

Cite this: *Chem. Sci.*, 2022, 13, 1957

All publication charges for this article have been paid for by the Royal Society of Chemistry

# How well does molecular simulation reproduce environment-specific conformations of the intrinsically disordered peptides PLP, TP2 and ONEG?<sup>†</sup>

Lauren M. Reid,<sup>abc</sup> Ileana Guzzetti,<sup>d</sup> Tor Svensson,<sup>d</sup> Anna-Carin Carlsson,<sup>e</sup> Wu Su,<sup>d</sup> Tomas Leek,<sup>d</sup> Lena von Sydow,<sup>d</sup> Werngard Czechtizky,<sup>d</sup> Marija Miljak,<sup>a</sup> Chandra Verma,<sup>bfg</sup> Leonardo De Maria<sup>id</sup> and Jonathan W. Essex<sup>id</sup>\*<sup>a</sup>

Understanding the conformational ensembles of intrinsically disordered proteins and peptides (IDPs) in their various biological environments is essential for understanding their mechanisms and functional roles in the proteome, leading to a greater knowledge of, and potential treatments for, a broad range of diseases. To determine whether molecular simulation is able to generate accurate conformational ensembles of IDPs, we explore the structural landscape of the PLP peptide (an intrinsically disordered region of the proteolipid membrane protein) in aqueous and membrane-mimicking solvents, using replica exchange with solute scaling (REST2), and examine the ability of four force fields (ff14SB, ff14IDPSFF, CHARMM36 and CHARMM36m) to reproduce literature circular dichroism (CD) data. Results from variable temperature (VT) <sup>1</sup>H and Rotating frame Overhauser Effect Spectroscopy (ROESY) nuclear magnetic resonance (NMR) experiments are also presented and are consistent with the structural observations obtained from the simulations and CD. We also apply the optimum simulation protocol to TP2 and ONEG (a cell-penetrating peptide (CPP) and a negative control peptide, respectively) to gain insight into the structural differences that may account for the observed difference in their membrane-penetrating abilities. Of the tested force fields, we find that CHARMM36 and CHARMM36m are best suited to the study of IDPs, and accurately predict a disordered to helical conformational transition of the PLP peptide accompanying the change from aqueous to membrane-mimicking solvents. We also identify an  $\alpha$ -helical structure of TP2 in the membrane-mimicking solvents and provide a discussion of the mechanistic implications of this observation with reference to the previous literature on the peptide. From these results, we recommend the use of CHARMM36m with the REST2 protocol for the study of environment-specific IDP conformations. We believe that the simulation protocol will allow the study of a broad range of IDPs that undergo conformational transitions in different biological environments.

Received 27th June 2021  
Accepted 3rd January 2022

DOI: 10.1039/d1sc03496k

rsc.li/chemical-science

## Introduction

Intrinsically disordered proteins and peptides (IDPs) do not adopt stable secondary or tertiary structures, and are best structurally characterised as ensembles of flexible conformations.<sup>1</sup> While the traditional paradigm holds that biological processes are mediated by macromolecules with fixed three-dimensional structures, it is now becoming increasingly evident that IDPs play a very important role.<sup>2</sup> In the absence of a stable structure and characterised by high levels of flexibility, IDPs can easily interconvert between multiple conformations. This enables them to bind to different macromolecules, resulting in diverse functionality.<sup>2–6</sup> For example, the intrinsically disordered N-terminal transactivation domain of the protein p53 is structured into a helix upon binding to a hydrophobic cleft in the proteins MDM2 or p300.<sup>3,7–9</sup> Similarly the

<sup>a</sup>School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. E-mail: jwe1@soton.ac.uk

<sup>b</sup>Bioinformatics Institute (A\*STAR), 30 Biopolis Street, Matrix, 138671, Singapore

<sup>c</sup>MedChemica Ltd, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

<sup>d</sup>Medical Chemistry, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

<sup>e</sup>Early Chemical Development, Pharmaceutical Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

<sup>f</sup>Department of Biological Sciences, National University of Singapore, 16 Science Drive 4, 117558, Singapore

<sup>g</sup>School of Biological Sciences, Nanyang Technological University, 60 Nanyang Dr, 637551, Singapore

<sup>†</sup> Electronic supplementary information (ESI) available. See DOI: 10.1039/d1sc03496k

disordered C-terminal region of p53 has been shown to adopt diverse secondary structures when complexed to different protein partners.<sup>10</sup> The study of the structural dynamics of IDPs and their participation in protein–protein interactions (PPIs) using computational and experimental studies is increasingly illuminating the subtle yet complex nature of molecular interactions governing biomolecular mechanisms.<sup>10</sup>

The ability of IDPs to assume different shapes depending on the interacting partner or the environment may also enable them to play a significant role in the multimerisation of membrane proteins. For example, the 19-residue C-terminus of the membrane proteolipid protein (referred to as the PLP peptide henceforth) is intrinsically disordered and resides on the cytoplasmic side of the membrane.<sup>6</sup> Structural studies of the PLP peptide in different biologically-relevant solvent environments have suggested that it is involved in the multimerisation of the full protein through its flexible rearrangement into a membrane-associated helix, followed by its aggregation and  $\beta$ -sheet formation with the C-terminal peptides from other PLP monomers.<sup>6</sup> The aggregation of IDPs is also known to be associated with several diseases,<sup>11</sup> including Alzheimer's disease in which intrinsically disordered amyloid  $\beta$  (A $\beta$ ) peptides form misfolded oligomers that lead to the formation of fibrils.<sup>12</sup> Studies have shown that the conformational ensembles of A $\beta$  monomers are very sensitive to osmolyte concentrations, and it is thought that such environment-dependent conformational transitions promote oligomerisation.<sup>13</sup>

There are many known cell-penetrating and antimicrobial peptides (CPPs and AMPs, respectively) that are intrinsically disordered in aqueous solution but adopt ordered structures upon binding to, or insertion into, cell membranes.<sup>6,14,15</sup> CPPs are a class of small peptides that are found to translocate across cell membranes and can deliver large, polar molecules into cells,<sup>16</sup> while AMPs are found to interact with and destabilise bacterial cell membranes, causing cell death.<sup>17</sup> The conformational sensitivity of such peptides to their biological environment is essential for the membrane-penetration process, since the arrangement of hydrophobic and hydrophilic residue side chains and the shielding of the polar backbone through the formation of intramolecular hydrogen bonds, allow the peptides to interact favourably with the polar headgroups and the lipid chain regions of cell membranes. There is much interest in deducing the exact mechanisms by which CPPs and AMPs interact with cell membranes, as this knowledge could prove crucial in aiding the design of intracellular delivery systems, biomedical imaging agents and/or new antimicrobial treatments in the fight against antibiotic resistance.<sup>16–19</sup>

From the examples given above, it is obvious that IDPs are involved in a broad range of biological processes and offer an exciting line of study. However, the very nature of IDPs (their flexibility and lack of defined structure) makes them a difficult target for conventional experimental techniques, such as X-ray crystallography that is used to capture the crystalline structure of proteins, and instead techniques that enable the measurement of dynamic structures are necessary. Examples of such techniques that have been used to study IDPs include nuclear magnetic resonance (NMR),<sup>20</sup> circular dichroism (CD),<sup>20</sup> Raman

spectroscopy,<sup>21</sup> Fourier transform infrared spectroscopy (FT-IR),<sup>22</sup> small-angle X-ray scattering (SAXS)<sup>23</sup> and static and dynamic light scattering (SLS/DLS).<sup>24</sup> NMR is particularly popular in the study of IDPs as it can provide atomistic detail and detect the presence or absence of structure in solution-phase peptides; however, interpretation of NMR output can be challenging and time-consuming, especially since the diverse conformational ensembles of IDPs lead to noisy signals. CD is also frequently used for IDP characterisation and provides spectra that are characteristic of secondary structure or disorder within the peptide; however, only global structural descriptors can be identified, making CD a low-resolution technique. For a recent review of the structural characterisation techniques used to study IDPs, see ref. 25.

In light of these challenges, atomistic molecular dynamics (MD) simulation, which involves modelling biological systems at the atomistic level and calculating the dynamic evolution of the atoms, provides an alternative or complementary approach to help elucidate the dynamic structures and mechanisms of IDPs.<sup>26,27</sup> However, biomolecular MD comes with its own set of challenges suggesting that simulation results must be interpreted with caution and validated against experimental data.<sup>28</sup> One of the biggest challenges is the computational expense involved in modelling biological systems atomistically, with the implication that simulations of large systems can only access  $\mu$ s timescales with currently available hardware. In reality, many biological processes occur over much longer timescales as they can involve multiple complex macromolecular folding, aggregation, binding and insertion events, which are often associated with large free energy barriers.<sup>28</sup> The simulation of IDPs presents a further challenge since the conformational space available to the peptides is vast and long timescales may be needed to exhaustively sample the full structural landscape. Recent studies have improved IDP conformational sampling by applying enhanced sampling techniques, such as parallel tempering<sup>29</sup> or replica exchange,<sup>30</sup> or by incorporating experimental data to restrain the simulation or reweight the resulting structural ensembles.<sup>31</sup>

Another important consideration for biomolecular MD is the choice of force field. Force field accuracy greatly depends on the data and protocol used in the parameterisation procedure and it is often reported that different force fields are better suited to different systems and properties of interest.<sup>32,33</sup> It is currently accepted that many standard protein force fields are better suited to globular proteins than IDPs, due to the greater availability of globular experimental data for parameterisation.<sup>34–36</sup> With increasing interest in IDPs, the correction of force fields to better sample disordered structures is a hot topic.<sup>34–40</sup> Established by Best *et al.*,<sup>39</sup> one technique has been to refine the protein–water Lennard-Jones parameters to improve solvation and reduce the sampling of compact structures, which was originally applied to the Amber force field ff03w. The resulting force field, ff03ws, was shown to sample more extended and flexible conformations.<sup>35,39</sup> Similar approaches, where the protein–water interactions were scaled, were adopted by Robustelli *et al.*<sup>36</sup> and Huang *et al.*<sup>38</sup> to improve IDP sampling of the Amber ff99SB-ILDN and CHARMM36m force fields,



respectively. Approaches that tackle the protein interaction parameters directly have also been utilised to improve IDP sampling, for example some recent force fields include a CMAP correction term (a matrix of energy corrections applied to the  $\phi/\psi$  backbone dihedral space) to encourage the simulation to sample disordered structures.<sup>34,37,38</sup> This approach has been applied in multiple iterations of the CHARMM force fields,<sup>41,42</sup> resulting in the most recent CHARMM36m,<sup>38</sup> and to correct the Amber ff14SB force field<sup>43</sup> for IDP sampling to create ff14IDPSFF.<sup>44</sup> However, as has been discussed, many IDPs are involved in environment-dependent conformational transitions, so it is also important to test force fields in a variety of biologically-relevant environments to determine their conformational sensitivity.

In this work, we test the ability of the enhanced conformational sampling approach replica exchange with solute scaling (REST2),<sup>45</sup> and a number of force fields, to explore the conformations accessed by IDPs in aqueous and hydrophobic solvents, as would be relevant in the study of membrane proteins, CPPs, AMPs, and even for peptides that bind to hydrophobic pockets in other proteins (e.g. the p53-MDM2 interaction). With validation in mind, we apply the simulation protocol to the PLP peptide, which has literature circular dichroism (CD) data in aqueous and a variety of membrane-mimicking/hydrophobic environments.<sup>6</sup> To provide additional experimental indications of the different PLP peptide conformations, variable temperature (VT) <sup>1</sup>H and Rotating frame Overhauser Effect Spectroscopy (ROESY) nuclear magnetic resonance (NMR) experiments were recorded in aqueous and membrane-mimicking solvent mixtures. Following this, we apply the validated simulation protocol to TP2, a spontaneous membrane translocating peptide (SMTP) that is known to penetrate cells and artificial lipid vesicles as a monomer at low concentrations, and compare the results against a negative control peptide, ONEG, which has a similar amino acid sequence but does not possess membrane-translocating properties.<sup>46</sup> We show that the CHARMM36<sup>42</sup> and CHARMM36m<sup>38</sup> force fields are best able to reproduce the conformational transition that occurs in the PLP peptide as the solvent changes from aqueous to membrane-mimicking. ff14SB<sup>43</sup> overpredicts the helical structure of the PLP peptide and TP2 in aqueous solvent, and ff14IDPSFF<sup>44</sup> is unsuccessful in predicting the PLP peptide structure in any of the tested solvents. In contrast to a previously reported experimental CD study,<sup>47</sup> we identify an  $\alpha$ -helical structure of TP2 in the membrane-mimicking solvents and discuss the possible mechanistic significance of this observation; such a contradiction, and the additional mechanistic detail that is inferred from simulations, highlights the importance of complementing experimental structural studies with simulation results.

## Methods

### Computational methods

**System setups.** Initial coordinate files for the PLP peptide, TP2 and ONEG were built in extended conformations using the

Amber tleap program.<sup>48</sup> The sequences of the peptides are as follows:<sup>6,47</sup>

PLP: Ac-IAATYNFAVLKLMGRGTFK

TP2: PLIYLRLLRGQF-NH<sub>2</sub>

ONEG: PLGRPQLRRGQF-NH<sub>2</sub>

Where Ac – represents an acetyl N-terminal cap and –NH<sub>2</sub> represents an amide C-terminal cap.

A single monomer of each peptide was solvated with three different solvent systems to model the different biological environments the peptides encounter during membrane association and penetration:

- Water, neutralising Cl<sup>−</sup> ions and 0.15 M Na<sup>+</sup>Cl<sup>−</sup> ions were used to model the extracellular aqueous medium.
- Trifluoroethanol (TFE) : water, ~1 : 1 mol% (~8 : 2 vol%) + neutralising Cl<sup>−</sup> ions were used as a model of the hydrophobic interior of a cell membrane.
- Chloroform : methanol : water, ~0.25 : 0.48 : 0.27 mol% (~4 : 4 : 1 vol%) + neutralising Cl<sup>−</sup> ions were used as a second model of the hydrophobic interior of a cell membrane for the Amber force fields.
- Chloroform : methanol : water, ~0.24 : 0.64 : 0.11 mol% (~4.1 : 5.5 : 0.4 vol%) + neutralising Cl<sup>−</sup> ions were used with the CHARMM force fields.

The choices of solvent models and ratios are justified in the discussion.

Four different force fields were used to model the PLP peptide in each solvent system:

- ff14SB<sup>43</sup>
- ff14IDPSFF<sup>44</sup>
- CHARMM36<sup>42</sup>
- CHARMM36m<sup>38</sup>

Two force fields were subsequently used to model TP2 and ONEG in each solvent system:

- ff14SB
- CHARMM36m

For the Amber force fields (ff14SB and ff14IDPSFF), GAFF parameters of TFE were taken from ref. 49 and GAFF parameters for chloroform and methanol were taken from ref. 50. TIP3P water was used with the Amber force fields in all cases, except in the TFE : water mixture where the TIP4P-ew model was used, since this was one of the water models used in the optimisation and validation of the TFE model.<sup>49</sup> The standard Amber ion parameters were used with the Amber force fields.<sup>43</sup> For the CHARMM force fields, CGenFF parameters of TFE and methanol were taken from ref. 51 and the Dietz-Heinzinger (DH) model of chloroform<sup>52</sup> was used (as tested for CHARMM force fields in reference<sup>53</sup>). The CHARMM modified TIP3P water model and the standard CHARMM ion parameters were used with the CHARMM force fields in all cases.<sup>42</sup>

To assess simulation convergence, two simulations were performed for each peptide/force field/solvent combination; (i) one starting from the extended conformation built in tleap, and (ii) one starting from a helical structure found for each peptide in this study. Details of each system setup and simulation length can be found in Table S1.†

**Simulation protocol.** GROMACS-5.1.4<sup>54</sup> patched with Plumed2-2.2<sup>55</sup> was used for all simulations. The leap-frog



integrator was used with a time step of 2 fs and periodic boundary conditions were employed. Coulomb interactions were treated with the Particle Mesh Ewald (PME) method, with a real space cutoff of 1.0 nm for the Amber force fields and 1.2 nm for the CHARMM force fields, as is common practice for these force fields. van der Waals interactions were treated with cutoffs of 1.0 nm and 1.2 nm for the Amber and CHARMM force fields respectively. The systems were first energy minimised using the steepest-descent algorithm until the maximum force was  $<1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  and subsequently using the conjugate gradient method until the maximum force was  $<50 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . The LINCS algorithm was used to constrain all hydrogen-containing bonds except when the DH chloroform was used, in which case the SHAKE algorithm was employed. NVT equilibration was performed on each replica with the Nose–Hoover thermostat<sup>56,57</sup> at 300 K for 1 ns. This was followed by NPT equilibration with the Parrinello–Rahman barostat<sup>58</sup> at 1 bar for 1 ns. Each system was simulated using the REST2 protocol described in the ESI,<sup>†</sup> with the replicas spaced according to Tables S2 and S3<sup>†</sup> for the PLP peptide and TP2/ONEG respectively. Replica coordinate exchanges were attempted every 1 ps. Each simulation was run until structural convergence was observed (convergence criteria discussed below) or until 1  $\mu\text{s}$  was reached, resulting in the simulation lengths given in Table S1.<sup>†</sup> The resulting total simulation time reported here is over 225  $\mu\text{s}$ .

**Analysis methods.** The following analyses were performed on the base replica of each simulation:

- Dictionary of protein secondary structure (DSSP)<sup>59</sup> was implemented using mdtraj,<sup>60</sup> which assigns to each residue of each frame a secondary structure identifier based on the atomic coordinates of the backbone. These were plotted against simulation time to monitor convergence and were used to calculate the structural content from the equilibrated parts of each trajectory. The average structural content percentages and standard errors were calculated from the two independent simulations of each system.

- Principal component analysis (PCA)<sup>61</sup> was performed on the backbone  $\phi$  and  $\psi$  dihedral angles (using a  $(\sin(x), \cos(x))$  transformation on the angles to eradicate wrap around prior to PCA) taken from the full set of base replica data for the given peptide using PyEMMA.<sup>62</sup> PCA is a data dimensionality reduction technique that assembles a series of linear, non-correlated coordinates that is used to extract the most extreme motions in the data. By combining the base replica data from the different simulation conditions, a generic set of PCs were built for each peptide, such that the results obtained under different simulation conditions for the given peptide could be compared using the same PCs; for example, the 24 base replicas for the PLP peptide (Table S1<sup>†</sup>) were combined to produce PCs that could be used to compare the conformational space visited by the PLP peptide in the different solvents. The base replica trajectories were divided in half and projected onto the first two principal components (PC1 and PC2) in order to assess convergence.

- Two-dimensional free energy surfaces (FESs) were built from the equilibrated parts of the trajectories with respect to

PC1 and PC2, using PyEMMA that implements the following relation:

$$F(x) = -k_{\text{B}}T \ln P(x)$$

where  $P(x)$  is the probability distribution of coordinate  $x$ , taken as the 2 dimensional histogram of PC1 and PC2.

- MDASH-3 was used to employ the DASH clustering algorithm on the combined equilibrated parts of the base replica trajectory of each peptide/force field/solvent combination.<sup>63</sup> The DASH algorithm clusters molecular conformations based on the  $\phi$  and  $\psi$  backbone dihedral angles explored throughout the simulation. First, the dihedral angle histograms are binned by placing bin edges midway between maxima, which are identified when they are  $>2.4\%$  of the trajectory frames. Each trajectory frame can then be described by a sequence of dihedral angle states, creating discrete conformational states. Conformational states present in  $>1\%$  of the trajectory frames are retained as clusters. In this work, DASH clusters have been further clustered into macrostates based on their location on the PC FES.

Simulation convergence was assessed and equilibration times chosen using several criteria:

- When the DSSP plots from the two simulation repeats showed similar secondary structure content.
- When the first and second halves of the equilibrated portions of the two simulations explored a similar PC1 and PC2 space, assessed using histograms.
- When the percentage structural assignment of the two simulations provided an average with a minimal standard error.

## Experimental methods

**PLP peptide preparation.** 2-Chlorotriyl chloride resin (300 mg, 0.3 mmol,  $1.0\text{--}1.8 \text{ mmol g}^{-1}$ ) was swelled in DCM (5 mL) for 1 h. A solution of Fmoc-Phe-OH (81 mg, 0.21 mmol) and DIPEA (209  $\mu\text{L}$ , 1.20 mmol) in DCM (5 mL) was added. The reaction was shaken for 21 h at room temperature. The resin was filtered off and washed with DCM ( $1\times$ ), DMF ( $1\times$ ) and DCM ( $1\times$ ). Subsequently, the resin was added DCM/MeOH/DIPEA (17 : 2 : 1, 10 mL) and the reaction was shaken for 30 min at room temperature. The resin was filtered off and washed with DCM ( $2\times$ ), DMF ( $1\times$ ), DCM ( $1\times$ ), MeOH ( $1\times$ ), DCM ( $1\times$ ), MeOH ( $1\times$ ), DCM ( $2\times$ ) and  $\text{Et}_2\text{O}$  ( $2\times$ ), and the resin was dried *in vacuo*. The rest of the peptide was synthesised using the Biotage Alstra synthesiser (automated microwave peptide synthesiser) in a 30 mL reactor vial by swelling of Rink Amide ChemMatrix resin (298 mg, 0.2 mmol,  $0.48 \text{ mmol g}^{-1}$ ) in DCM at  $40^\circ\text{C}$  for 30 min using orthogonally protected Fmoc-amino acids (0.15 M, 4 eq.) and 1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo [4,5-*b*]pyridinium 3-oxide hexafluorophosphate (0.5 M, 3.92 eq.) in DMF and *N,N*-diisopropylethylamine (2 M, 8 eq.) in NMP. After the last Fmoc-deprotection an *N*-acetylation was carried out using  $\text{Ac}_2\text{O}$  (5 M, 40 eq.) in DMF and *N,N*-diisopropylethylamine (2 M, 8 eq.) in NMP. The purification was done by preparative HPLC using a Waters Atlantis T3 ODB column (5  $\mu\text{m}$ ,  $19 \times 150 \text{ mm}$ ) with a gradient of 5% B for 0.5 min, 5–21% B in 1.5 min, 21–26% B in 14 min (A:  $\text{H}_2\text{O}/\text{HOAc}$  100 : 0.5, B:





MeCN) and a flow of 30 mL min<sup>-1</sup> at room temperature. The volume injected was 650 µL with a concentration of 95 mg mL<sup>-1</sup>. Fractions were collected at 230 nm. Pure compound was identified by analytical HPLC with a Waters Acquity BEH C18 column (1.7 µm, 2.1 × 50 mm) using a linear gradient 5–60% in 9.3 min (A: H<sub>2</sub>O/TFA 100 : 0.1, B: MeCN/H<sub>2</sub>O/TFA 95 : 5 : 0.1) and a flow of 0.4 mL min<sup>-1</sup> at 45 °C. Detection was done at 230 nm. The pure product (amount 42 mg, yield 6.6%, UV purity 97%) was found to be a white solid. ESI MS (+) (*m/z*): calculated for C<sub>100</sub>H<sub>159</sub>N<sub>25</sub>O<sub>25</sub>S = 1072.8, found = 1072.7 (M + 2H)<sup>2+</sup> and ESI MS (–) (*m/z*): calculated for C<sub>100</sub>H<sub>159</sub>N<sub>25</sub>O<sub>25</sub>S = 715.5, found = 715.5 (M – 3H)<sup>3–</sup>.

**PLP peptide NMR.** All experiments were collected on a Bruker AVANCE NEO 600 MHz spectrometer equipped with a 5 mm detection cryoprobe (CP TCI H-C/N-D-05 Z). The PLP peptide was dissolved in two different solvents: 90%-H<sub>2</sub>O/10%-D<sub>2</sub>O and 80%-TFE/20%-H<sub>2</sub>O (v/v), at 1 mM concentration,<sup>64</sup> due to observed aggregational behaviour at higher concentration in aqueous solution. Deuterated trifluoroethanol was obtained from Alfa Aesar, Karlsruhe Germany. For the experiments performed in aqueous solution, 3-(trimethylsilyl)-propionate-d<sub>4</sub>, sodium salt (TSP) was used as external standard;<sup>65</sup> tetramethylsilane (TMS) was used as internal standard for the experiments in TFE/H<sub>2</sub>O. <sup>1</sup>H 2D DQF-COSY, TOCSY, ROESY and HSQC were recorded at 298 K using the following Bruker pulse programs, cosydfesgpph, mlevesgpph, roesyegpph and hsqc-detgpsisp2. ROESY spectra were recorded using 512 increments of 8 K data points and 64 scans per increment. A mixing time of 200 ms was used for the ROESY spectra and a mixing time of 120 ms was used for TOCSY spectra. Excitation sculpting sequence was used for water suppression in <sup>1</sup>H 2D DQF-COSY, TOCSY and ROESY. Spectra were processed in Topspin 4.0.8. For ROESY spectra, t1 and t2 dimensions were both zero-filled respectively to 1 K and 8 K real data points, sine-bell squared window functions were applied in both dimensions. The spectra were baseline corrected, using Bernstein Polynomial and Polynomial fit methods, and analyzed in MestReNova 14.1.1.

## Results and discussion

### The PLP peptide

In order to assess the ability of different force field/solvent combinations to predict aqueous and membrane-associated peptide conformational ensembles, we chose the PLP peptide as a test case.<sup>6</sup> As discussed in the introduction, the PLP peptide is part of the C-terminus of the PLP protein, which is a myelin protein that spans the membrane.<sup>66</sup> The C-terminus resides on the cytoplasmic side of the membrane and has been shown to be a random coil in aqueous medium but to adopt structure in membrane-mimicking environments.<sup>6,66</sup> Specifically, CD data indicate that the PLP peptide adopts an  $\alpha$ -helix in TFE : water mixtures and a combination of  $\alpha$ - and  $3_{10}$ -helices in the presence of lipid vesicles at low peptide concentrations (Table 1).<sup>6</sup> The peptide is also shown to adopt a  $\beta$ -sheet structure in the presence of lipid vesicles at high peptide concentrations but

this is not relevant for this study, since we simulate only one peptide in each system.<sup>6</sup>

The PLP peptide was simulated using the REST2 protocol in three solvent systems, with four force fields (Table 1). The choice of TFE : water (~8 : 2 vol%) and chloroform : methanol : water (~4 : 4 : 1 vol%) as membrane-mimicking solvents was made because the use of these solvent mixtures for membrane-protein extraction, CD and/or NMR has precedent in the experimental literature.<sup>6,67–70</sup> Hydrophobic solvents have also been used to mimic the low polarity of protein binding pockets for CD and NMR experiments of intrinsically disordered protein-binding peptides.<sup>4,71</sup> Following initial simulations, it was observed that the chloroform : methanol : water (~4 : 4 : 1 vol%) mixture separates into two phases when using the CGenFF/DH models (Fig. S1†) but not when using the GAFF models (Fig. S2†). The biphasic nature of the solvent is not necessarily incorrect, since inspection of the associated experimental phase diagram<sup>72</sup> reveals that the ~4 : 4 : 1 vol% ratio falls on the boundary between monophasic and biphasic behaviour. However, we chose to slightly alter the solvent ratio for the CGenFF/DH models to create a monophasic mixture, as this is more consistent with the solvent behaviour observed in NMR and CD experiments (as noted in ref. 69). Therefore, a monophasic ratio of ~4 : 4 : 1 vol% was used for the GAFF solvent and a monophasic ratio of ~4.1 : 5.5 : 0.4 vol% was used for the CGenFF/DH solvent (Fig. S3†).

NMR experiments were also conducted to investigate the conformational propensity of the PLP peptide in H<sub>2</sub>O : D<sub>2</sub>O (9 : 1 vol%) and TFE : H<sub>2</sub>O (8 : 2 vol%), in order to provide further experimental data to compare against our simulation results. Specifically, the H<sub>2</sub>O : D<sub>2</sub>O (9 : 1 vol%) NMR solvent was used to provide an experimental benchmark for the simulations in water, and the TFE : H<sub>2</sub>O (8 : 2 vol%) NMR solvent was used to provide an experimental benchmark for the simulations in TFE : water.

Given the vast amounts of simulation data produced in this study, the following sub-sections will firstly present the results from two of the PLP peptide simulations as examples, followed by a comparison with the results from the NMR experiments, before providing a concise comparison of the results from each force field/solvent combination with the NMR and literature CD data.<sup>6</sup> Full analyses of each trajectory can be found in the ESI.†

**Simulation analysis.** Fig. 1 shows analyses of the PLP peptide simulations using the CHARMM36m force field in water. The figure includes results from the two simulations, one starting from an extended conformation and the other starting from a helical conformation. The two simulations were performed to assess convergence: if both simulations explored a similar conformational space after an initial equilibration period (determined using the criteria described in the methods), they were taken to be in equilibrium.

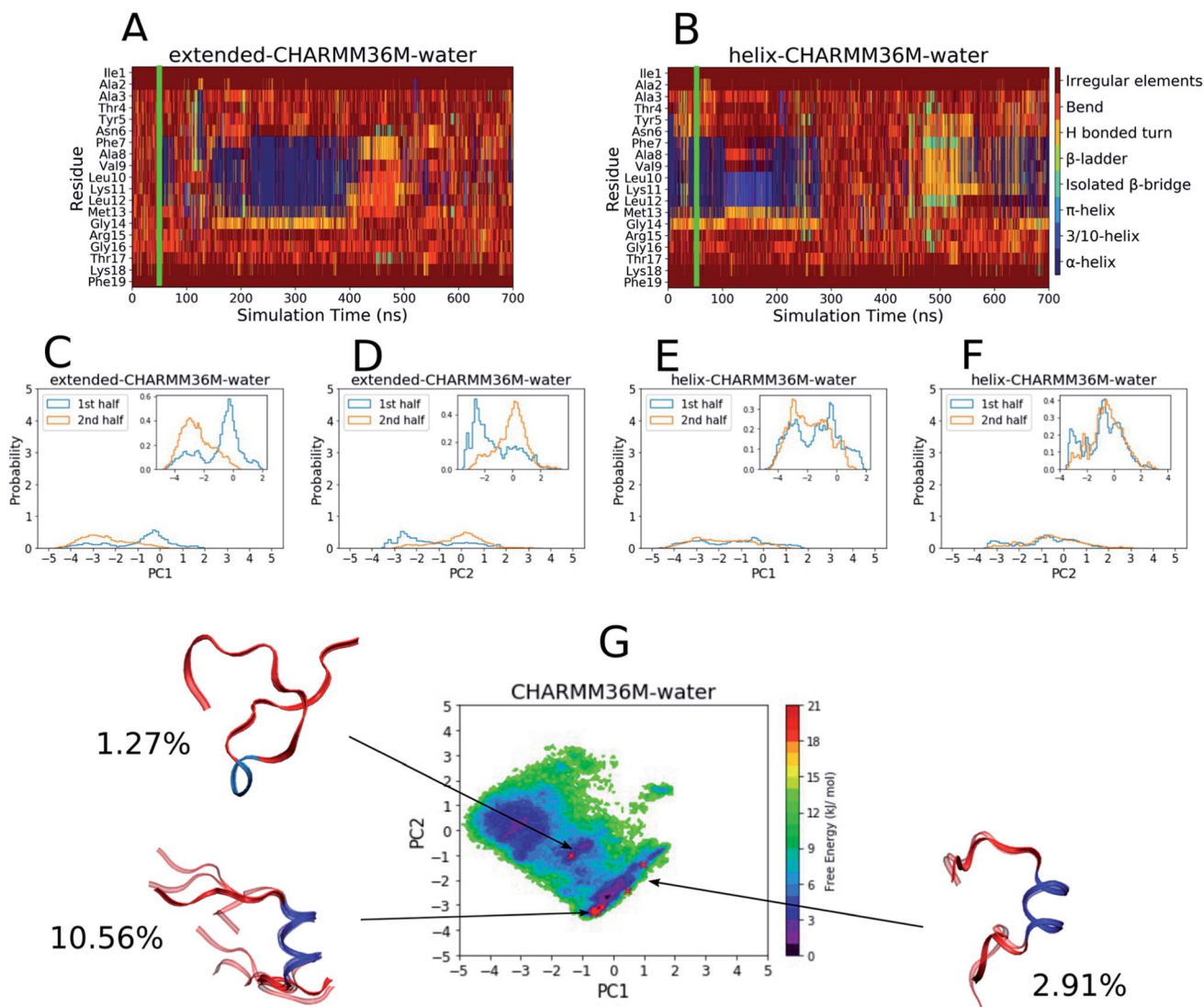
To define the conformational space, DSSP assignment and the first two PCs from dihedral backbone PCA were used. The DSSP assignment throughout both simulations can be seen in Fig. 1A and B with a green bar showing the point at which the equilibration periods end and from which the remaining



analyses were performed. The plots show that the peptide mainly visited random coil structures but some helicity was occasionally observed between residues 7–13 in both simulations. The PC histograms in Fig. 1C–F show that, although the simulation starting from an extended conformation visited a different PC space in the first and second half of the trajectory, the full PC space covered by both trajectories was roughly similar, and the simulation starting from a helical conformation visited a similar PC space in both halves of the trajectory. This can also be seen in the DSSP plot (Fig. 1B), where the

peptide unfolds and then refolds towards the end of the simulation. Although the exploration of conformational space is changing throughout both simulations, we believe that it is fair to end the equilibration periods at 50 ns to include both the random coil and helical sampling that is consistent in both simulations.

The FES with respect to the PLP peptide PC1 and PC2 coordinates was built from the combined equilibrated trajectory frames and can be seen in Fig. 1G, where there are several wide and shallow energy minima as a result of the diffuse sampling

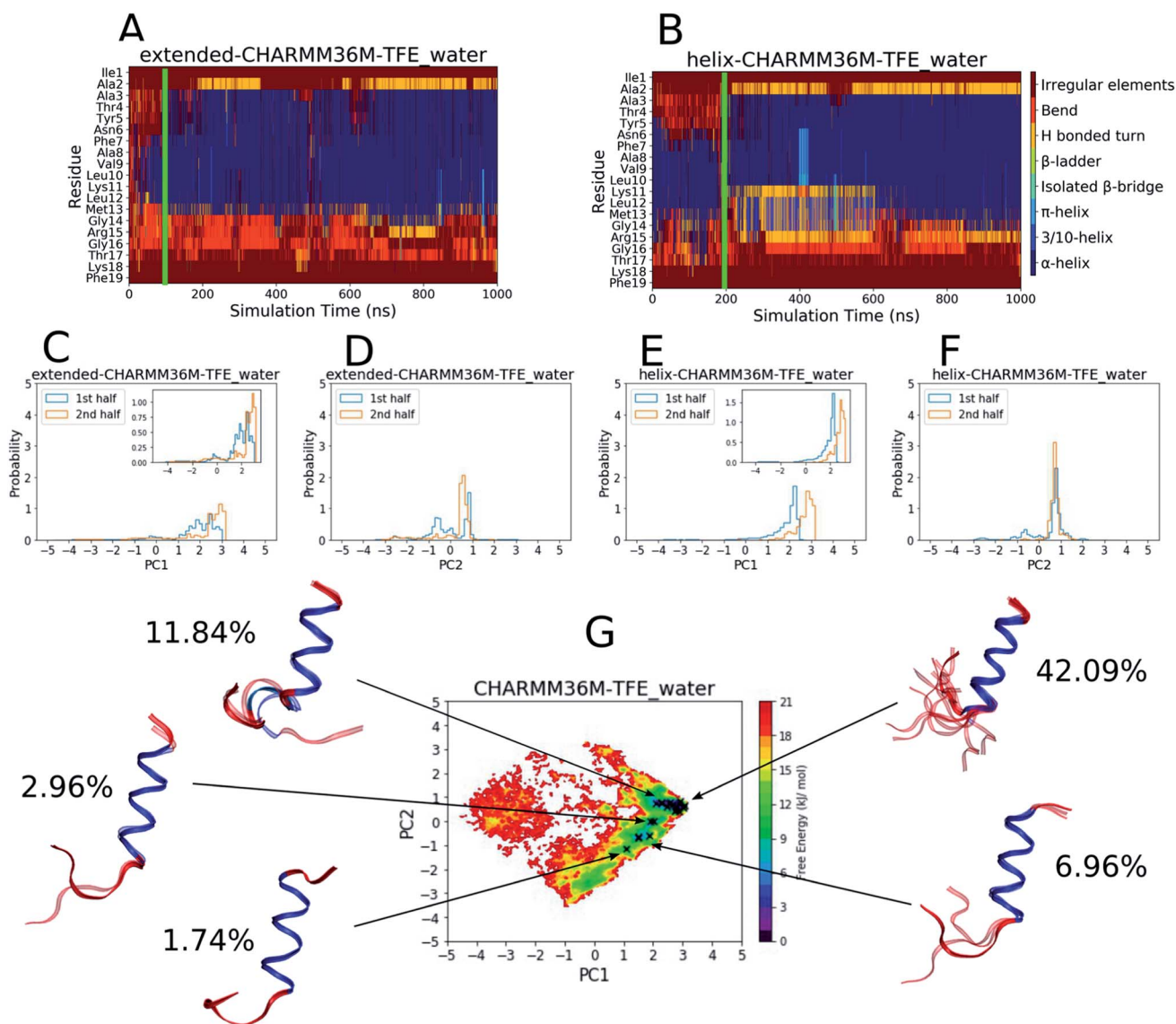


**Fig. 1** Results from the PLP peptide CHARMM36m water simulations. (A and B) DSSP analysis of the simulations starting from (A) an extended conformation and (B) a helical conformation. The green bar on each plot indicates the time from which equilibrium analysis was performed. (C–F) Histograms of the projections of the simulation coordinates onto PC1 (C and E) and PC2 (D and F) built from the first and second halves of the equilibrated part of the extended (C and D) and helical (E and F) trajectories. The y axes' maxima are set to 5 to allow comparison between all of the PLP peptide histograms (Fig. 2C–F and S19–S28C–F†), however zoomed in inserts are provided when the largest peak is less than 2. (G) The FES with respect to PC1 and PC2 built from the combined equilibrated parts of the two trajectories. The energy minimum is set to 0 and the colourbar range is fixed at 0–21 kJ mol<sup>−1</sup> to allow comparison between all of the PLP peptide FESs (Fig. 2G and S19–S28G†). The DASH clusters are overlaid on the surface as red crosses, and those that occupy a similar PC and DSSP space are grouped into macrostates with the corresponding structural representatives shown as a superposition of the cluster centroids. The structures are shown as ribbons with random coil, turn and bend residues shown in red,  $\alpha$ -helical residues shown in dark blue,  $3_{10}$ -helical residues shown in light blue,  $\beta$ -bridge residues shown in cyan and  $\beta$ -ladder residues shown in lime. The percentage of trajectory frames occupied by each macrostate is also shown.



of PC space. The DASH clusters are labelled on the FES and grouped into macrostates based on their DSSP assignment and PC values, and the percentage of trajectory frames that each macrostate represents is shown. The macrostate structures are shown on the FES as overlaid DASH cluster centroids. Only about 15% of the trajectory frames belong to a given cluster, since most of the trajectory is occupied by disordered structures that are too sparsely populated for the DASH clustering

algorithm to identify; for example the shallow energy minimum at  $[PC1 \sim -2 \text{ to } -4, PC2 \sim -1 \text{ to } 1]$  does not contain any clusters because this is a disordered region of PC space and the energy minimum is too diffuse to contain heavily populated states. The macrostates reside in or near the remaining two shallow FES minima, two of which have a central  $\alpha$ -helix region (dark blue) with random coil residues (red) on either side, while the third is



**Fig. 2** Results from the PLP peptide CHARMM36m TFE : water ( $\sim 8 : 2$  vol%) simulations. (A and B) DSSP analysis of the simulations starting from (A) an extended conformation and (B) a helical conformation. The green bar on each plot indicates the time from which equilibrium analysis was performed. (C–F) Histograms of the projections of the simulation coordinates onto PC1 (C and E) and PC2 (D and F) built from the first and second halves of the equilibrated part of the extended (C and D) and helical (E and F) trajectories. The y axes maxima are set to 5 to allow comparison between all of the PLP peptide histograms (Fig. 1C–F and S19–S28C–F†), however zoomed in inserts are provided when the largest peak is less than 2. (G) The FES with respect to PC1 and PC2 built from the combined equilibrated parts of the two trajectories. The energy minimum is set to 0 and the colourbar range is fixed at 0–21 kJ mol<sup>−1</sup> to allow comparison between all of the PLP peptide FESs (Fig. 1G and S19–S28G†). The DASH clusters are overlaid on the surface as black crosses, and those that occupy a similar PC and DSSP space are grouped into macrostates with the corresponding structural representatives shown as a superposition of the cluster centroids. The structures are shown as ribbons with random coil, turn and bend residues shown in red,  $\alpha$ -helical residues shown in dark blue,  $3_{10}$ -helical residues shown in light blue,  $\beta$ -bridge residues shown in cyan and  $\beta$ -ladder residues shown in lime. The percentage of trajectory frames occupied by each macrostate is also shown.





mainly random coil with a short  $3_{10}$ -helical region in the middle.

As a comparison, the equivalent analyses of the PLP peptide simulations using the CHARMM36m force field in TFE : water ( $\sim 8 : 2$  vol%) are shown in Fig. 2. In this case, simulation convergence was easier to establish because the DSSP plots (Fig. 2A and B) and PC histograms (Fig. 2C–F) have even greater similarity between the two simulations than those for the water simulations (Fig. 1A–F). The DSSP plots reveal a longer and more pronounced  $\alpha$ -helix that was maintained throughout most of both simulations. The FES in Fig. 2G is less diffuse and has a more-defined and deeper minimum at  $[PC1 \sim 2.8, PC2 \sim 0.6]$ , where longer helical conformations reside, compared to that in water (Fig. 1G). 21 DASH clusters were identified and grouped into five macrostates. In this case, the DASH clusters/macrostates represent a total of about 65% of the trajectory, showing that a higher percentage of the trajectory is occupied by well-defined structures, compared to the 15% in water. The macrostate structures have longer  $\alpha$ -helical regions and fewer random coil residues than those found in water.

The results in Fig. 1 and 2 show that the CHARMM36m PLP peptide adopted longer and more-populated  $\alpha$ -helical structures in TFE : water than in pure water. Although the water simulations still predicted some weak and short-lived helicity, the overall trend is consistent with the observation taken from experimental CD: the PLP peptide increases in helicity in the membrane-mimicking environment.<sup>6</sup> These results suggest that the CGenFF TFE : water solvent provides a good membrane-mimic for CHARMM36m simulations.

**NMR results.** Fig. 3 shows an expansion of the NH/NH region of the ROESY spectra obtained for the PLP peptide in aqueous solution (a) and in TFE/H<sub>2</sub>O mixture (b). In both solvents, from the analysis of the amide proton regions,  $d_{NN}$  ROE connectivities can be observed and were used for the sequential walk assignment (Fig. S12/Table S4 and Fig. S17/Table S6†).  $d_{NN}$  ROE connectivities, even if not always sufficient, are necessary

evidence of helical content.<sup>73</sup> In the TFE/H<sub>2</sub>O mixture, all of the amide protons have observed  $d_{NN}$  ROE connectivities with strong to medium intensities (Fig. 3b), indicating a uniform conformational behaviour and suggestive of helical content. For the peptide in aqueous solution, even though  $d_{NN}$  ROE connectivities are present, their intensities are more varied; weak, medium and strong intensity ROE can be observed (Fig. 3a). This suggests a greater conformational heterogeneity of the peptide in water.<sup>74</sup> Temperature coefficients ( $\Delta\delta/\Delta T$ ) were also calculated for the amide protons of the peptide in aqueous solution (Fig. S14/Table S5†). High absolute values of  $\Delta\delta/\Delta T$  were found, suggesting that none of the amide protons are involved in stable intramolecular H-bonds<sup>75</sup> and providing supporting evidence for the greater conformational heterogeneity of the peptide in H<sub>2</sub>O/D<sub>2</sub>O.

To more closely compare the simulation results with the ROESY spectra, distance heat maps between the backbone amide protons from the CHARMM36m simulations in water and TFE : water are shown in the ESI (Fig. S4–S7†), revealing the presence of a more ordered signal for the PLP peptide in TFE : water compared to water. The presence of a strong  $i$  to  $i + 3$  signal in the TFE : water heat map (Fig. S5†), along with the associated single peak histograms (Fig. S7†), results from the long and stable helix that is sampled, and this corresponds to the strong and ordered  $d_{NN}$  connectivities in the TFE : water ROESY spectrum (Fig. 3b). However, the presence of a weak  $i$  to  $i + 3$  signal in the water heat map (Fig. S4†), along with the associated multiple peak histograms (Fig. S6†), suggests that a short and unstable helix is sampled as well as other disordered states, and this is in agreement with the conformational heterogeneity suggested by the  $d_{NN}$  connectivities observed in the water ROESY spectrum (Fig. 3a) and the high absolute values of the temperature coefficients,  $\Delta\delta/\Delta T$  (Fig. S14/Table S5†).

An analysis of ROE correlations in the NH/aliphatic proton regions (Fig. S13 and S18†) was also performed for the PLP

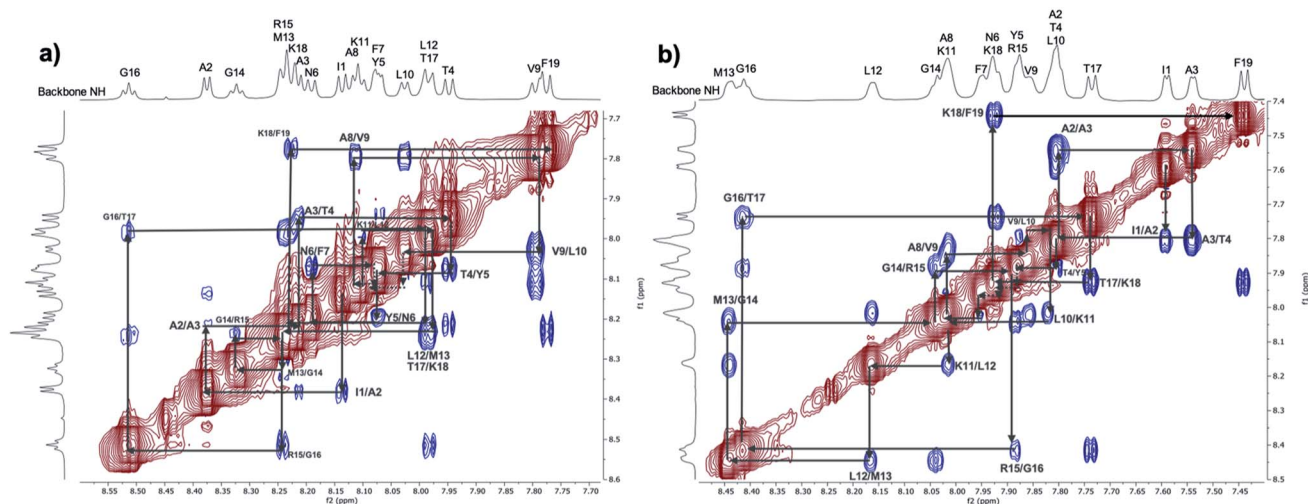


Fig. 3 ROESY spectra (normalised by largest peak, value: 100) at 298 K: expansion of the NH/NH region for the 1 mM PLP peptide sample in aqueous solution (a) and in mixed solvent TFE/H<sub>2</sub>O (b).





peptide in both aqueous solution and TFE/H<sub>2</sub>O. ROEs due to sequential and medium-range distances, like  $d_{\alpha N}(i, i + 3)$  and  $d_{\alpha N}(i, i + 4)$ , are observable for the PLP peptide in TFE/H<sub>2</sub>O, whereas only ROEs due to sequential distances are present for the peptide in aqueous solution, except for a few, weak correlations for some central residues. These regions of the ROESY spectra can be compared directly with the simulations using the distance heat maps shown in Fig. S8 and S9† that show the interactions between the backbone NH and aliphatic protons found in the CHARMM36m simulations of the PLP peptide in water and TFE : water. The heat maps correlate well with the experimental ROESY: the presence of a strong, medium-range signal ( $d_{\alpha N}(i, i + 3)$  and  $d_{\alpha N}(i, i + 4)$ ) can be seen in TFE : water, where a helical structure is present, and only a weak, medium-range signal can be observed for the central residues of the peptide in water, where a short-lived helix is formed in the simulation.

Overall, the observations made by NMR, along with the comparisons made between the ROESY spectra and the simulation heat maps, are consistent with what has been observed in the previously reported CD data<sup>6</sup> and the CHARMM36m simulations discussed in the previous section (Fig. 1 and 2), confirming that the PLP peptide does indeed adopt a more ordered helix in the membrane-mimicking than the aqueous environment.

**Comparison of solvent models and force fields with CD and NMR experiments.** The simulation analyses shown in Fig. 1 and 2 were applied to each force field/solvent combination and are reported in the ESI (Fig. S19–S28†). Table 1 shows a concise

evaluation of the PLP peptide results, with the secondary structure content of each combination taken as a percentage of the DSSP assignment of all the residues in all the equilibrated trajectory frames.

Inspection of the percentage secondary structure content predicted by the ff14SB force field shows that the PLP peptide has a greater helical propensity in chloroform : methanol : water than in water ( $59.7 \pm 0.2\%$  as opposed to  $28.9 \pm 2.1\%$ , respectively) (Table 1). However, the helical content of the ff14SB PLP peptide only very slightly (and with low statistical certainty due to the overlapping error bars) increased in TFE : water ( $34.7 \pm 5.0\%$ ), which does not reflect the structural distinction found between the aqueous and membrane-mimicking solvents in the NMR and CD experiments (Fig. 3).<sup>6</sup> The absence of this distinction in our simulations is likely to be a consequence of the known oversampling of secondary structures by the ff14SB force field, due to the fact that generic protein force fields (including the Amber series) have historically been parameterised using globular protein structures,<sup>44</sup> meaning that our water simulations were unable to sufficiently sample the disordered states.

The ff14IDPSFF force field, which was developed to add a CMAP correction term to ff14SB to remove the bias towards globular structures and increase the sampling of intrinsically disordered structures,<sup>44</sup> performed badly for the PLP peptide in the given solvents (Table 1). A fairly significant  $\beta$ -ladder content was observed for the peptide in water ( $25.4 \pm 3.5\%$ ) and the random coil content actually increased from  $73.2 \pm 2.7\%$  to  $\sim 91.0 \pm 4.0\%$  in the membrane-mimicking solvents, which is

**Table 1** Secondary structure content of the PLP peptide with four force fields in three solvents, taken as a percentage of the DSSP assignment of all the residues in all the equilibrated trajectory frames, with error bars calculated from the two simulations conducted for each force field/solvent combination. For simplicity, the DSSP assignment is categorised into helical elements (H) containing  $\alpha$ -,  $\pi$ - or 3/10-helices (although the majority of the helical sampling in this study was  $\alpha$ -helical),  $\beta$  elements (B) containing isolated  $\beta$ -bridges or full  $\beta$ -ladders, and coil elements (C) containing hydrogen bonded turns, bends, loops and irregular elements. Standard errors are taken between the two simulations for each force field/solvent combination. The structural observations made from the NMR experiments (Fig. 3) and the CD experiments in ref. 6 are included for comparison

% Secondary structure content				
	Water	TFE : water ( $\sim 8 : 2$ vol%)	Chloroform : methanol : water ( $\sim 4 : 4 : 1$ vol% (Amber); $\sim 4.1 : 5.5 : 0.4$ vol% (CHARMM))	Lipid vesicles
ff14SB	H = $28.9 \pm 2.1$ B = $1.3 \pm 0.8$ C = $69.7 \pm 1.3$	H = $34.7 \pm 5.0$ B = $0.6 \pm 0.6$ C = $64.7 \pm 4.4$	H = $59.7 \pm 0.2$ B = $0.0 \pm 0.0$ C = $35.4 \pm 0.2$	—
ff14IDPSFF	H = $1.4 \pm 0.7$ B = $25.4 \pm 3.5$ C = $73.2 \pm 2.7$	H = $1.0 \pm 0.0$ B = $7.7 \pm 3.6$ C = $91.3 \pm 3.7$	H = $0.7 \pm 0.3$ B = $8.0 \pm 3.0$ C = $91.4 \pm 3.4$	—
CHARMM36	H = $7.9 \pm 1.0$ B = $4.0 \pm 0.1$ C = $88.1 \pm 1.0$	H = $58.8 \pm 2.8$ B = $0.0 \pm 0.0$ C = $41.1 \pm 2.8$	H = $64.9 \pm 2.6$ B = $0.0 \pm 0.0$ C = $35.1 \pm 2.6$	—
CHARMM36m	H = $13.3 \pm 0.8$ B = $2.1 \pm 0.5$ C = $84.5 \pm 0.3$	H = $52.9 \pm 0.1$ B = $0.1 \pm 0.0$ C = $47.0 \pm 0.1$	H = $57.9 \pm 2.0$ B = $0.0 \pm 0.0$ C = $42.1 \pm 2.0$	—
NMR (ROESY)	Conformationally heterogeneous (possible presence of helical content)	Ordered helical structure	—	—
CD expt. <sup>6</sup>	Random coil with small amount of helicity	$\alpha$ -helix	—	$\alpha$ -/ $3_{10}$ -helix



inconsistent with the experimental NMR and CD data (Fig. 3).<sup>6</sup> There are no significant helical structures found in any of the simulations. Although ff14IDPSFF was developed to allow greater sampling of IDP structures, our results using ff14SB and ff14IDPSFF in water reveal that the CMAP term actually biased the sampling towards  $\beta$ -ladder structures and the random coil content remained roughly the same (Table 1). A closer look at the ff14IDPSFF parameterisation procedure reveals that parameters were fitted to coil/loop regions of protein crystal structures,<sup>44</sup> which may not be appropriate for the description of standalone intrinsically disordered peptides that do not have a structured protein environment. Furthermore, X-ray crystallography captures static structures that may not be the most appropriate way to model dynamic coil and loop regions of proteins. Static crystal structures of dynamic protein regions are often not well defined.<sup>76</sup>

The CHARMM36 and CHARMM36m results were comparable, with both force fields correlating well with the experimental NMR and CD data (Table 1). The secondary structure content in water was mainly random coil ( $88.1 \pm 1.0\%$  for CHARMM36 and  $84.5 \pm 0.3\%$  for CHARMM36m) and mainly helical in TFE : water ( $58.8 \pm 2.8\%$  for CHARMM36 and  $52.9 \pm 0.1\%$  for CHARMM36m) and chloroform : methanol : water ( $64.9 \pm 2.6\%$  for CHARMM36 and  $57.9 \pm 2.0\%$  for CHARMM36m). CHARMM36 was developed as a correction to the CHARMM22/CMAP force field for better protein backbone and side chain sampling,<sup>42</sup> and CHARMM36m provided a further refinement to the CMAP potential for better sampling of IDP structures.<sup>38</sup> This is noticeable in the membrane-mimicking solvents, where there was a slightly greater random coil content predicted by CHARMM36m compared to CHARMM36. Despite this, it is pleasing that CHARMM36m was able to correctly sample disordered structures and weak helicity of the PLP peptide in water and a more-ordered  $\alpha$ -helix in the membrane-mimicking solvents. Additional radius of gyration ( $R(g)$ ) and root mean square deviation (RMSD) analyses of the simulations of the PLP peptide in water and TFE–water with the CHARMM36m force field are provided in Fig. S43 and S44.†

At this stage, it is necessary to note that conformation-dependent replica trapping was observed as an effect of using the REST2 protocol in the membrane-mimicking solvents, and was most noticeable when using the CHARMM force fields. An example of the phenomenon is explained in more detail in the ESI (ESI Section: Replica Trapping Effect; Fig. S41 and S42†) but, briefly, we observed a reduction in the ability of the base replica to travel across the full replica space upon folding of the PLP peptide in the membrane-mimicking solvents. While it is important to consider that this effect reduces the conformational sampling in the simulation, we believe that the simulation procedure still provides a useful approach for the investigation of membrane-associated peptide conformations because the conformational sampling before the system relaxed to a folded structure was sufficient. Furthermore, preliminary simulations revealed that using hydrophobic solvent systems as membrane-mimicking models, rather than phospholipid bilayers, reduced the replica trapping effect and increased conformational sampling. Using REST2 with hydrophobic solvent

models, therefore, provides an effective compromise between membrane-model accuracy and conformational sampling.

In summary, the PLP benchmark study revealed that REST2 can be used with certain force fields and membrane-mimicking solvents to reproduce membrane-associated IDP conformations. ff14IDPSFF had the worst correlation with the NMR and CD experiments, ff14SB behaved reasonably well with the chloroform : methanol : water solvent, but less well in water or TFE : water, and CHARMM36/CHARMM36m showed the best correlation with the experimental data. Considering that conformational prediction simulations of membrane-active peptides are likely to be used as a precursor to simulations that involve lipid bilayers, CHARMM36m is the better choice of force field because both its protein and lipid parameters were tested with the CHARMM modified TIP3P water model.<sup>77</sup> CHARMM36, however, could be problematic for peptide–lipid simulations because its protein parameters were shown to work best with the standard TIP3P model, while the lipid parameters were parameterised using the CHARMM modified TIP3P model,<sup>78</sup> meaning it could be difficult to choose an appropriate water model for simulations that include both peptides and lipids.

For these reasons, we chose to use the ff14SB and CHARMM36m force fields to study TP2 and ONEG (discussed in the next section).

## TP2 and ONEG

Having established a satisfactory protocol for investigating the environment-specific effects on the structural dynamics of the disordered PLP peptide, we next extend our investigation to a CPP called TP2.<sup>46</sup> TP2 was identified in a high-throughput screen (HTS) as a spontaneous membrane translocating peptide (SMTP) that penetrates through artificial and cell membranes.<sup>46,79</sup> We also investigate ONEG, a peptide that has a sequence similar to that of TP2 and yet was found to not penetrate through artificial or cell membranes and this will serve as a negative control.<sup>47</sup> We hope that the structural differences between the two peptides in aqueous and membrane-mimicking environments could help identify the properties that are important for SMTP activity.

The secondary structure of TP2 has previously been characterised in aqueous buffer and in the presence of POPC vesicles by CD, which showed a transition from mainly random coil in water to an unknown secondary structure as the peptide interacts with POPC.<sup>47</sup> Oriented-CD (OCD) of TP2 in a multi-bilayer POPC stack has also been performed and resulted in the observation of a  $\beta$ -sheet-like spectrum.<sup>47</sup> However, the authors of the characterisation study noted that TP2 binds to and translocates membranes in a concentration-independent manner, meaning that it is likely that TP2 acts as a monomer and is unlikely to form  $\beta$ -sheet oligomers during translocation.<sup>47</sup> This raises uncertainty as to the functional structure of TP2, which is why we chose to study it using MD simulations; in contrast, ONEG has been characterised as a random coil in both aqueous and membrane environments.<sup>47</sup>

**Comparison of solvent models and force fields with CD experiments.** REST2 was performed using the ff14SB and



CHARMM36m force fields in all three solvent systems (water, chloroform : methanol : water and TFE : water) for TP2 and ONEG (Table S1†). Full analyses of the results can be found in the ESI (Fig. S29–S40†), while a concise evaluation of the secondary structure content is given in Tables 2 and 3 for TP2 and ONEG respectively.

For TP2, ff14SB predicted a strong helical content in water ( $47.4 \pm 0.6\%$ ) that actually decreased in TFE : water ( $23.7 \pm 2.1\%$ ) and only increased very slightly in chloroform : methanol : water ( $51.9 \pm 1.6\%$ ) (Table 2). This is inconsistent with the CD data that shows that the peptide is mainly a random coil in water but gets increasingly structured as it interacts with the membrane.<sup>47</sup> No significant  $\beta$  structures were found with the ff14SB force field in any of the solvents, and this is not in agreement with the OCD data.<sup>47</sup>

The CHARMM36m force field predicted a mainly random coil structure in water ( $88.1 \pm 0.0\%$ ) with a very small amount of helical content ( $11.3 \pm 0.1\%$ ) and vanishingly small  $\beta$  content ( $0.6 \pm 0.1\%$ ); this provided a better correlation with the TP2 experiments (Table 2).<sup>47</sup> The helical structure increased to  $22.3 \pm 10.7\%$  in TFE : water and  $32.2 \pm 22.1\%$  in chloroform : methanol : water, which agrees with the observed increase in secondary structure as the peptide interacts with POPC in the CD study.<sup>47</sup> However, the  $\beta$  content did not increase in either of the membrane-mimicking solvents, which does not agree with the  $\beta$ -sheet-like structure found by OCD.<sup>47</sup> A discussion of this contradiction and what it might suggest for the translocation structure of TP2 is included later.

As can be seen in Table 2, the standard errors for the structural percentages of the CHARMM36m TP2 are large, which is due to the large differences in conformational sampling that is seen between the two independent simulations for these systems. This can be seen in Fig. S32–S34,† where it is obvious that the structural content of TP2 continues to change throughout the 1  $\mu$ s simulations, suggesting that the simulations are not yet converged. With computational expense in

mind, the decision was taken not to continue any of the simulations past 1  $\mu$ s and, where conformational convergence could not be established, to set a standard equilibration cutoff of 100 ns. Despite the poor convergence and large error bars, it can be noted that the membrane-mimicking solvents stabilise the helical structure of the CHARMM36m TP2, as can be seen from the slower unfolding of the helix in Fig. S33B and S34B, compared to that in Fig. S32B.†

Comparing the results for TP2 with those for the PLP peptide highlights how different force fields appear to be suited to different biological systems. For example, although it is likely that Amber14SB overpredicted the PLP peptide helicity in water, an obvious increase in helicity occurred as the solvent changed to chloroform : methanol : water (Table 1). However, the same force field predicted similar helical contents for TP2 in both solvents. The ability of the force field to accurately predict a structural transition of one peptide but not another is a reminder that there is not a “best” generic force field, at least for biopolymers.

The results for ONEG (Table 3) appear to be less force-field-dependent and more closely aligned with the experimental CD data than for TP2. As expected, the peptide is mainly a random coil with both force fields in all three solvents. Despite the prediction of a small amount of helicity in water with the ff14SB force field ( $8.0 \pm 2.10\%$ ), both ff14SB and CHARMM36m were successful in reproducing the experimental differences showing that ONEG is less structured than TP2 in aqueous and membrane-mimicking environments.<sup>47</sup> Additional  $R(g)$  and RMSD analyses of the simulations of TP2 and ONEG in water and TFE–water with the CHARMM36m force field are provided in Fig. S45–S48.†

**Comments on the TP2 structure.** It is already known experimentally that TP2 can penetrate through artificial and cellular membranes in a concentration-independent manner and without causing significant membrane disruption.<sup>46</sup> This suggests that TP2 interacts favourably with the membrane as a monomer and

**Table 2** Secondary structure content of TP2 with two force fields in three solvents, taken as a percentage of the DSSP assignment of all the residues in all the equilibrated trajectory frames, with error bars calculated from the two simulations conducted for each force field/solvent combination. For simplicity, the DSSP assignment is categorised into helical elements (H) containing  $\alpha$ -,  $\pi$ - or 3/10-helices (although the majority of the helical sampling in this study was  $\alpha$ -helical),  $\beta$  elements (B) containing isolated  $\beta$ -bridges or full  $\beta$ -ladders, and coil elements (C) containing hydrogen bonded turns, bends, loops or irregular elements. Standard errors are taken between the two simulations for each force field/solvent combination. The experimental CD assignment taken from ref. 47 is included for comparison

% Secondary structure content				
	Water	TFE : water (~8 : 2 vol%)	Chloroform : methanol : water (~4 : 4 : 1 vol% (Amber); ~4.1 : 5.5 : 0.4 vol% (CHARMM))	POPC
Amber14SB	H = $47.4 \pm 0.6$ B = $0.7 \pm 0.1$ C = $51.9 \pm 0.7$	H = $23.7 \pm 2.1$ B = $1.4 \pm 0.8$ C = $74.9 \pm 1.3$	H = $51.9 \pm 1.6$ B = $0.4 \pm 0.4$ C = $47.7 \pm 1.2$	—
CHARMM36m	H = $11.3 \pm 0.1$ B = $0.6 \pm 0.1$ C = $88.1 \pm 0.0$	H = $22.3 \pm 10.7$ B = $0.1 \pm 0.1$ C = $77.6 \pm 10.7$	H = $32.2 \pm 22.1$ B = $0.5 \pm 0.4$ C = $67.3 \pm 21.6$	—
CD expt. <sup>47</sup>	Random coil with “small amount of unidentifiable secondary structure”	—	—	Increased secondary structure (CD) $\beta$ -sheet-like (OCD)



**Table 3** Secondary structure content of ONEG with two force fields in three solvents, taken as a percentage of the DSSP assignment of all the residues in all the equilibrated trajectory frames, with error bars calculated from the two simulations conducted for each force field/solvent combination. For simplicity, the DSSP assignment is categorised into helical elements (H) containing  $\alpha$ -,  $\pi$ - or 3/10-helices (although the majority of the helical sampling in this study was  $\alpha$ -helical),  $\beta$  elements (B) containing isolated  $\beta$ -bridges or full  $\beta$ -ladders, and coil elements (C) containing hydrogen bonded turns, bends, loops or irregular elements. Standard errors are taken between the two simulations for each force field/solvent combination. The experimental CD assignment taken from ref. 47 is included for comparison

	% Secondary structure content			POPC
	Water	TFE : water ( $\sim 8 : 2$ vol%)	Chloroform : methanol : water ( $\sim 4 : 4 : 1$ vol% (Amber); $\sim 4.1 : 5.5 : 0.4$ vol% (CHARMM))	
Amber14SB	H = $8.0 \pm 2.1$ B = $2.2 \pm 0.3$ C = $89.8 \pm 1.8$	H = $2.4 \pm 1.0$ B = $0.7 \pm 0.4$ C = $97.0 \pm 1.4$	H = $2.8 \pm 1.1$ B = $0.1 \pm 0.0$ C = $97.2 \pm 1.0$	—
CHARMM36m	H = $0.4 \pm 0.2$ B = $0.2 \pm 0.0$ C = $99.4 \pm 0.2$	H = $0.4 \pm 0.3$ B = $0.0 \pm 0.0$ C = $99.5 \pm 0.3$	H = $3.0 \pm 0.7$ B = $0.2 \pm 0.2$ C = $96.8 \pm 0.9$	—
CD expt. <sup>47</sup>	Random coil	—	—	Random coil

partakes in a monomeric translocation pathway. The  $\beta$ -sheet-like spectrum of TP2 in POPC identified by OCD<sup>47</sup> would suggest that the monomeric structure is a  $\beta$ -hairpin but we were unable to identify any significant  $\beta$ -hairpin structures in our simulations. In contrast, we consistently observed  $\alpha$ -helical structures in the membrane-mimicking solvents.

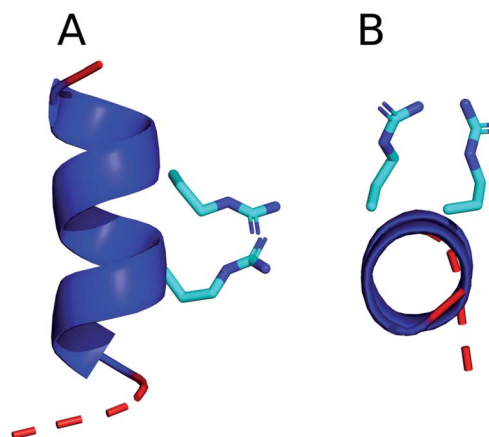
Our helical observation is in agreement with implicit membrane simulations performed by Lazaridis *et al.* on TP2, which revealed that the helical conformation of the peptide in the membrane was characterised by a lower energy compared to an extended structure.<sup>80</sup> They did, however, identify a low energy  $\beta$ -hairpin at the membrane interface and suggested that TP2 flips from a  $\beta$ -hairpin at the interface to an  $\alpha$ -helix inside the membrane.<sup>80</sup> It is possible that the  $\beta$ -sheet-like OCD spectrum of TP2<sup>47</sup> is actually a measurement of the peptide at the membrane surface.

Alternatively, the  $\beta$ -sheet-like spectrum could result from unexpected TP2 aggregation in the OCD experiment. Macchi *et al.* investigated the effect of covalent cargo on the self-aggregation of TP2, showing that the critical micelle concentration (CMC) (the concentration above which aggregation occurs) increases when the fluorescent dye, TAMRA, is attached.<sup>81</sup> TAMRA was covalently attached to TP2 in the translocation assays<sup>46</sup> but not in the OCD experiments,<sup>47</sup> so it could be possible that a monomeric TP2-TAMRA complex translocated in the cell penetration assays but a  $\beta$ -sheet TP2 oligomer formed in the OCD experiment. If this is the case, our monomeric simulations are relevant for TP2 below the CMC and it would be interesting to extend our simulation protocol to study peptide aggregation in the future.

Our hypothesis that TP2 adopts a helix inside the membrane gains further confidence because the peptide contains a specific motif made up of hydrophobic ( $\Phi$ ) and arginine (R) residues ( $\Phi$ R $\Phi$  $\Phi$ R) that is also present in the S4 helix of the voltage-gated potassium channel, K<sub>v</sub>AP.<sup>82</sup> Despite containing many positive charges, the S4 helix has been shown to move across the hydrophobic membrane during channel gating and to act as an SMTP as a stand-alone peptide.<sup>82</sup> Studies have revealed the importance of this specific arrangement of hydrophobic and

arginine residues in inducing SMTP behaviour,<sup>83</sup> so it is reasonable to assume that TP2 has a similar structure and membrane-translocating mechanism as the S4 helix.

Furthermore, upon inspection of some of the TP2 structures identified in this study, it is obvious that the alignment of the two arginines in the  $\alpha$ -helix allows the guanidinium groups to form a pair (Fig. 4). Guanidinium pairs are thermodynamically stable<sup>84</sup> and are predicted to improve CPP activity because they result in non-additive membrane disruption, *i.e.* the overall water defect caused by a guanidinium pair moving into a membrane is less than that caused by two separate guanidinium groups.<sup>84</sup> Additionally, the other TP2 residues are mainly hydrophobic, suggesting that the peptide amphipathicity is increased with the two positively charged residues pointing in the same direction, and this is known to be an important property for CPP activity.<sup>85</sup> These observations support the theory that the TP2 monomer translocates as a helix.



**Fig. 4** A structural cluster identified in the CHARMM36m TP2 simulation in chloroform : methanol : water, viewed from (A) the side of the helix and (B) down the axis of the helix. The backbone is represented as a cartoon with coil residues shown in red and  $\alpha$ -helical residues shown in dark blue. The arginine side chains are shown in stick representation using the standard CPK colouring.





In contrast, the simulation results suggest that ONEG does not form a stable secondary structure in the membrane, which is in agreement with the experimental CD and OCD data.<sup>47</sup> The absence of a helical structure in ONEG could result from the presence of the central proline residue that disrupts the ability of the backbone to partake in ordered hydrogen bonding. The disordered structure of ONEG in the membrane would leave its polar backbone atoms open to the hydrophobic environment, which is unfavourable. Additionally, the inability of ONEG to form amphipathic secondary structures reduces its ability to orient itself to interact favourably with both the hydrophilic and hydrophobic parts of the membrane. Furthermore, ONEG has three arginine residues, as opposed to the two arginines present in TP2, and its disordered structure reduces its ability to form guanidinium pairs. To further test the hypothesis that helical structure is important for SMTP activity, it would be interesting to study the structures of the different hydrophobic/arginine motifs tested in ref. 83 to observe whether changing the arrangement of residues affects their ability to form stable, amphipathic helical structures and/or guanidinium pairs.

## Conclusion

IDPs are an important part of the proteome and are paramount in a range of biological processes including PPIs, the multimerisation of membrane proteins, amyloid aggregation, cell penetration and the disruption of bacterial membranes. The unique flexible nature of IDPs sensitises their conformational ensembles to the biological environment, enabling them to easily assume different functional conformations and bind to different macromolecules. For PPIs, this may involve the folding of an intrinsically disordered peptide upon binding to another protein, or for membrane-active peptides (such as membrane proteins, CPPs or AMPs), upon penetrating into the membrane. In both cases, the IDPs transition from an aqueous to a hydrophobic environment, resulting in a different conformational ensemble that allows them to perform their biological function.

MD is often used to study biological systems but it is challenging to simulate the conformational transitions of IDPs because of the long timescales required to sample all of the relevant disordered conformational space and to overcome the free energy barriers associated with peptide folding. In this paper, we tested the ability of four force fields to predict the conformational ensembles of aqueous and membrane-associated peptides, using the enhanced sampling method REST2 in aqueous and membrane-mimicking solvents, and conducted NMR experiments for the PLP peptide in two solvent systems to add information to be compared with the data obtained from the simulation protocol. There is also experimental precedent for using hydrophobic solvents to mimic the low polarity of protein binding sites,<sup>4,71</sup> suggesting that our simulation protocol could be applicable for the conformational study of IDPs involved in PPIs. Furthermore, it would be interesting to use the protocol to simulate the conformational ensembles of cell-permeable macrocyclic drugs, which are also known to change their conformations as they pass through the hydrophobic cell membrane.<sup>86</sup>

We chose to benchmark the simulation protocol and force fields using the PLP peptide, which has relevant experimental CD data reported in the literature,<sup>6</sup> and for which we produced additional experimental data with NMR. CD experiments have previously indicated that the PLP peptide is intrinsically disordered in the cytoplasm and binds to the membrane as an  $\alpha$ - or  $3_{10}$ -helix.<sup>6</sup> The NMR results presented in this paper suggest a more ordered helix in TFE/H<sub>2</sub>O than in H<sub>2</sub>O/D<sub>2</sub>O, which further validates the hypothesis that the peptide adopts a helical structure upon binding to the membrane. Out of the force fields tested in this study, we found that the CHARMM force fields were best able to capture the transition from disordered structures in water to helical structures in the membrane-mimicking solvents. It is likely that the ff14SB force field overstabilised the helix in water, however it was successful in capturing the transition to a more helical structure in chloroform : methanol : water, and the ff14IDPSFF force field was unable to predict a helical structure in the membrane-mimicking solvents. It would also be interesting to extend the simulation protocol to include multiple copies of the peptide in order to test the ability of the force fields to capture the transition from helical monomers to  $\beta$ -sheet oligomers observed in the CD experiments.<sup>6</sup>

Following the validation study on the PLP peptide, we applied the simulation protocol to TP2, an SMTP that is found to penetrate through artificial and cellular membranes at low concentrations. Our simulations led to an  $\alpha$ -helical assignment of the peptide in the membrane, which contradicts the  $\beta$ -sheet-like OCD spectrum reported in the literature.<sup>47</sup> We have included a discussion of this contradiction and have suggested mechanistic reasoning to support the possible  $\alpha$ -helical assignment of TP2. Our simulations also predicted that the negative control peptide, ONEG, does not form a significant secondary structure, which could explain the difference in SMTP ability between the two peptides.

In conclusion, the results presented in this paper show that, with enhanced sampling and the right force fields, MD is a useful tool to study the conformational transitions of IDPs in different biological environments. We recommend the use of the CHARMM36m force field with the water, TFE : water or chloroform : methanol : water solvent models used in this study for the investigation of IDP conformations in aqueous and hydrophobic environments. Although we have observed conformation-dependent replica trapping with the use of the REST2 protocol in membrane-mimicking solvents, we note the proficiency of the method in enhancing the conformational sampling of IDPs without having to provide user-defined collective variables that may bias the simulation towards preconceived ideas of the peptide structure. Although we have tested the protocol on membrane-active peptides, the possible scope of this approach could extend to any IDP that undergoes environment-dependent structural transitions, and we hope to utilise this in the future to increase our understanding of the interesting mechanisms of IDPs.

## Data availability

Data for this paper, including simulation trajectory datasets, are available at Zenodo at <https://doi.org/10.5281/zenodo.5830484>.



## Author contributions

Lauren Reid: Investigation, Formal Analysis, Writing – original draft; Ileana Guzzetti: Investigation, Formal Analysis, Writing – original draft; Tor Svensson: Investigation; Anna-Carin Carlsson: Investigation; Wu Su: Conceptualisation; Supervision; Writing – review & editing; Tomas Leek: Conceptualization; Supervision; Writing – review & editing; Lena von Sydow: Investigation; Werngard Czechtizky: Funding acquisition; Conceptualisation; Writing – review & editing; Marija Miljak: Software; Chandra Verma: Supervision, Funding acquisition, Conceptualisation, Project Administration; Leonardo De Maria: Conceptualisation; Project Administration; Supervision; Writing – review & editing; Jonathan W. Essex: Supervision, Funding acquisition, Conceptualisation. Project Administration, Writing – review & editing.

## Conflicts of interest

JWE's research is part funded by AstraZeneca. LMR is currently employed by MedChemica Ltd.

## Acknowledgements

The authors would like to acknowledge the University of Southampton and A\*STAR for PhD funding, as well as the IRI-DIS High Performance Computing Facility, and associated support services at the University of Southampton, and the National Supercomputing Centre (NSCC) Singapore for computational support. We also thank Stephen Fox, Pietro Aronica, Thomas Piggot and Richard Bradshaw for advice and guidance. We also thank Richard Lewis for helpful discussions regarding NMR spectra.

## References

- V. N. Uversky, *Front. Physiol.*, 2019, **7**, 10.
- V. N. Uversky, *J. Biomed. Biotechnol.*, 2010, **2010**, 14.
- S. Yadahalli, J. Li, D. P. Lane, S. Gosavi and C. S. Verma, *Sci. Rep.*, 2017, **7**, 15600.
- A. Emendato, R. Guerrini, E. Marzola, H. Wienk, R. Boelens, S. Leone and D. Picone, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **5**, 100.
- Z. A. Levine, L. Larini, N. E. LaPointe, S. C. Feinstein and J. E. Shea, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 2758–2763.
- A. Appadu, M. Jelokhani-Niaraki and L. DeBruin, *J. Phys. Chem. B*, 2015, **119**, 14821–14830.
- A. Hafner, M. L. Bulyk, A. Jambhekar and G. Lahav, *Nat. Rev. Mol. Cell Biol.*, 2019, **20**, 199–210.
- F. Mantovani, L. Collavin and G. Del Sal, *Cell Death Differ.*, 2019, **26**, 199–212.
- H. Feng, L. M. M. Jenkins, S. R. Durell, R. Hayashi, S. J. Mazur, S. Cherry, J. E. Tropea, M. Miller, A. Wlodawer, E. Appella and Y. Bai, *Structure*, 2009, **17**, 202–210.
- V. N. Uversky, *Int. J. Mol. Sci.*, 2016, **17**, 1874.
- P. Kulkarni and V. N. Uversky, *Biomolecules*, 2019, **9**, 147.
- X. Zhang, Z. Fu, L. Meng, M. He and Z. Zhang, *Front. Aging Neurosci.*, 2018, **10**, 359.
- A. Kumari, R. Rajput, N. Shrivastava, P. Somvanshi and A. Grover, *Int. J. Biochem. Cell Biol.*, 2018, **99**, 19–27.
- E. Eiríksdóttir, K. Konate, Ü. Langel, G. Divita and S. Deshayes, *Biochim. Biophys. Acta Biomembr.*, 2010, **1798**, 1119–1128.
- S. J. Fox, R. Lakshminarayanan, R. W. Beuerman, J. Li and C. S. Verma, *J. Phys. Chem. B*, 2018, **122**, 8698–8705.
- M. Gallo, S. Defaus and D. Andreu, *Arch. Biochem. Biophys.*, 2019, **661**, 74–86.
- J. Lei, L. C. Sun, S. Huang, C. Zhu, P. Li, J. He, V. Mackey, D. H. Coy and Q. Y. He, *Am. J. Transl. Res.*, 2019, **11**, 3919–3931.
- A. Borrelli, A. Tornesello, M. Tornesello, F. Buonaguro, A. Borrelli, A. L. Tornesello, M. L. Tornesello and F. M. Buonaguro, *Molecules*, 2018, **23**, 295.
- Y. Du, L. Wang, W. Wang, T. Guo, M. Zhang, P. Zhang, Y. Zhang, K. Wu, A. Li, X. Wang, J. He and J. Fan, *J. Biomed. Nanotechnol.*, 2018, **14**, 161–167.
- J. H. Na, W. K. Lee and Y. G. Yu, *Int. J. Mol. Sci.*, 2018, **19**, 381.
- S. Signorelli, S. Cannistraro and A. R. Bizzarri, *Appl. Spectrosc.*, 2017, **71**, 823–832.
- S. Koubaa, A. Bremer, D. K. Hinch and F. Brini, *Sci. Rep.*, 2019, **9**, 1–11.
- P. Bernadó and D. I. Svergun, *Mol. Biosyst.*, 2012, **8**, 151–167.
- K. Gast and C. Fiedler, *Methods Mol. Biol.*, 2012, **896**, 137–161.
- P. G. Argudo and J. J. Giner-Casares, *Nanoscale Adv.*, 2021, **3**, 1789–1812.
- D. J. Huggins, P. C. Biggin, M. A. Dämgen, J. W. Essex, S. A. Harris, R. H. Henchman, S. Khalid, A. Kuzmanic, C. A. Laughton, J. Michel, A. J. Mulholland, E. Rosta, M. S. P. Sansom and M. W. van der Kamp, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2019, **9**, e1393.
- E. Haensele, N. Mele, M. Miljak, C. M. Read, D. C. Whitley, L. Banting, C. Delépée, J. Sopkova-De Oliveira Santos, A. Lepailleur, R. Bureau, J. W. Essex and T. Clark, *J. Chem. Inf. Model.*, 2017, **57**, 298–310.
- L. M. Reid, C. S. Verma and J. W. Essex, *Drug Discov. Today*, 2019, **24**, 1821–1835.
- R. Appadurai, J. Nagesh and A. Srivastava, *Nat. Commun.*, 2021, **12**(12), 1–11.
- U. R. Shrestha, J. C. Smith and L. Petridis, *Commun. Biol.*, 2021, **4**(4), 1–8.
- M. Chan-Yao-Chong, D. Durand and T. Ha-Duong, *J. Chem. Inf. Model.*, 2019, **59**, 1743–1758.
- T. Piggot, Á. Piñeiro and S. Khalid, *J. Chem. Theory Comput.*, 2012, **8**, 4593–4609.
- A. Sandoval-Perez, K. Pluhackova and R. Böckmann, *J. Chem. Theory Comput.*, 2017, **13**, 2310–2321.
- J. Huang and A. D. MacKerell, *Curr. Opin. Struct. Biol.*, 2018, **48**, 40–48.
- J. Henriques, C. Cragnell and M. Skepö, *J. Chem. Theory Comput.*, 2015, **11**, 3420–3431.
- P. Robustelli, S. Piana and D. E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.*, 2018, **115**, E4758–E4766.



- 37 D. Song, W. Wang, W. Ye, D. Ji, R. Luo and H.-F. Chen, *Chem. Biol. Drug Des.*, 2017, **89**, 5–15.
- 38 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. de Groot, H. Grubmüller and A. MacKerell, *Nat. Methods*, 2016, **14**, 71–73.
- 39 R. Best, W. Zheng and J. Mittal, *J. Chem. Theory Comput.*, 2014, **10**, 5113–5124.
- 40 R. B. Best, *Curr. Opin. Struct. Biol.*, 2017, **42**, 147–154.
- 41 M. Buck, S. Bouguet-Bonnet, R. W. Pastor and A. D. MacKerell and Jr., *Biophys. J.*, 2006, **90**, L36.
- 42 J. Huang and A. MacKerell, *J. Comput. Chem.*, 2013, **34**, 2135–2145.
- 43 J. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
- 44 D. Song, R. Luo and H. F. Chen, *J. Chem. Inf. Model.*, 2017, **57**, 1166–1178.
- 45 L. Wang, R. Friesner and B. J. Berne, *J. Phys. Chem. B*, 2011, **115**, 9431–9438.
- 46 J. He, B. Kauffman, T. Fuselier, S. Naveen, T. Voss, K. Hristova and W. Wimley, *J. Biol. Chem.*, 2013, **288**, 29974–29986.
- 47 J. Cruz, M. Mihailescu, G. Wiedman, K. Herman, P. Searson, W. Wimley and K. Hristova, *Biophys. J.*, 2013, **104**, 2419–2428.
- 48 R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2013, **3**, 198–210.
- 49 J. Vymětal and J. Vondrášek, *J. Phys. Chem. B*, 2014, **118**, 10390–10404.
- 50 C. Coleman, P. van Maaren, M. Hong, J. Hub, L. Costa and D. van der Spoel, *J. Chem. Theory Comput.*, 2012, **8**, 61–74.
- 51 N. M. Fischer, P. J. van Maaren, J. C. Ditz, A. Yildirim and D. van der Spoel, *J. Chem. Theory Comput.*, 2015, **11**, 2938–2944.
- 52 W. Dietz and K. Heinzinger, *Berichte der Bunsengesellschaft für physikalische Chemie*, 1985, **89**, 968–977.
- 53 W. Yu, X. He, K. Vanommeslaeghe and A. D. MacKerell, *J. Comput. Chem.*, 2012, **33**, 2451–2468.
- 54 M. Abraham, T. Murtola, R. Schulz, S. Páll, J. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
- 55 G. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.
- 56 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.
- 57 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
- 58 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
- 59 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.
- 60 R. McGibbon, K. Beauchamp, M. Harrigan, C. Klein, J. Swails, C. Hernández, C. Schwantes, L.-P. Wang, T. Lane and V. Pande, *Biophys. J.*, 2015, **109**, 1528–1532.
- 61 I. T. Jolliffe and J. Cadima, *Philos. Trans. R. Soc., A*, 2016, **374**, 20150202.
- 62 M. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz and F. Noé, *J. Chem. Theory Comput.*, 2015, **11**, 5525–5542.
- 63 D. Salt, B. Hudson, L. Banting, M. Ellis and M. Ford, *J. Med. Chem.*, 2005, **48**, 3214–3220.
- 64 G. G. Kordopati, H. Tzoupis, A. N. Troganis, G. M. Tsigoulis, S. Golic Grdadolnik, C. Simal and T. V. Tselios, *J. Comput. Mol. Des.*, 2017, **31**, 841–854.
- 65 D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley and B. D. Sykes, *J. Biomol. NMR*, 1995, **6**, 135–140.
- 66 J. M. Greer and M. B. Lees, *Int. J. Biochem. Cell Biol.*, 2002, **34**, 211–215.
- 67 D. Warschawski, A. Arnold, M. Beaugrand, A. Gravel, É. Chartrand and I. Marcotte, *Biochim. Biophys. Acta Biomembr.*, 2011, **1808**, 1957–1974.
- 68 F. Penin, V. Brass, N. Appel, S. Ramboarina, R. Montserret, D. Ficheux, H. E. Blum, R. Bartenschlager and D. Moradpour, *J. Biol. Chem.*, 2004, **279**, 40835–40843.
- 69 O. Y. Dmitriev, K. Altendorf and R. H. Fillingame, *FEBS Lett.*, 2004, **556**, 35–38.
- 70 B. Chaubey, A. Dey, A. Banerjee, N. Chandrakumar and S. Pal, *J. Phys. Chem. B*, 2020, **124**, 5993–6003.
- 71 M. G. Bomar, S. J. Samuelsson, P. Kibler, K. Kodukula and A. K. Galande, *Biomacromolecules*, 2012, **13**, 579–583.
- 72 C. Breil, M. Abert Vian, T. Zemb, W. Kunz and F. Chemat, *Int. J. Mol. Sci.*, 2017, **18**, 708.
- 73 K. Wüthrich, *Eur. Newsl.*, 1986, **17**, 11–13.
- 74 G. Merutka, H. Jane Dyson and P. E. Wright, *J. Biomol. NMR*, 1995, **5**, 14–24.
- 75 T. Cierpicki and J. Otlewski, *J. Biomol. NMR*, 2001, **21**, 249–261.
- 76 M. V. Petoukhov, N. A. J. Eady, K. A. Brown and D. I. Svergun, *Biophys. J.*, 2002, **83**, 3113–3125.
- 77 W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey and M. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 78 S. Boonstra, P. Onck and E. van der Giessen, *J. Phys. Chem. B*, 2016, **120**, 3692–3698.
- 79 J. Marks, J. Placone, K. Hristova and W. Wimley, *J. Am. Chem. Soc.*, 2011, **133**, 8995–9004.
- 80 T. Lazaridis, J. Leveritt and L. PeBenito, *Biochim. Biophys. Acta Biomembr.*, 2014, **1838**, 2149–2159.
- 81 S. Macchi, G. Signore, C. Boccardi, C. Di Rienzo, F. Beltram and F. Cardarelli, *Sci. Rep.*, 2015, **5**, 16914.
- 82 J. He, K. Hristova and W. Wimley, *Angew. Chem., Int. Ed.*, 2012, **51**, 7150–7153.
- 83 T. Fuselier and W. Wimley, *Biophys. J.*, 2017, **113**, 835–846.
- 84 C. Allolio, K. Baxova, M. Vazdar and P. Jungwirth, *J. Phys. Chem. B*, 2015, **120**, 143–153.
- 85 B. Kauffman, T. Fuselier, J. He and W. Wimley, *Trends Biochem. Sci.*, 2015, **40**, 749–764.
- 86 E. Danelius, V. Poongavanam, S. Peintner, L. H. E. Wiese, M. Erdélyi and J. Kihlberg, *Chem.-Eur. J.*, 2020, **26**, 5231–5244.

