

Proline and water stabilization of a universal two-step folding mechanism for β -turn formation in solution

Nicola Steinke,[†] Anna Genina,[‡] Richard J. Gillams,^{†,¶} Christian D. Lorenz,^{*,‡} and Sylvia E. McLain^{*,†}

[†]*Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK*

[‡]*Department of Physics, King's College London, London, WC2R 2LS, UK*

[¶]*Current address: Structural Genomics Consortium, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7DQ, UK*

E-mail: chris.lorenz@kcl.ac.uk; sylvia.mclain@bioch.ox.ac.uk

Abstract

The atomic scale process by which proteins fold into their functional forms in aqueous solutions is still not well understood. While there is clearly an interplay between specific sequence of the protein and the surrounding water solvent that leads to highly-specific and reproducible folding in nature, there is still an ongoing debate concerning how water molecules aid in driving protein folding. By using a combination of techniques which provide information at the atomic level - neutron and X-ray diffraction and computer simulations - the mechanism of folding in a series of peptides which only vary with respect to the central side-chain residue has been determined. Specifically, β -turn formation for the KGXGK peptide (with X=P,G,S or L) occurs via a two-step water-driven attraction between specific sites on the peptide backbone. This proposed

mechanism suggests both that the site-specific hydration of the backbone facilitates the initial stages of protein folding and that this hydration interaction in combination with the presence of a proline residue in the $i + 1$ position helps to stabilize the folded and intermediate folding state of the peptide in aqueous solutions leading to a greater propensity for PG containing sequences to occur in β -turns in protein structures.

Introduction

Proteins are the essential working components in all living organisms. Their biological function is intrinsically linked to the specific three-dimensional structure the amino acid polymer adopts during folding. Although a universal mechanism by which proteins fold remains unknown,¹ it is clear that water plays an integral role in this process as folding occurs in the presence of water. There are two conflicting theories as to how hydration facilitates folding *in vivo*. One theory focuses on the entropically driven expulsion of water from the hydrophobic core as the fundamental phenomenon for this process.²⁻⁵ In contrast, more recent theories emphasize the importance of hydrophilic interactions - such as hydrogen bond formation and water-bridging interactions⁶⁻¹³ and that folding occurs prior to water being eliminated from the hydrophobic core of a protein.¹⁴

Turns are small secondary structure motifs that change the direction of the polypeptide chain and it has previously been suggested that these folding sites may initiate the formation of tertiary protein structure in solution.^{7,15-17} β -turns are four residue motifs. They are the most common type of turns with a direct hydrogen bond between the carbonyl group of the first residue and the amide group of the fourth residue ($i \rightarrow i + 3$), where proline and glycine residues are commonly found within β -turn sequences at the $i + 1$ and $i + 2$ positions, respectively.^{7,18-22} The sequential presence of these amino acids have been shown to stabilize β -turns,^{23,24} and the XPGX' sequence is found 20467 times in the UniProt Protein Database,²⁵ 41% in β -turns. While other sequences such as XLGX', XSGX' or XGGX' are found less often as β -turns among the deposited structures in the same database (<30%).

Both glycine and proline are somewhat unique among the amino acids, as glycine can easily adopt a wider range of possible dihedral angles due to its small side chain and proline has a relatively restricted ϕ angle. However, it is unclear how these conformational properties influence the process of folding and if water plays an active role in the initiation of β -turn formation.

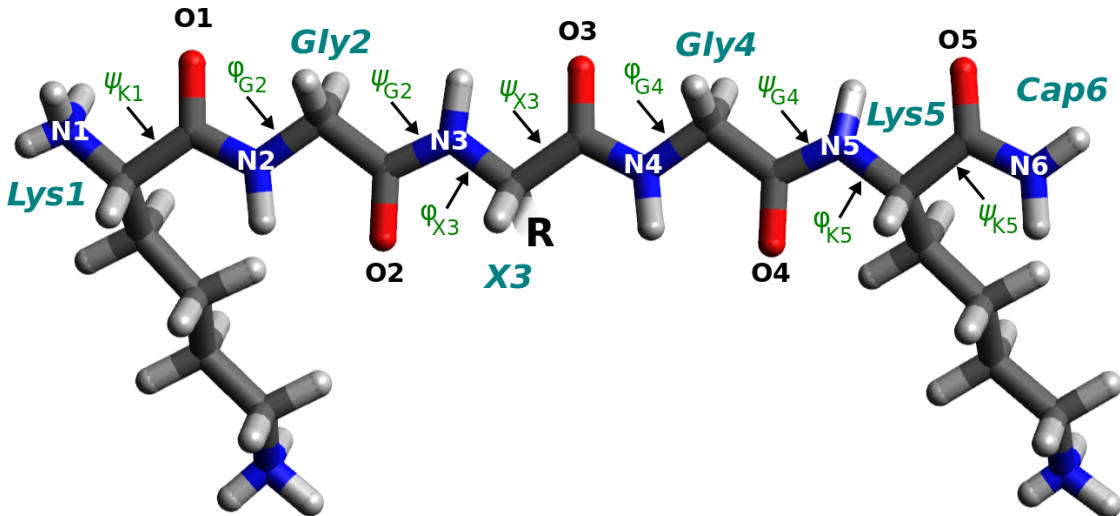


Figure 1 The molecular structure and salient atomic labels for the KGPGK, KGGGK, KGSGK, KGLGK peptides. The labeled bond refers to the respective rotational axes of the dihedral angles. The varied residue of the third amino acid is $R = H$ (for Gly3), $R = C_3H_6$ (for Pro3, -H on N3), $R = CH_2OH$ (for Ser3), $R = CH_2CH(CH_3)_2$ (for Leu3).

One difficulties in investigating the role that solvation plays in the protein folding process is due to the fact that naturally occurring proteins fold rapidly on the sub-second time scale.^{26,27} As a result, it is difficult to observe proteins in the process of folding, especially in the initial stages, by many experimental techniques. Further, the role of water in these processes is particularly elusive as many measurements are relatively blind to the interactions between water and peptides on the atomic scale where these interactions necessarily occur *in vivo*. The reason for this is that many structural techniques do not directly probe hydration and this involves hydrogen bonding, which is difficult to probe in solution by many experimental techniques and as such the structure must be inferred from dynamical data. Secondly, in solution, unlike what occurs in crystalline systems, individual water molecules

can move and out of their specific positions during the folding process and as such structural interactions, as measured by experiment, can be difficult to assess.

In the current work, the interplay between conformation and hydration in a series of KGXGK peptides (lysine-glycine-X-glycine-lysine-NH₂; X=P,L,G or S (Fig. 1)) in aqueous solution has been investigated on the atomic scale in order to determine how both sequence and the hydration of a specific sequence aid in the nucleation of β -turn formation. These peptides have been specifically chosen because in solution they will not be fully folded. Instead they will exist in a variety of folded states allowing for the observation of these peptides in the *process* of folding, so that the role of water in the process of folding can be probed directly. Further the variable central side chains have been chosen so that proline, which occurs most often in β -turns can be compared with a non-polar (L), a polar (S) and a highly flexible (G) residue. Finally, the lysine residues were chosen in order to add more solubility and flexibility: solubility as the diffraction experiments require a relatively high concentration of peptide to ensure that the scattering of the peptide itself is visible and flexibility in order to observe intermediate peptide conformations in the process of folding.

The atomic scale details of the hydration and how this is coupled to folding in KGPGK has been probed using a combination of diffraction measurements and computational techniques: both molecular dynamics (MD) and Empirical Potential Structure Refinement (EPSR);²⁸ a Monte Carlo-based simulation that is constrained by a set of neutron and X-ray diffraction data. The combination of diffraction experiments and computation provides experimentally consistent models of structure and hydration of the different KGPGK conformations in solution. The structure and hydration of the other KGXGK peptides have been assessed using MD. The use of MD is particularly useful in the present work as it allows for an understanding not only of the structure but also for a more direct assessment of the folding process for all of the peptides investigated.

Materials and Methods

Diffraction measurements and Empirical Potential Structure Refinement

Neutron diffraction with isotopic substitution is the premier technique for determining the structure and hydration of small molecules in a variety of solutions.^{7,9,29–38} Importantly, it is one of the few techniques that can give direct structural assessment of hydrogen bonding on the atomic length scale in solution due to the large difference in scattering from hydrogen and deuterium labeled samples.³⁹ This allows for the measurement of a set of isotopically unique yet chemically identical samples, where the hydrogen-containing correlations are weighted differently in each diffraction pattern.⁴⁰ The quantity obtained in a diffraction experiment, after relevant corrections is the so-called static structure factor ($F(Q)$) viz. $F(Q) = \sum_{\alpha, \beta \geq \alpha} (2 - \delta_{\alpha\beta}) c_\alpha c_\beta b_\alpha b_\beta (S_{\alpha\beta}(Q) - 1)$ where c_i and b_i are the relative concentration and scattering length of atom i , α or β , $\delta_{\alpha\beta}$ is the Kronecker delta function, Q is the scattering vector, $Q = (4\pi \cdot \sin \theta) / \lambda$ with neutron wavelength λ and scattering angle 2θ . $S_{\alpha\beta}(Q)$ are the individual partial structure factors for each atomic pair within the measured solution. These are related to the atomic distances in real space, the radial distribution functions (RDFs; $g_{\alpha\beta}(r)$ s), by Fourier transformation *via* $S_{\alpha\beta}(Q) = 1 + \frac{4\pi\rho}{Q} \int r \cdot (g_{\alpha\beta}(r) - 1) \cdot \sin(Qr) dr$. Neutron diffraction data from 5 isotopically unique samples and a complimentary X-ray diffraction pattern for the peptide KGPGK-NH₂ in aqueous solutions were collected at the ISIS Neutron and Muon Source (STFC, UK) and analyzed as previously described.³⁰ For the neutron diffraction experiments, each solution contained the same molecular ratio of KGPGK:water and varied only with respect to the H/D ratio in the solutions. Specifically each solution was prepared, by weight in a ratio of KGPGK⁺³:3TFA⁻:346H₂O. For the X-ray diffraction experiments the fully protiated sample was measured at the same molecular ratio as for the neutron diffraction experiments. A full description of the neutron diffraction measurements for this work are described elsewhere.³⁰

As stated above, these particular systems were designed to ‘trap’ the peptides in the process of folding in order to assess the structural role water has in this process. As fully functional proteins and some peptides fold on milli to microsecond timescale^{26,27} and as such it would not be possible to assess the different conformational states during folding as folding would occur faster than the experiment. Specifically, KGPGK contains an YPGX sequence which are often found in β -turning motifs in proteins,,^{7,18–22,25} yet also contains terminal lysine side chains which carry a charge in the solution. This not only aids in keeping the molecule in solution at a reasonable concentration for the neutron diffraction measurements but also will minimize the occurrence of fully-folded KGPGK conformations in solution. That there are a range of intermediate folding states for KGPGK in the measured solution is evident from the analysis of the diffraction data below.

Empirical potential structure refinement (EPSR) is a reverse Monte Carlo technique where the atomic conformation of the simulated system is constrained to fit a set of diffraction data. EPSR uses a box of molecules at the same concentration, density and temperature as the experimental diffraction measurements. Starting potentials for each unique atom are refined iteratively until the EPSR-simulated diffraction data shows agreement with the experimental data.^{28,41} For the work presented here, the KGPGK peptide was modeled in a box containing 20 KGPGK molecules, 60 TFA[−] ions and 6920 water molecules at the measured density ($\rho = 0.101$ atoms \AA^{-3}) and temperature of the experiments (298K) as previously described.³⁰ The EPSR simulation contained a mixture of *cis* and *trans* KGPGK molecules (with respect to the Gly-Pro bond) in ratios which correspond with that measured by ¹H NMR (15%*cis*; 85%*trans*). Starting potentials for the KGXGK molecules were taken from the CHARMM36 forcefield,^{42–44} for the TFA[−] ions from the CHARMM General force field^{45,46} and the water molecule starting potentials from the TIP3P model⁴⁷ which has been modified for the CHARMM force field.⁴⁸ The same potentials were used for the MD simulations and are listed in the SI. The peptide starting conformation was identical to the KGPGK simulation in MD.

Molecular dynamics

MD simulations of KGPGK in solution were performed at the same molecular ratios as the diffraction measurements and EPSR simulations. Each system contained 20 KGXGK-NH₂ molecules (with *trans* G-P peptide bonds), 60 TFA⁻ counter ions and 6920 water molecules. All of the bonds and angles for the water molecules were constrained using the SHAKE algorithm⁴⁹ and the simulations were conducted using GROMACS 4.⁵⁰ MD simulations were also performed on three further peptides - KGGGK, KGSGK and KGLGK (each with a C-terminal amide cap -NH₂ and as a *trans*-conformer). Each simulation contained 20 peptides, 60 TFA molecules and 6920 water molecules using the same force fields as for KGPGK.

Each of the 4 MD simulation boxes were constructed using the Packmol software,⁵¹ and for each, in order to eliminate atomic overlaps, an energy minimization simulation was used. Subsequently, a 2 ns simulation using the NVT ensemble with a target temperature of 300 K was performed for temperature equilibration, followed by a 2 ns NPT simulation at 300K and 1 atm, for a volume and pressure equilibration. A NPT production simulation was performed at 300 K and 1 atm for 50 ns with a time step of 2 fs for each of the four peptides. For all simulations, the Nose-Hoover thermostat^{52,53} was used to control the temperature and the Martyna-Tuckerman-Tobias-Klein (MTTK) barostat⁵⁴ was used in the NPT simulations to control the pressure. A cut-off of 14 Å was used for the van-der-Waals interactions, and the long range Coulombic interactions were calculated using the particle mesh Ewald (PME) algorithm.^{55,56}

It should be noted that neutron and X-ray diffraction measures the instantaneous structure of molecules in solution and ‘snapshots’ of these instantaneous structures are collected during the diffraction experiments. As a result the EPSR simulation, which is constrained by the measured data, will depict the average distribution of peptides throughout the simulation but not the transition between the peptides in the process of folding. Conversely, MD will measure the peptides in the process of folding where any one given peptide can be

tracked over the course of the simulation.

In order to evaluate the stability of the ‘folding states’ of each KGXGK peptide, the probability of a peptide of remaining in each of the three identified conformational states, e.g. their lifetimes, as well as the probability that a peptide would transition from one conformational state to another was determined. In order to do this, the conformation that each peptide adopted (based on the $O2 \cdots N5$ distance) was determined for each snapshot of the trajectory. To calculate the *lifetime probabilities*, once a peptide transitioned into a new conformation, the number of consecutive snapshots (separated by 4 ps timesteps) where the peptide remained in its particular conformational category - ‘open’, ‘medium’ or closed - were used to give the lifetime of that peptide in that particular conformation. The distributions of these conformational lifetimes across the entirety of the production simulation for each KGXGK peptide was normalized to the number of times that the peptides in the system adopted a configuration in a given state. Similarly, the transition probabilities from one ‘folding state’ (i) to another (j) were determined by counting the number of times that a peptide is in conformation i in configuration t and then found in conformation j in configuration $t + 1$, and then normalized by the product of the number of configurations in the entire trajectory and to the number of peptides in the system.

ANGULA analysis

Figure 1 shows the labeling of the peptide residues and backbone dihedral angles. 4000 configurations of the EPSR and MD simulations were analyzed using ANGULA,^{57,58} a total of 80,000 KGXGK peptide molecules being analyzed for each system. Specifically, the distribution of the β -turn $O2 \cdots N5$ distances was determined and used as a ‘folding criteria’. The cut-offs for the ‘folding groups’ were set at 4.5 and 6.0 Å (7 s.f.), where specifically ‘closed’ conformations ranged from 0-4.5 Å, ‘medium’ conformations from 4.5 - 6.0 Å and ‘open’ for all configurations with an intra-peptide $O2 \cdots N5 > 6.0$ Å.

The underlying conformational changes that accompany the closing of the $O2 \cdots$ distance,

forming a direct intra-peptide hydrogen bond, are analyzed using the peptide backbone dihedral angles. These dihedral angles can also be correlated with each other (Fig. 4) or used as a criterion for quantification of specific peptide backbone conformations (Fig. 5) using the ANGULA program. It is also possible to determine standard deviation of whole dihedral angle distributions or individual peaks.

ANGULA also allows for the relative positions and orientations of molecules and atoms to each other to be obtained.^{59,60} By combining nearest neighbor water analysis of two peptide sites it is also possible to identify peptide sites that share the same water molecule, i.e. are bridged and stabilized by a water molecule (Fig. 8).

Results and Discussion

Figure 2 shows the measured X-ray and neutron diffraction patterns, after appropriate corrections ($F(Q)$), for the various isotopomers of KGPGK in aqueous solution and the EPSR fits to these data as have been previously presented and described.³⁰ In addition, this figure shows the Fourier transformation of each data set, the total pair-correlation functions $G(r)$ s, again compared with the EPSR fits in real space. The difference curves for the $F(Q)$ data show good agreement between the data and EPSR model and the comparison between Fourier transformation of both model and fit is in excellent agreement.

Peptide ‘folding’ groups

As the KGXGK peptides are in the process of folding in solution, there will naturally be a variable distribution of peptide conformations throughout the course of the diffraction experiments and within the subsequent simulations. In order to classify the different conformations of the ‘folding’ states, the peptides have been assigned to one of three different groups (‘closed’, ‘medium’ and ‘open’) throughout all of the simulations using ANGULA. The criteria for these groupings is the $O2 \cdots N5$ (Fig. 1) intra-peptide distance as this is the

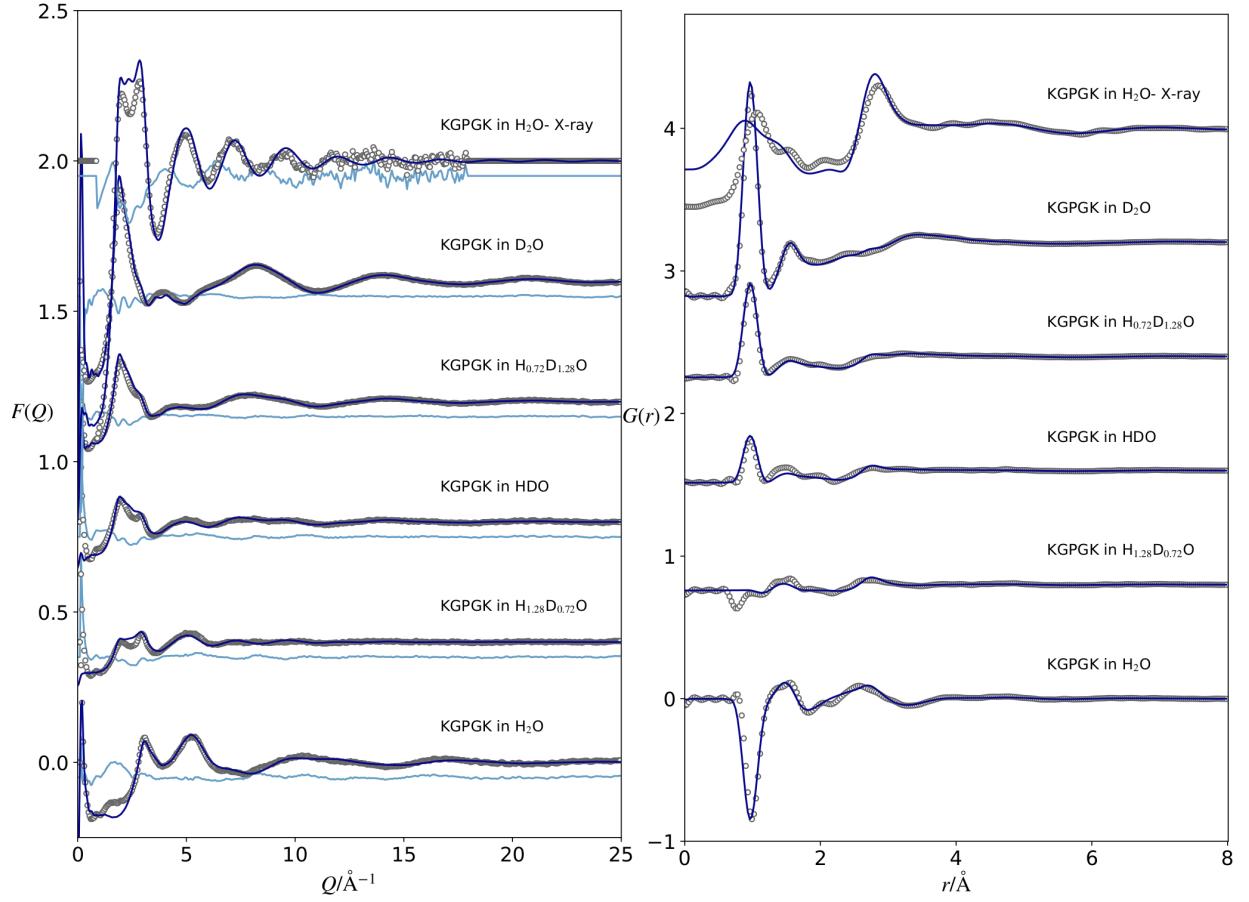


Figure 2 Left panel: Measured neutron and X-ray diffraction data $F(Q)$ (circles) for KGPGK in aqueous solution compared to the EPSR fits to this data (solid lines) along with the differences between data and fit (cyan lines). Right panel: Fourier transformation of the data (circles) and EPSR fits (solid lines) in real space ($G(r)$).³⁰ Both $F(Q)$ s and $G(r)$ s have been shifted vertically for clarity.

site of the i to $i + 3$ β -turn.^{7,19} The percentage of peptides within each folding group are shown for the KGXGK peptides in Table 1. The ‘closed’ distance range is up to the maximum distance that could be considered a direct $i \rightarrow i + 3$ hydrogen bond, thus giving a fully folded peptide.

Table 1 Percentages of KGXGK peptides in the ‘closed’, ‘medium’ and ‘open’ folding conformations based on the O2 \cdots N5 distance ($i \rightarrow i + 3$ β -turn position).

		closed / % (0 – 4.5 Å)	medium / % (4.5 – 6.0 Å)	open / % (6.0 – 10 Å)
EPSR	KGPGK	7.6	31.2	61.2
MD	KGPGK	2.2	34.5	63.3
	KGLGK	2.7	13.8	83.5
	KGSGK	1.2	14.4	84.4
	KGGGK	1.3	14.0	84.7

The ‘medium’ and ‘open’ grouping were designated by reference to the RDFs for the KGPGK intra-peptide O2 \cdots N5 distances (Fig. 3) from the MD simulations, where these two folding groups can be roughly distinguished. The EPSR simulations for this peptide show a broader range of peptide O2 \cdots N5 distances, particularly in the ‘closed’ and ‘medium’ folded states from 3.0 to 5.0 Å. This is most likely a result of the difference in the simulation techniques (as described in the Methods) where ‘short-lived’ intermediate conformations are more likely to be observed in the diffraction data and thus the EPSR simulations. Neutron and X-ray diffraction measures a series of snapshots in solution, where each snapshot gives the instantaneous structure of KGPGK in solution. EPSR, as it is constrained to fit this experimental data, will give a distribution of peptides which captures the peptide molecules at a variety of conformational states as the molecules are moving in solution. Comparatively, as a result of the classical forcefields used to solve Newton’s classical equations of motion,⁵² MD can show a more limited distribution of molecular conformations for these molecules in solution. This difference in conformational distributions is apparent in Fig. 3 for the ‘medium’ peptides. However even though there is a broader distribution of ‘medium’ folded peptides, with respect to the O2 \cdots N5 distance in EPSR, overall the % of these peptides as

shown in Table 1 within each range is consistent with the MD simulations.

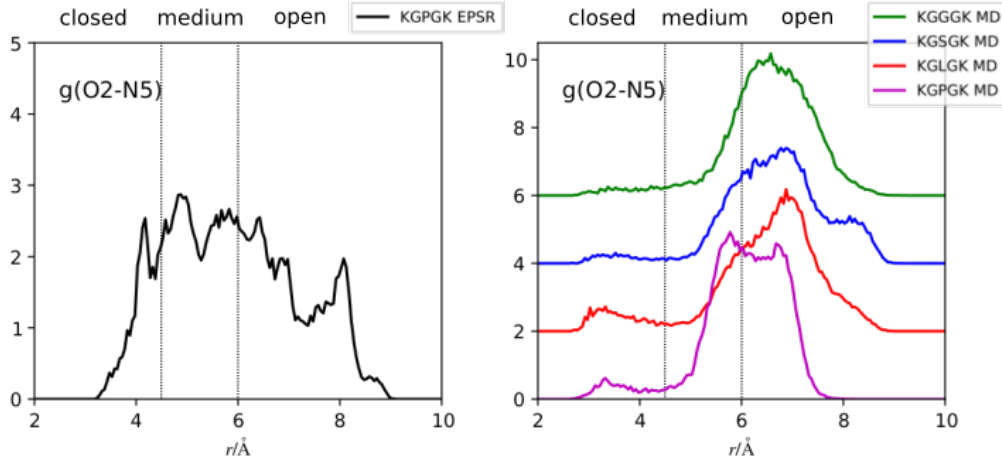


Figure 3 Intra-peptide RDFs for the O2 \cdots N5 distance ($i \rightarrow i + 3$ β -turn interaction) for KGXGK peptides. The distribution limits for each conformational group - ‘open’, ‘medium’ and ‘closed’ - as listed in Table 1 are indicated with dotted lines.

As expected, Table 1 shows that most of the KGXGK peptides are not closed or fully folded in solution. Of these the few peptides which do fully fold, KGPGK and KGLGK are more likely to be found in ‘closed’ conformations. While in comparison, KGGGK and KGSGK have relatively fewer fully folded motifs. Interestingly, the largest difference between the different KGXGK peptides is that KGPGK has a significantly higher proportion of peptides within the ‘medium’ O2 \cdots N5 distance range in both MD and EPSR fits to the diffraction data, with around 30-35% of the peptides in the intermediate (‘medium’) stage of folding.

Figure 4 shows Ramachandran plots⁶¹ ($[\phi, \psi]$ plot) for the Pro3 and Gly4 residues in the ‘open’, ‘medium’ and ‘closed’ conformations for KGPGK from both EPSR fits to the diffraction data and from the MD simulations. These two residues are the only ones within KGPGK which show significant changes upon the transition from open to closed, other than the X3 dihedrals (*vide infra*; the remaining dihedral angles are shown in the SI). In this figure, the conformational groups from the EPSR fits to the diffraction data are less clearly defined compared with MD, again due to the broader distribution of conformations observed in the diffraction experiments. Interestingly and most notably, the conformational changes

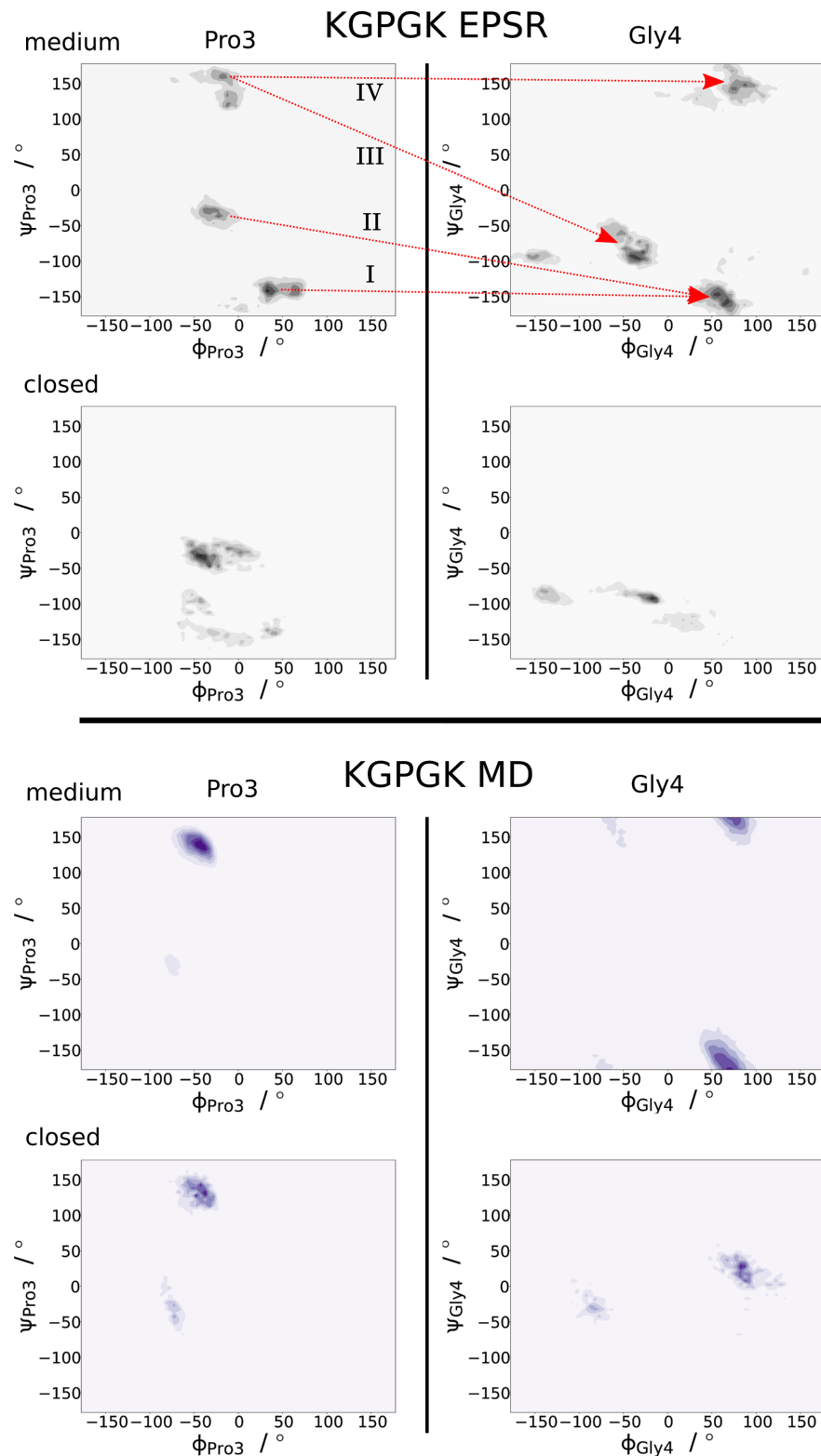


Figure 4 Ramachandran plots for the Pro3 and Gly4 residues in KGPGK for ‘medium’ and ‘closed’ conformation groups from EPSR (left) and MD (right). Also depicted for the EPSR plots are the Pro3/Gly4 correlations giving rise to the I, II, III and IV ‘medium’ conformations as determined using ANGULA.

between ‘open’, ‘medium’ and ‘closed’, are not only found for ϕ_{Gly4} and ψ_{Gly4} but also within ϕ_{Pro3} .

Table 2 Dihedral angle values for KGPGK from MD and EPSR fits to the diffraction data from Ramachandran plots shown in Fig. 4.

	Conformation	ϕ_{Pro3}	ψ_{Pro3}	ϕ_{Gly4}	ψ_{Gly4}
EPSR	medium I	$45(\pm 17)^\circ$	$-142(\pm 11)^\circ$	$72(\pm 37)^\circ$	$-129(\pm 51)^\circ$
	medium II	$-27(\pm 13)^\circ$	$-32(\pm 12)^\circ$	$72(\pm 37)^\circ$	$-129(\pm 51)^\circ$
	medium III	$-18(\pm 14)^\circ$	$146(\pm 15)^\circ$	$-50(\pm 22)^\circ$	$-68(\pm 18)^\circ$
	medium IV	$-18(\pm 14)^\circ$	$146(\pm 15)^\circ$	$89(\pm 18)^\circ$	$137(\pm 18)^\circ$
	closed	$-36(\pm 14)^\circ$	$-32(\pm 12)^\circ$	$-19(\pm 60)^\circ$	$-112(\pm 20)^\circ$
MD	medium	$-53(\pm 16)^\circ$	$129(\pm 25)^\circ$	$73(\pm 22)^\circ$	$-172(\pm 36)^\circ$
	closed	$-51(\pm 17)^\circ$	$-134(\pm 20)^\circ$	$82(\pm 32)^\circ$	$32(\pm 29)^\circ$

From Figure 4, roughly 4 ‘medium’ peptide- conformations can be identified and quantified using ANGULA from the EPSR, and only one medium is present in the MD simulations. Table 2 delineates the Ramachandran angles shown in Fig. 4 for each of these ‘medium’ conformations. Again it is worth noting that EPSR is constrained to the diffraction data and as such is expected to contain a comparatively larger range of conformations for KGPGK in solution. Conformations I and II show two distinct proline backbone conformations for the Pro3 angles where these two folded states are both linked to a single conformation with respect to the Gly4 backbone dihedral angles. Conversely, a single conformation for Pro3 is linked to two conformations for the Gly4 backbone dihedral angles to give conformers III and IV. The corresponding plots in Figure 4 from MD, as expected, shows more limited distributions with respect to the ‘medium’ KGPGK peptides. In MD, a single predominant conformation is present for the peptides in this intermediate stage of folding, where this roughly corresponds to conformations III and IV with respect to the Pro3 Ramachandran angles and conformations IV and I for the Gly4 Ramachandran angles in EPSR. The ‘closed’ conformation Ramachandran plots show more limited KGPGK conformations found from the EPSR fits to the diffraction data, compared with the distribution of ‘medium’ conformations, with the largest population of the ‘closed’ conformation in EPSR shown in Table 2. MD shows a somewhat different ‘closed’ conformation when compared with the predominant

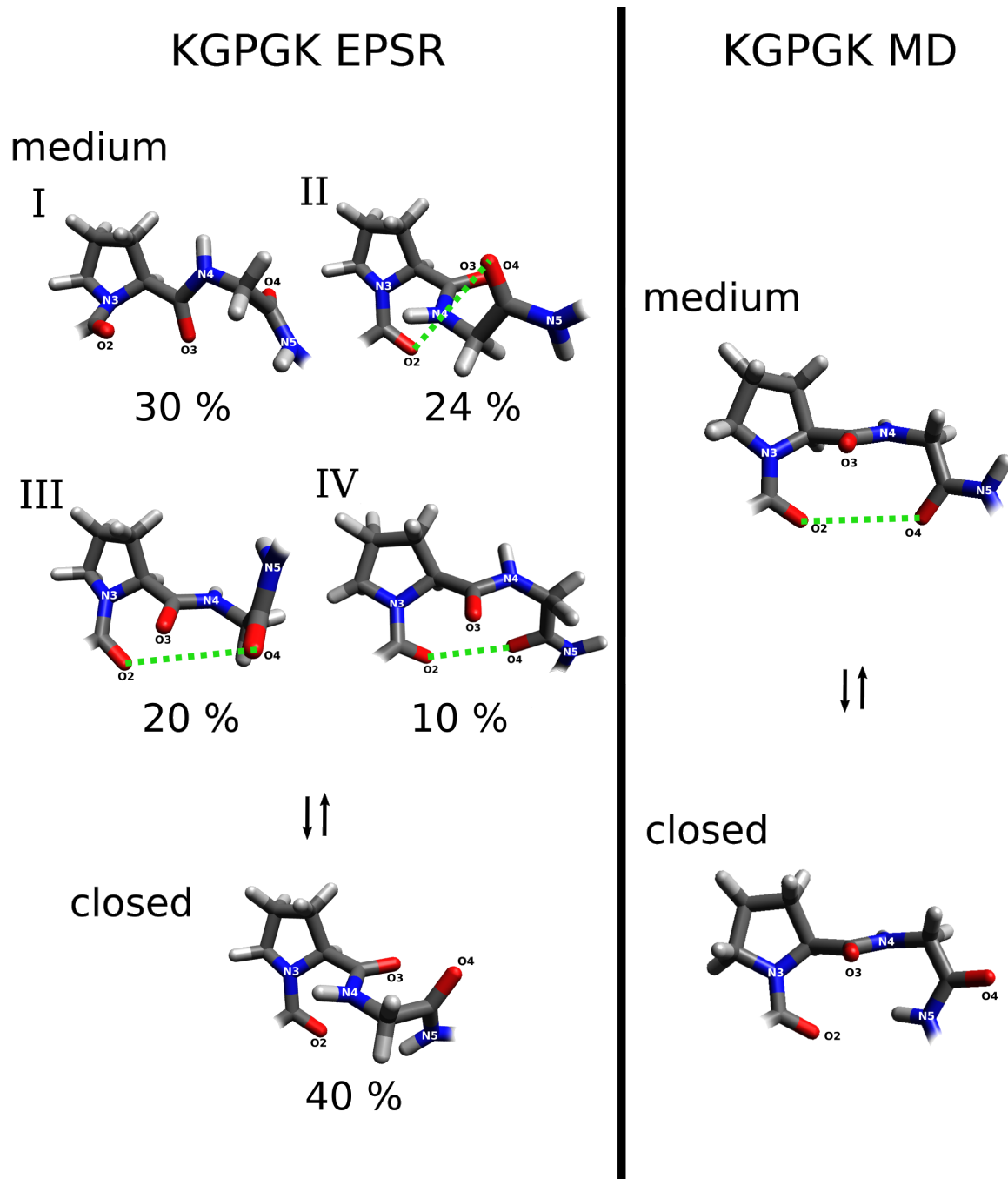


Figure 5 Representative KGPGK peptide ‘medium’ and ‘closed’ conformations from EPSR (left) and MD (right), corresponding dihedral angles found in Figure 4. The green lines between O2 and O4 show the open ‘pathway’ between these two atoms on the KGPGK molecules.

conformer from the EPSR fits to the experimental data.

Figure 5 shows the representative molecular structures for the ‘medium’-conformations of the KGPGK peptides from the Ramachandran data shown in Fig. 4. In addition, the populations of the 4 medium conformations observed in the EPSR fits to the diffraction data have been quantified using ANGULA by using a larger angle deviation than that reported in Table 2 (where the deviation was calculated) in order to ensure inclusion of all of the peptides within the medium conformation group. The representative peptides were chosen by random selection from each group as within each group the peptides are highly similar as observed in Fig. 4. Although the peptides exhibit a broader range of possible conformations in EPSR, as expected from the differences in simulation techniques where EPSR is more representative of the measured diffraction data, it is clear that conformations III & IV are similar to the single medium conformation observed in the MD simulations especially with respect to the O2, O3 and O4 atoms. Specifically, the O2 and O4 atoms are pointed towards the ‘inside’ of the peptide backbone where there is a clear ‘pathway’ between the O2 and O4 atoms, as indicated by the green dotted line in Fig. 5. There is a similar pathway between O2 and O4 apparent in conformation II for EPSR, although this is somewhat impeded by the Pro2-Gly3 amide group being pointed in towards the ‘inside’ of the molecule. Conformation I on the other hand shows the O2 and O4 atoms much further apart in space where these two atoms are roughly oriented so they are pointing completely away from one another, with the Pro2-Gly3 carbonyl between these two atoms. Despite having slightly different conformations, KGPGK conformations II, III and IV in EPSR fits to the diffraction data and the MD medium conformation suggest a ‘pathway’, as marked by the green lines between O2 and O4 in Fig. 5.

Reflective of their Ramachandran plots, the ‘closed’ conformations from the EPSR fits to the diffraction data and MD are somewhat different, especially with respect to the Gly4 orientation. In EPSR, this results in the most likely ‘closed’ conformation having the Pro3-Gly4 peptide bond carbonyl (O3) pointed in the opposite direction from the O2 at β -turn

position and in MD the ‘closed’ conformation shows O3 atom pointed towards the O2 atom. For MD this closed peptide is remarkably similar to the ‘medium’ conformation with only the O4 atom flipped out of the plane once the β -turn is formed. Similarly, the ‘closed’ conformer from EPSR shows O4 flipped away from the β -turn site. That both ‘medium’ and ‘closed’ conformers are similar in MD is likely due to a result of the dihedral constraints within the simulation itself. Yet importantly, the relative position of the O4 atom during the the ‘medium’ to ‘closed’ transitions is surprisingly similar between the two techniques.

Folding conformation transitions for KGXGK peptides

Figure 6 shows a comparison of the dihedral angles ϕ_{Gly4} and ψ_{Gly4} within the folding groups from the MD simulations on the KGXGK peptides compared with KGPGK, where these dihedral angles are the same as those used to generate the Ramachandran plot for the Gly4 residue (Fig. 4 (left column)). The Gly4 angles are depicted as these are the dihedral angles which are most significantly affected by the change in the $\text{O2} \cdots \text{N5}$ distance (all of the dihedral angles for ‘open’, ‘medium’ and ‘closed’ conformations of all of the KGXGK peptides are included in the SI).

Generally, in all KGXGK peptide conformations from MD there are two preferred ϕ_{Gly4} dihedral angle maxima at $\simeq -75^\circ$ and $\simeq 75^\circ$. The universal trend is that in the ‘open’ conformations the overall preferred ϕ_{Gly4} dihedral angle value is -75° and upon folding from ‘open’ to ‘medium’ this preference shifts from -75° to 75° . Subsequently, the folded (‘closed’) KGXGK peptides retain this preference for $\phi_{\text{Gly4}} \simeq 75^\circ$ and the ψ_{Gly4} -maximum shifts from 180° to 30° as the peptides transitions from ‘medium’ to ‘closed’. Interestingly, the ψ_{Gly4} dihedral angle is largely unchanged between the ‘open’ and ‘medium’ conformations. Across the series of peptides the folding appears to occur via a specific intermediate transition state where the conformation adopted by the peptides is depicted in Figure 6. For the ‘medium’ conformation of the peptides, rather than proceeding rapidly from open to closed, the peptides fold in a two-step process: that is from open \leftrightarrow medium \leftrightarrow closed. The

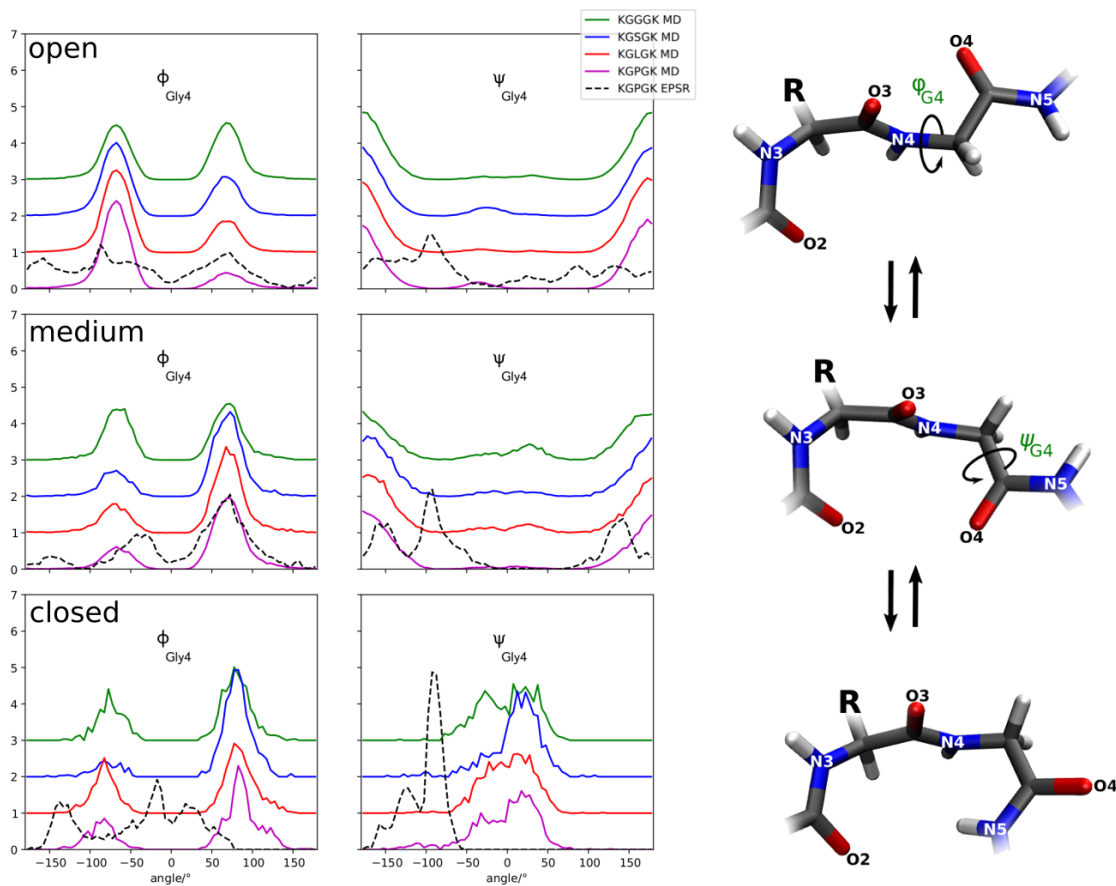


Figure 6 Dihedral angle distributions of ϕ_{Gly4} and ψ_{Gly4} , their change upon a closing $\text{O2} \cdots \text{N5}$ distance (left) and representative conformations observed in all KGXGK peptides in MD (right).

transition probabilities between the three folded states -‘open’, ‘medium’ and ‘closed’ - in all of the KGXGK peptides are consistent with this two-step mechanism. Table 3 shows that there is a very low probability that a peptide does not proceed to folding through the intermediate ‘medium’ conformation as evidenced by the virtual absence of open \leftrightarrow closed transitions in all of the peptides (far right column of row 1 and first column of row 3 in Table 3).

Table 3 Transition probabilities in % between 4 ps snapshots in the MD simulations for KGPGK (purple), KGLGK (red), KGSGK (blue) and KGGGK (green).

state(t) \ state(t+4 ps)	closed	medium	open	KGXGK
closed	92.6	7.4	<0.1	KGPGK
	87.7	12.1	0.2	KGLGK
	85.0	14.4	0.6	KGSGK
	74.8	24.8	0.4	KGGGK
medium	0.6	77.1	22.3	KGPGK
	2.3	55.9	41.8	KGLGK
	1.3	57.5	41.2	KGSGK
	2.1	48.6	49.3	KGGGK
open	<0.1	9.6	90.4	KGPGK
	<0.1	6.6	93.4	KGLGK
	<0.1	6.6	93.4	KGSGK
	<0.1	7.4	92.6	KGGGK

While KGPGK has the highest probability to remain ‘closed’ (92.6 %) compared to the other peptides (≤ 88 %), all of the peptides once closed tend to remain folded despite the number of folded peptides being low throughout all of the simulations (Table 1). KGPGK has the lowest probability to transition into the ‘closed’ state from ‘medium’ (0.6 %) compared to the other peptides (≥ 1.3 %), yet KGPGK has, by far, the highest probability to stay in the ‘medium’ state (77.0 %) compared to the other peptides (≤ 56 %). KGPGK has the highest probability to transition from ‘open’ to ‘medium’ (9.6 %) compared to the other peptides (≤ 7.4 %) and the lowest probability to remain in the ‘open’ state (90.4 %) compared to the other peptides (≥ 92.6 %). This data shows that the intermediate folding state is stabilized in KGPGK relative to the other peptides within the KGXGK series.

Figure 7 shows the lifetime of each of the folding conformations for KGPGK compared with the other KGXGK peptides from the MD simulations. This figure gives the probability that the KGXGK peptide remains in a certain ‘folding state’ for a given time. In all of KGXGK peptides, the ‘open’ conformation has the longest survival time and the presence of long-lived ‘open’-conformations is similar between the series of peptides. Both KGPGK and KGLGK show a larger proportion of peptides with quite long survival times (>1 ns) in their ‘closed’ conformations resulting in an overall longer average survival time for these two peptides with more hydrophobic residues. The relative longevity of the fully-folded state for the series is in the order $\text{KGGGK} < \text{KSGSK} < \text{KGLGK} < \text{KGPGK}$. The most significant difference between the four peptides is the survival time of peptides in the intermediate folding conformation (‘medium’), which shows that $\text{KGGGK}=\text{KSGSK}=\text{KGLGK} \ll \text{KGPGK}$, with KGPGK having a fairly significantly longer average survival time in this ‘medium’ folding conformation.

Water stabilization of the ‘medium’ state

Compared to the other KGXGK peptides, the ‘medium folding intermediate’ in KGPGK is more stable in solution. In the medium conformation, the KGXGK peptides are not fully folded, but rather are in the process of folding to form a fully folded peptide. Given that this intermediate has a relatively long survival time, it is likely to be stabilized somehow by the surrounding water molecules. From Figures 5 & 6 the intermediate ‘medium’ conformations II and III from the EPSR fits to the diffraction data and in the intermediate ‘medium’ conformation from MD show similar conformations with both the O2 and O4 atom pointed towards the ‘inside’ of the conformation. This suggests that the $\text{O2} \cdots \text{O4}$ distance is somehow relevant to stabilization of these intermediate states in all of the peptides. To test this hypothesis a nearest neighbor water analysis was performed using ANGULA. In this analysis, all of the KGXGK peptides were analyzed to quantify how many peptides had water molecules that were first nearest neighbors to both O2 and O4 atoms. Table 4 shows the percentage of

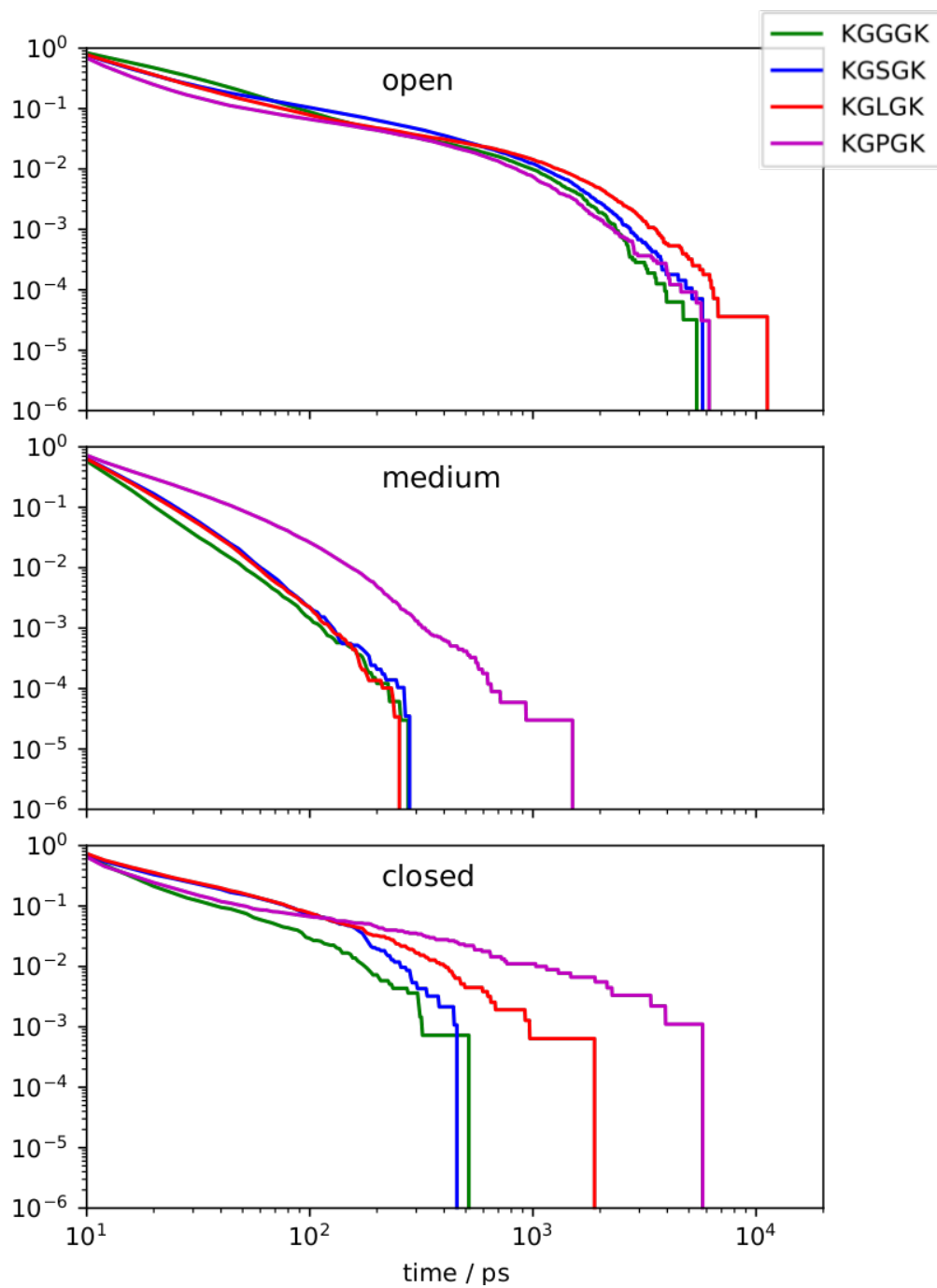


Figure 7 Distribution of average survival time of KGXGK peptide ‘folding states’ in MD.

peptides in each ‘folding state’ that have one mediating water molecule between O2 and O4. In all of the MD simulations $\simeq 25\%$ of the ‘medium’ KGXGK peptides are stabilized by the a first nearest neighbor water molecule which bridges the distance between O2 and O4. In EPSR, which has a higher population of peptides within the ‘closed’ O2 \cdots N5 distance range, there is still a water molecule present between O2 and O4 even when KGPGK is close to being fully folded. It should be noted that within this range in EPSR the peptides are not completely folded as they are for MD as can be seen in Figure 3 where KGPGK in EPSR shows no minimum at 4.5 Å and relatively few close nearest neighbor O2 \cdots N5 contacts.

Table 4 Percentages of peptides with a bridging water molecule between O2 and O4 in relation to the total number of peptides within the groups: open, medium and closed.

		closed / %	medium / %	open / %
EPSR	KGPGK	10.1	13.7	1.8
MD	KGPGK	1.7	27.2	4.7
	KGLGK	<0.1	23.5	3.7
	KGSGK	<0.1	23.8	3.0
	KGGGK	<0.1	24.0	7.2

The ‘medium’ KGPGK peptides which had a shared nearest neighbor water molecule were further analyzed in order to understand the positioning and orientation of this shared water molecule using ANGULA. Figure 8 shows the results of this analysis. The top row of this figure shows the average O2 \cdots O4 distance for the KGPGK peptides which have a single mediating water. The plots below show the water position (relative to O2 in spherical polar coordinates; row 2) and the Euler angle plots (θ_{OR}/ϕ_{OR} (row 3) & θ_{OR}/ψ_{OR} (row 4)) which depict the water orientation around O2. It is evident in this figure that the water position and orientation is virtually identical for both MD and EPSR fits to the diffraction data, despite the differences in the ‘medium’ conformations found in the two different simulations as shown in Figure 5. As shown in Figure 8, a nearest neighbor water molecule with the determined relative position and orientation clearly forms a bridge between O2 and O4 in the intermediate folding peptides. The two observed water orientations: 1. $\cos(\theta_{OR}) \approx -0.9$, $\phi_{OR} \approx -150^\circ$ $\psi_{OR} \approx 70^\circ$ and 2. $\cos(\theta_{OR}) \approx 0.1$, $\phi_{OR} \approx -50^\circ$ $\psi_{OR} \approx -30^\circ$ are in fact identical

as ANGULA does not take the symmetry of water molecules into account.

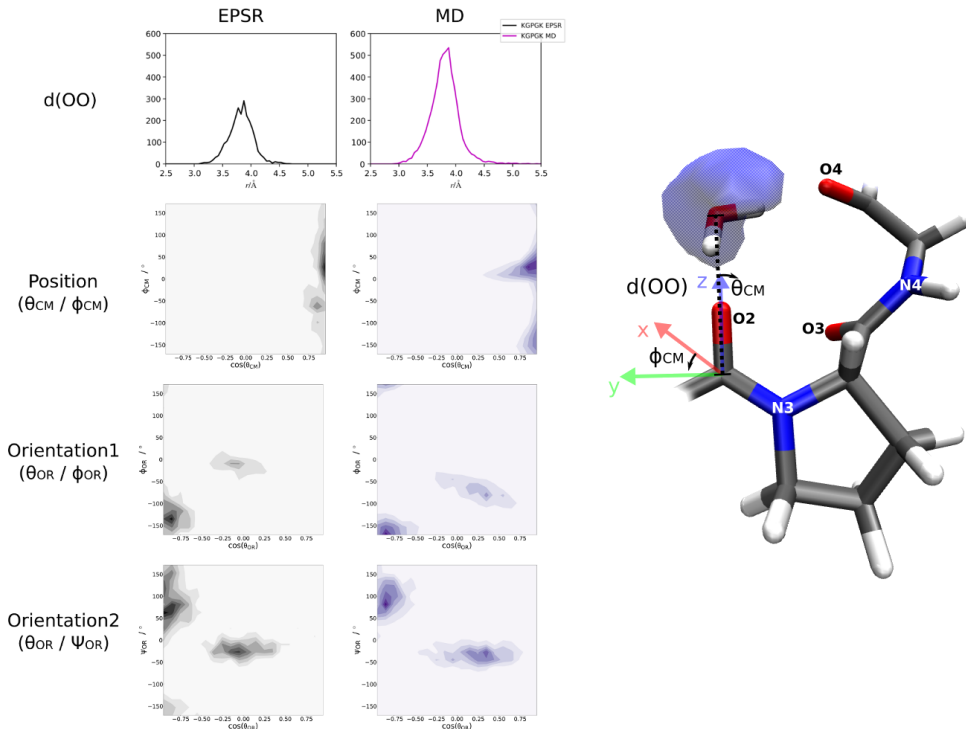


Figure 8 (Left) Mediating water between O2 and O4: Position and orientation relative to the coordinate system assigned to O2. (Right) Most likely position and orientation of a mediating water molecule.

This same water orientation between O2 and O4 is also observed for the other folding 'medium' KGXGK peptides (see SI). Interestingly, this $O2 \cdots O4$ interaction in the 'medium'-range peptides is sometimes bridged by the $-NH_3^+$ group on the lysine side chain (K1) rather than the water molecules (picture in the SI). This is unsurprising given the proximity, polarity and flexibility of this side chain and is likely responsible for the fact that only 25% of the peptides are observed to have a bridging water stabilizing this intermediate state as the K1 side chain bridge does not occur concomitantly with the bridging water. Further, it should be noted that individual water molecules will also move in and out of the $O2 \cdots O4$ individual binding sites and as such an individual water molecule may be exchanged throughout the process of the simulation. These dynamics may also lead to the lower observed percentage

of water-bridging observed for all of the peptides in Fig. 8.

What is special about proline in the $i+1$ position?

From the above analysis, the intermediate folding state in all of the KGXGK peptides appears to be stabilized by a mediating water molecule that helps bridge between the polar motifs of the peptide backbone. As observed for GPG in solution,⁷ it would appear that water is likely to either facilitate or to stabilize the initiation of the β -turning process. Further, the folding pathway appears to proceed from open \rightarrow medium \rightarrow closed and it may be that this water-mediated bridge between O2 and O4 in the intermediate stage of folding is a prerequisite in the formation of the β -turn between O2 \cdots N5. The mechanism is consistent between the series of peptides, yet the intermediate state is much more stable for KGPGK compared to the other peptides (Fig. 7 & Table 3). So the question remains what is special about proline in this position?

Figure 9 shows the peptide backbone dihedral angles in the variable residue X3 (ϕ_{X3}). Unsurprisingly, these dihedral angles were found to be the most variable between the KGXGK peptides in the MD simulations. The observed trend from the ‘open’ to ‘medium’ conformation is a shift in ϕ_{X3} to $\simeq -75^\circ$ throughout the series of peptides. The only exception to this is KGPGK which shows an average ϕ_{X3} angle of $\simeq -50^\circ$ in all of the conformations, be they open, medium or closed. While this clearly does not aid in the mechanism of folding, as this mechanism appears to be consistent between all peptides (Fig. 6), it may provide stability to the peptide structure as it begins to turn, and therefore prevent the KGPGK peptide from unfolding. Thus, this stability may serve as a nucleation point for β -turn formation, which may, in turn, nucleate the protein folding process.

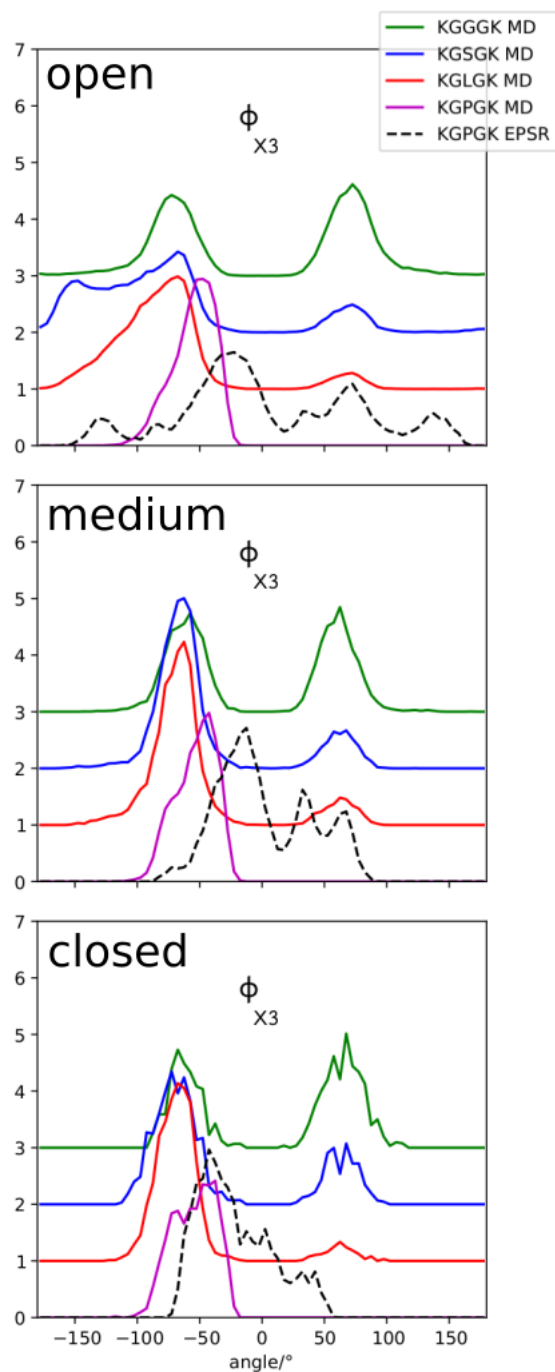


Figure 9 Dihedral angle distributions for the ϕ_{X3} angle (as labeled in Fig.1) for ‘open’, ‘medium’ and ‘closed’ conformations for the KGXGK peptides.

Conclusions

From the data presented here it appears that -GXG- containing peptides follow a consistent mechanism of folding, which proceeds via a two-step process. As the peptides transition from a fully unfolded state (open) to a fully folded conformation (closed) they do so by means of an intermediate folding state rather than just closing in a single step. This two step mechanism of folding appears to proceed by means of a water bridge between two of the backbone carbonyl oxygens within the peptides, specifically the oxygens between the second and fourth peptide bonds of the molecule (O2 & O4; Fig. 1). This initiation of folding yields a peptide conformation where the peptide motifs that participate in the $i \rightarrow i + 3$ β -turn formation, namely O2 and N5, are close enough to one another to nucleate folding into the physiological structure found in proteins.

Although the EPSR fits to the experimental data and MD simulations of KGPGK show somewhat different conformations in solution, as is expected from the difference in the simulations techniques- where one fits the experimental data which measures a larger range of intermediates and the other (MD) has more limited transitions between conformations the mechanism appears to be consistent between the two simulations of KGPGK. Further, and perhaps most importantly, the orientation of the water molecules in the bridging site, between O2 and O4 in the intermediate folding transition state for KGPGK is identical between the two simulation techniques. This confirms that site-specific hydration of water around these groups is important in the folding process as it can stabilize and facilitate β -turn formation in solution. Further, the same two-step folding process between ‘open’, ‘medium’ and ‘closed’ is consistent between all of the KGXGK peptides - suggesting a universal β -turn folding mechanism for -GXG- containing motifs in proteins. Whilst the folding mechanism is consistent regardless of the side chain in the X_3 position, only KGPGK shows a highly stabilized intermediate folding state.

The hypothesis that short-range interactions such as hydrogen bonds form secondary structure elements before the hydrophobic collapse and subsequent formation of the tertiary

structure of the protein has previously been suggested as a vital part of the framework model.²⁶ Although the KGXGK peptides does not form an organized secondary structure - such as an α -helix or a β -pleated sheets, the results here suggest the formation of β -turns as a possible nucleation site for protein folding in solution and the formation of intra-peptide hydrogen bonds independent of a hydrophobic collapse. The results presented here are supportive of the ‘framework’ rather than the ‘hydrophobic collapse’ model,^{62–64} especially with respect to the initial stages of protein folding.

As suggested by Anfinsen, protein folding is most likely to be initiated at certain sites within the polypeptide chain *"that can participate in conformational equilibria between random and cooperatively stabilized arrangements."*⁶⁵ The data presented here suggests that β -turn formation could be one of the initiation sites in protein folding as the -GXG- sequences measured here can transition from folded to unfolded, or vice versa. That is they fluctuate in and out of their native conformation. Similarly, individual water molecules will also likely flicker in and out of the nucleation site. Further, the peptides which contain the -GPG- sequence are less likely to unfold, where the unique structure of proline causes the -GPG- to adopt a more constrained set of conformations which prevent this peptide from unfolding in solution. As a result, perhaps this sequence aids to stabilize this water-mediated interaction so that the β -turn formation is nucleated to initiate protein folding *in vivo*. That is, once the KGPGK peptides have formed this intermediate state they are less likely to fluctuate to ‘open’ conformations than the other peptides, which perhaps explains the prevalence of the -GPG- sequence in β -turn sequences²⁵ observed in fully folded proteins.

Supporting Information

Supporting information is provided which contains a list of the seed-potentials used in the EPSR fits to the diffraction data, a full atomic labeling scheme for the KGXGK molecules, and details of the Euler angles used in the ANGULA analysis. A full set of the backbone

dihedral angles for all of the KGPGK molecules in their conformational groups and RDFs for the intra-peptide O2-O4 interactions for all of the ‘medium’ peptides and the distribution of survival times from the MD simulations is also provided.

Acknowledgements

The authors thank the Leverhulme Trust (RPG-2015-135), the UK Engineering and Physical Sciences Research Council (EP/J002615/1) and the Department of Physics at KCL for funding, the ISIS Facility (STFC, UK) for the allocation of neutron beam time. Also, through our membership of the UK’s HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202), this work used the ARCHER UK National Supercomputing Service (<http://www.archer.ac.uk>).

Keywords: protein folding · neutron scattering · simulation · hydration · peptide

References

- (1) Dill, K. A.; MacCallum, J. L. *Science* **2012**, *338*, 1042–1046.
- (2) Daidone, I.; Ulmschneider, M. B.; Di Nola, A.; Amadei, A.; Smith, J. C. *Proc. Nat. Ac. Sci.* **2007**, *104*, 15230–15235.
- (3) Huang, D. M.; Chandler, D. *J. Phys. Chem. B* **2002**, *106*, 2047–2053.
- (4) Chandler, D. *Nature* **2005**, *437*, 640–647.
- (5) Zhou, R.; Huang, X.; Margulis, C. J.; Berne, B. J. *Science* **2004**, *305*, 1605–1609.
- (6) Huggins, D. J. *J. Comp. Chem.* **2012**, *33*, 1383–1392.

- (7) Busch, S.; Bruce, C. D.; Redfield, C.; Lorenz, C. D.; McLain, S. E. *Angew. Chem. Int. Ed.* **2013**, *52*, 13091–13095.
- (8) Bignucolo, O.; Leung, H. T. A.; Grzesiek, S.; Bernèche, S. *J. Am. Chem. Soc.* **2015**, *137*, 4300–4303.
- (9) Steinke, N.; Gillams, R. J.; Pardo, L. C.; Lorenz, C. D.; McLain, S. E. *Phys. Chem. Chem. Phys.* **2016**, *18*, 3862–3870.
- (10) Ben-Naim, A. *Open J. Biophys.* **2011**, *1*, 1.
- (11) Tarek, M.; Tobias, D. *Phys. Rev. Letts* **2002**, *88*, 138101.
- (12) Huggins, D. J. *J. Struc. Biol.* **2016**, *196*, 394–406.
- (13) Levy, Y.; Onuchic, J. N. *Ann. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389–415.
- (14) Ben-Naim, A. *J. Biomol. Struct. Dynam.* **2012**, *30*, 113–124.
- (15) Blanco, F. J.; Rivas, G.; Serrano, L. *Nature Struc. Mol. Bio.* **1994**, *1*, 584.
- (16) McCallister, E. L.; Alm, E.; Baker, D. *Nature Structural and Molecular Biology* **2000**, *7*, 669.
- (17) Thukral, L.; Smith, J. C.; Daidone, I. *J. Am. Chem. Soc.* **2009**, *131*, 18147–18152.
- (18) Cheung, M. S.; García, A. E.; Onuchic, J. N. *Proc. Nat. Ac. Sci* **2002**, *99*, 685–690.
- (19) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Proc. Nat. Ac. Sci* **1971**, *68*, 2293–2297.
- (20) Chou, P. Y.; Fasman, G. D. *Biochem.* **1974**, *13*, 211–222.
- (21) Wilmot, C.; Thornton, J. *J. Mol. Bio.* **1988**, *203*, 221–232.
- (22) Guruprasad, K.; Rajkumar, S. *J. Biosci.* **2000**, *25*, 143–156.
- (23) Trevino, S. R.; Schaefer, S.; Scholtz, J. M.; Pace, C. N. *J. Mol. Bio.* **2007**, *373*, 211–218.

- (24) Hsu, H.-J.; Chang, H.-J.; Peng, H.-P.; Huang, S.-S.; Lin, M.-Y.; Yang, A.-S. *Structure* **2006**, *14*, 1499–1510.
- (25) Website, <http://www.ebi.ac.uk/pdbe-site/pdbemotif/>; 19th October 2016.
- (26) Kim, P. S.; Baldwin, R. L. *Ann. Rev. Biochem.* **1982**, *51*, 459–489, PMID: 6287919.
- (27) Ben-Nissan, G.; Sharon, M. *Chem. Soc. Rev.* **2011**, *40*, 3627–3637.
- (28) Soper, A. K. *Mol. Sim.* **2012**, *38*, 1171–1185.
- (29) Lenton, S.; Rhys, N.; Towey, J.; Soper, A.; Dougan, L. *Nature Comm.* **2017**, *8*, 919.
- (30) Steinke, N.; Genina, A.; Lorenz, C. D.; McLain, S. E. *J. Phys. Chem. B* **2017**, *121*, 1866–1876.
- (31) Johnston, A. J.; Busch, S.; Pardo, L. C.; Callear, S. K.; Biggin, P. C.; McLain, S. E. *Phys. Chem. Chem. Phys.* **2016**, *18*, 991–999.
- (32) Martinek, T.; Duboué-Dijon, E.; Timr, S.; Mason, P. E.; Baxová, K.; Fischer, H. E.; Schmidt, B.; Pluharova, E.; Jungwirth, P. *J. Chem. Phys.* **2018**, *148*, 222813.
- (33) Duboué-Dijon, E.; Mason, P. E.; Fischer, H. E.; Jungwirth, P. *J. Chem. Phys.* **2017**, *146*, 185102.
- (34) McGregor, J.; Li, R.; Zeitler, J. A.; D’Agostino, C.; Collins, J. H. P.; Mantle, M. D.; Manyar, H.; Holbrey, J. D.; Falkowska, M.; Youngs, T. G. A.; Hardacre, C.; Stitt, E. H.; Gladden, L. F. *Phys. Chem. Chem. Phys.* **2015**, *17*, 30481–30491.
- (35) Mancinelli, R.; Botti, A.; Bruni, F.; Ricci, M. A.; Soper, A. K. *The Journal of Physical Chemistry B* **2007**, *111*, 13570–13577.
- (36) Headen, T. F.; Cullen, P. L.; Patel, R.; Taylor, A.; Skipper, N. T. *Phys. Chem. Chem. Phys.* **2018**, *20*, 2704–2715.

- (37) Seel, A. G.; Swan, H.; Bowron, D. T.; Wasse, J. C.; Weller, T.; Edwards, P. P.; Howard, C. A.; Skipper, N. T. *Angew. Chem. Int. Ed.* **2017**, *129*, 1583–1587.
- (38) Shephard, J. J.; Soper, A. K.; Callear, S. K.; Imberti, S.; Evans, J. S. O.; Salzmann, C. G. *Chem. Commun.* **2015**, *51*, 4770–4773.
- (39) Sears, V. *Neutron News* **1992**, *3*, 26–37.
- (40) McLain, S. E.; Benmore, C. J.; Siewenie, J. E.; Urquidi, J.; Turner, J. F. C. *Ange. Chem. Int. Ed.* **2004**, *43*, 1952–1955.
- (41) Soper, A. *Chem. Phys.* **1996**, *202*, 295–306.
- (42) MacKerell, A. D. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (43) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell Jr, A. D. *J. Chem. Theor. Comp.* **2012**, *8*, 3257–3273.
- (44) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Comp. Chem.* **2004**, *25*, 1400–1415.
- (45) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J. E.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (46) Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D. *J. Comp. Chem.* **2012**, *33*, 2451–2468.
- (47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (48) Reiher, W. E. Doctoral Thesis, Harvard University, Cambridge, Massachusetts, 1985.
- (49) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comp. Phys.* **1977**, *23*, 327–341.

- (50) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *J. Chem. Theo. Comp.* **2008**, *4*, 435–447.
- (51) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. *J. Comp. Chem.* **2009**, *30*, 2157–2164.
- (52) Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.
- (53) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
- (54) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117–1157.
- (55) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (56) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (57) <https://gcm.upc.edu/en/members/luis-carlos/angula/ANGULA>.
- (58) Busch, S.; Lorenz, C. D.; Taylor, J.; Pardo, L. C.; McLain, S. E. *J. Phys. Chem. B* **2014**, *118*, 14267–14277.
- (59) Johnston, A.; Zhang, Y.; Busch, S.; Pardo, L.; Imberti, S.; McLain, S. *J. Phys. Chem. B* **2015**, *119*, 5979–5987.
- (60) Gillams, R. J.; Lorenz, C. D.; McLain, S. E. *Chem. Phys. Letts.* **2017**, *676*, 58 – 64.
- (61) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Bio.* **1963**, *7*, 95–99.
- (62) Rose, G. D.; Dworkin, J. E. In *Prediction of Protein Structure and the Principles of Protein Conformation*; Fasman, G. D., Ed.; Springer US: Boston, MA, 1989; pp 625–633.

- (63) Dill, K. A. *Biochem.* **1985**, *24*, 1501–1509.
- (64) Mok, K.; Kuhn, L.; Goetz, M.; Day, I.; Anderson, N.; Hore, P. *Nature* **2007**, *447*, 106–109.
- (65) Anfinsen, C. B. *Science* **1973**, *181*, 223–230.

Graphical TOC Entry

