0.000          3.000      same as X -c3-
c3-X
N -CT-CT-H1   1    0.156         0.000          3.000      same as X -c3-
c3-X
N -CT-CT-OH   1    0.156         0.000          3.000      same as X -c3-
c3-X
H -N -CT-H1   1    0.000         0.000          2.000      same as X -c3-n
-X
H -N -CT-CT   1    0.000         0.000          2.000      same as X -c3-n
-X
CT-CT-CA-CA   1    0.000         0.000          2.000      same as X -c2-
c3-X
CT-CT-OH-HO   1    0.167         0.000          3.000      same as X -c3-
oh-X
H1-CT-CT-HC   1    0.156         0.000          3.000      same as X -c3-
c3-X
H1-CT-CT-CA   1    0.156         0.000          3.000      same as X -c3-
c3-X
H1-CT-CT-H1   1    0.156         0.000          3.000      same as X -c3-
c3-X
H1-CT-CT-OH   1    0.156         0.000          3.000      same as X -c3-
c3-X
CT-CT-CT-H1   1    0.156         0.000          3.000      same as X -c3-
c3-X
CT-CT-CT-OH   1    0.156         0.000          3.000      same as X -c3-
c3-X
CT-CA-CA-HA   1    6.650       180.000          2.000      same as X -c2-
c2-X
CT-CA-CA-CA   1    6.650       180.000          2.000      same as X -c2-
c2-X
```

```
HC-CT-CT-CT   1    0.156         0.000          3.000      same as X -c3-
c3-X
HC-CT-CA-CA   1    0.000         0.000          2.000      same as X -c2-
c3-X
CA-CT-CT-CT   1    0.156         0.000          3.000      same as X -c3-
c3-X
CA-CA-CA-HA   1    6.650       180.000          2.000      same as X -c2-
c2-X
CA-CA-CA-CA   1    3.625       180.000          2.000      same as X -ca-
ca-X
HA-CA-CA-HA   1    6.650       180.000          2.000      same as X -c2-
c2-X
H1-CT-OH-HO   1    0.167         0.000          3.000      same as X -c3-
oh-X

IMPROPER
CA-CA-CA-CT         1.1         180.0         2.0
CA-CA-CA-HA         1.1         180.0         2.0

NONBON
  N           1.8240  0.1700
  H           0.6000  0.0157
  CT          1.9080  0.1094
  H1          1.4870  0.0157
  HC          1.4870  0.0157
  CA          1.9080  0.0860
  HA          1.4870  0.0157
  OH          1.7210  0.2104
  HO          0.0000  0.0000
```

## Leuol (Leucinol) parameter file

```
remark goes here
MASS
N  14.010        0.530              same as n
H  1.008         0.161              same as hn
CT 12.010        0.878              same as c3
H1 1.008         0.135              same as hc
HC 1.008         0.135              same as hc
OH 16.000        0.465              same as oh
HO 1.008         0.135              same as ho

BOND
N -H   410.20   1.009       same as hn-n
N -CT  330.60   1.460       same as c3-n
CT-H1  337.30   1.092       same as c3-hc
CT-CT  303.10   1.535       same as c3-c3
CT-HC  337.30   1.092       same as c3-hc
CT-OH  314.10   1.426       same as c3-oh
OH-HO  369.60   0.974       same as ho-oh

ANGLE
N -CT-H1   49.800     109.500   same as hc-c3-n
N -CT-CT   65.900     112.130   same as c3-c3-n
H -N -CT   46.000     116.780   same as c3-n -hn
CT-CT-HC   46.400     110.050   same as c3-c3-hc
CT-CT-CT   63.200     110.630   same as c3-c3-c3
CT-CT-H1   46.400     110.050   same as c3-c3-hc
CT-CT-OH   67.700     109.430   same as c3-c3-oh
HC-CT-HC   39.400     108.350   same as hc-c3-hc
```

```
CT-OH-HO   47.100     108.160   same as c3-oh-ho
H1-CT-H1   39.400     108.350   same as hc-c3-hc
H1-CT-OH   51.100     109.500   same as hc-c3-oh

DIHE
N -CT-CT-HC   1    0.156         0.000          3.000      same as X -c3-
c3-X
N -CT-CT-CT   1    0.156         0.000          3.000      same as X -c3-
c3-X
N -CT-CT-H1   1    0.156         0.000          3.000      same as X -c3-
c3-X
N -CT-CT-OH   1    0.156         0.000          3.000      same as X -c3-
c3-X
H -N -CT-H1   1    0.000         0.000          2.000      same as X -c3-n
-X
H -N -CT-CT   1    0.000         0.000          2.000      same as X -c3-n
-X
CT-CT-CT-HC   1    0.156         0.000          3.000      same as X -c3-
c3-X
CT-CT-CT-CT   1    0.156         0.000          3.000      same as X -c3-
c3-X
CT-CT-OH-HO   1    0.167         0.000          3.000      same as X -c3-
oh-X
H1-CT-CT-HC   1    0.156         0.000          3.000      same as X -c3-
c3-X
H1-CT-CT-CT   1    0.156         0.000          3.000      same as X -c3-
c3-X
H1-CT-CT-H1   1    0.156         0.000          3.000      same as X -c3-
c3-X
H1-CT-CT-OH   1    0.156         0.000          3.000      same as X -c3-
c3-X
CT-CT-CT-OH   1    0.156         0.000          3.000      same as X -c3-
c3-X
HC-CT-CT-HC   1    0.156         0.000          3.000      same as X -c3-
c3-X
H1-CT-OH-HO   1    0.167         0.000          3.000      same as X -c3-
oh-X

IMPROPER

NONBON
  N           1.8240  0.1700            same as n
  H           0.6000  0.0157            same as hn
  CT          1.9080  0.1094            same as c3
  H1          1.4870  0.0157            same as hc
  HC          1.4870  0.0157            same as hc
  OH          1.7210  0.2104            same as oh
  HO          0.0000  0.0000            same as ho
```

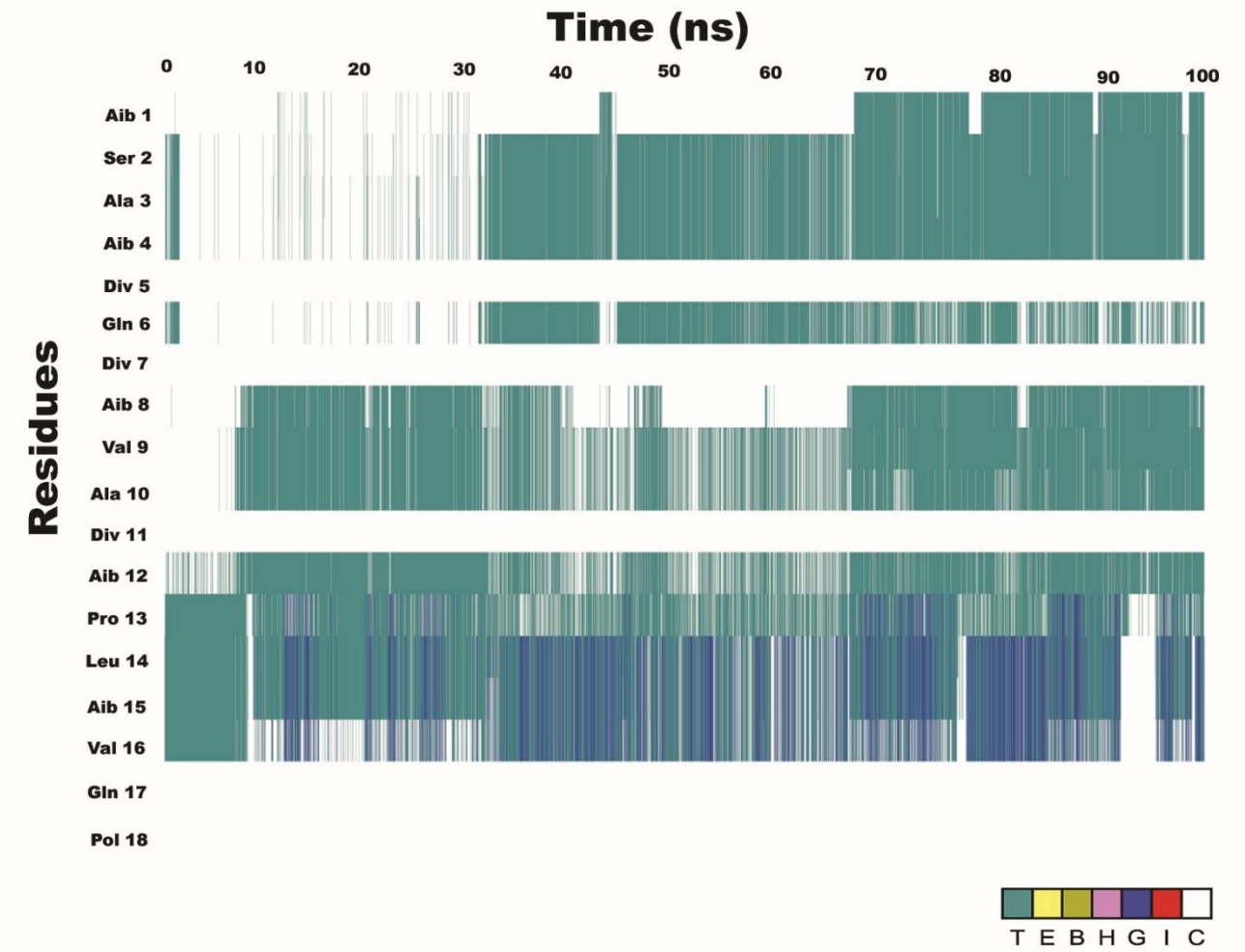


**Figure S1.** Timeline secondary structure calculation of TPN XIIc in implicit solvent for 100 ns. T=turn, E=beta extended, B=beta bridge, H=α-helix, G=$3_{10}$ helix, I=pi helix, C=coil.

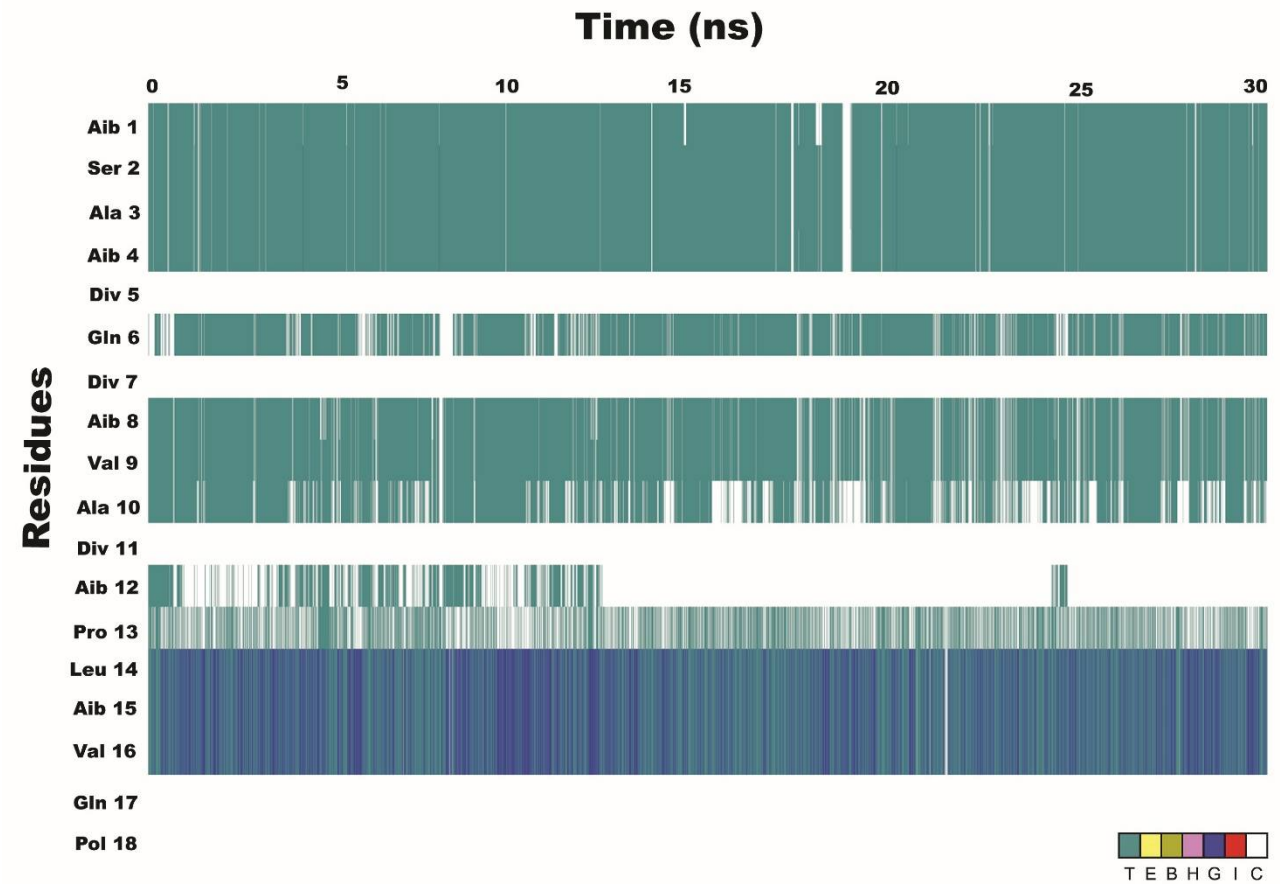


**Figure S2.** Timeline secondary structure calculation of TPN XIIc in explicit water solvent (TIP3P) for 30 ns.

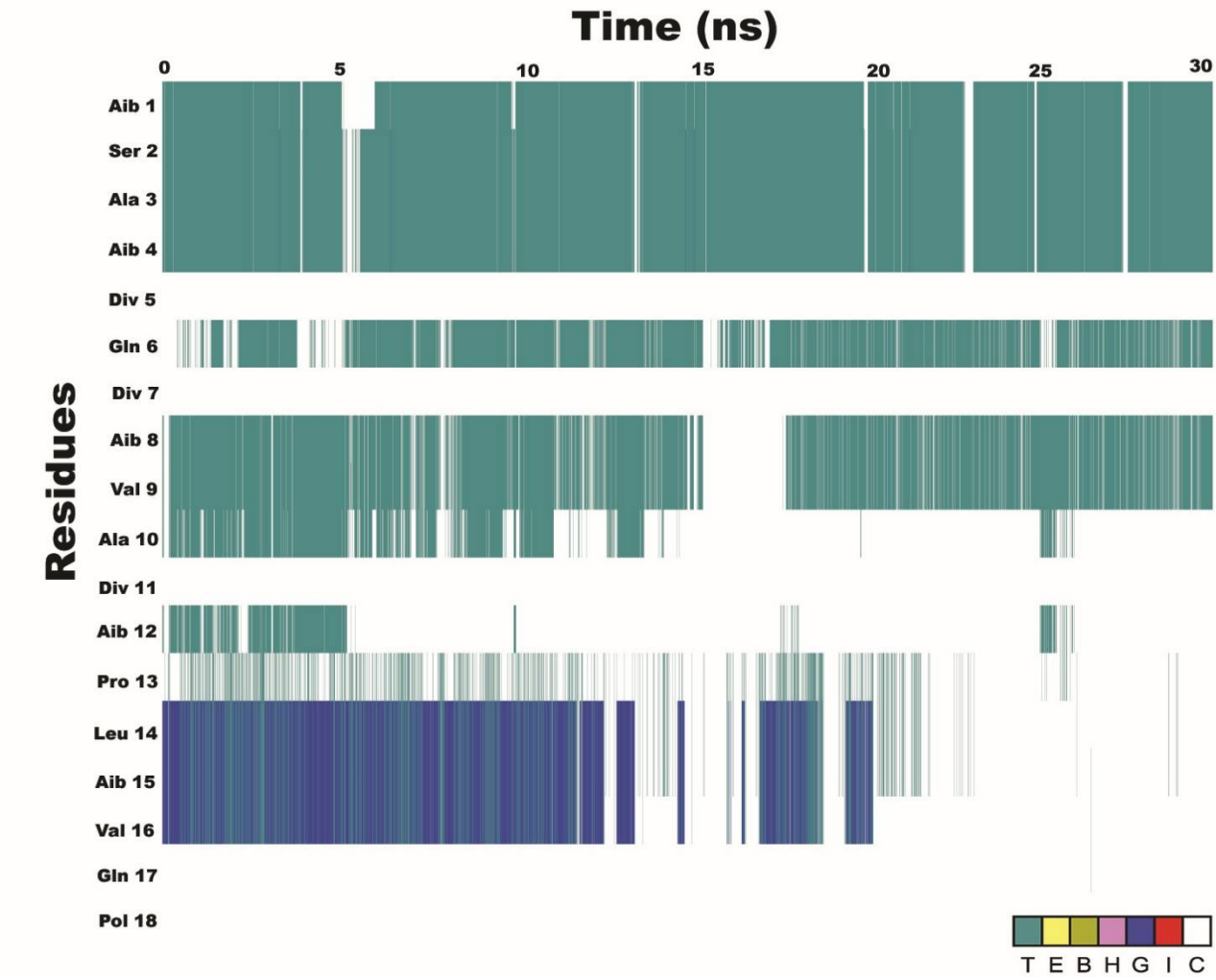


**Figure S3.** Timeline secondary structure calculation of TPN XIIc in explicit methanol solvent (MEOH) for 30 ns.

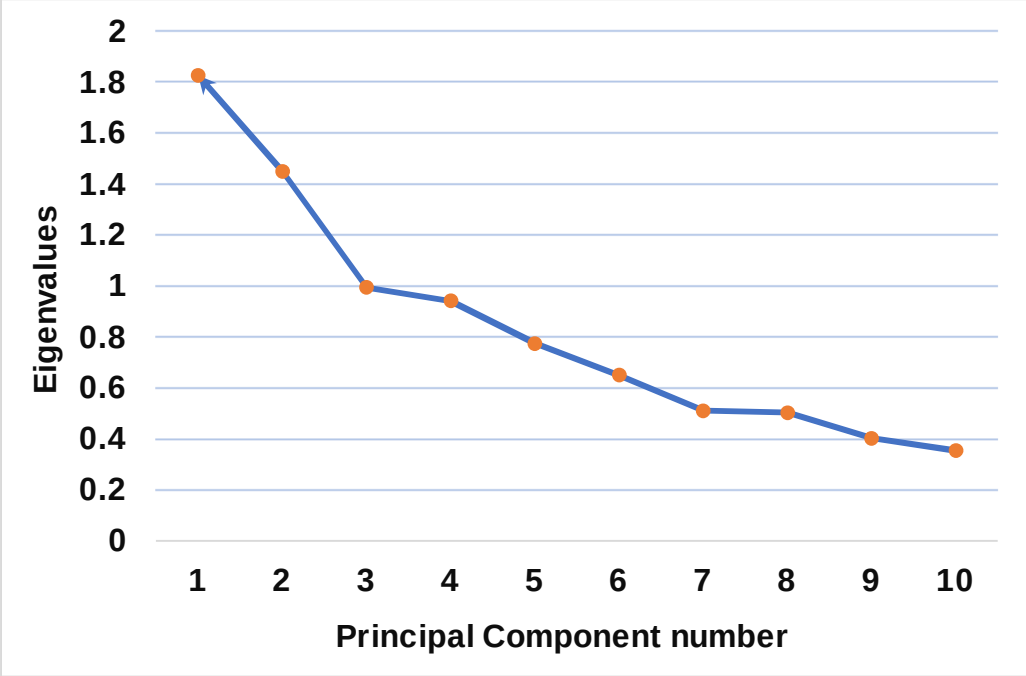


**Figure S4.** The scree plot for eigenvalues of PC1 and PC2 of TPN XIIc dihedral PCA in water solvent.

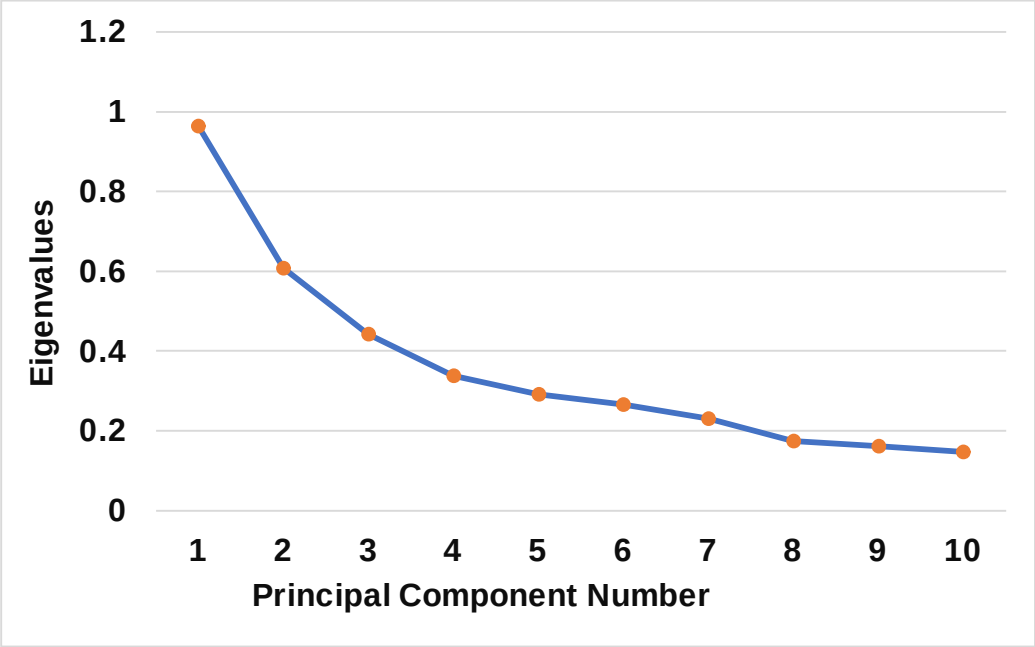


**Figure S5.** The scree plot for eigenvalues of PC1 and PC2 of TPN XIIc dihedral PCA in chloroform solvent.