Journal of Peptide Science
J. Pept. Sci. 2007; **13**: 679–692
Published online 5 September 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.907

JOURNAL OF PEPTIDE SCIENCE

Loop propensity of the sequence YKGQP from staphylococcal nuclease: implications for the folding of nuclease

SUNITA PATEL[a] and YELLAMRAJU U. SASIDHAR[a,b]*

[a] Department of Chemistry, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India
[b] School of Biosciences & Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India

Received 26 May 2007; Accepted 17 June 2007

Abstract: Recently we performed molecular dynamics (MD) simulations on the folding of the hairpin peptide DTVKLMYKGQPMTFR from staphylococcal nuclease in explicit water. We found that the peptide folds into a hairpin conformation with native and nonnative hydrogen-bonding patterns. In all the folding events observed in the folding of the hairpin peptide, loop formation involving the region YKGQP was an important event. In order to trace the origins of the loop propensity of the sequence YKGQP, we performed MD simulations on the sequence starting from extended, polyproline II and native type I' turn conformations for a total simulation length of 300 ns, using the GROMOS96 force field under constant volume and temperature (NVT) conditions. The free-energy landscape of the peptide YKGQP shows minima corresponding to loop conformation with Tyr and Pro side-chain association, turn and extended conformational forms, with modest free-energy barriers separating the minima. To elucidate the role of Gly in facilitating loop formation, we also performed MD simulations of the mutated peptide YKAQP (Gly → Ala mutation) under similar conditions starting from polyproline II conformation for 100 ns. Two minima corresponding to bend/turn and extended conformations were observed in the free-energy landscape for the peptide YKAQP. The free-energy barrier between the minima in the free-energy landscape of the peptide YKAQP was also modest. Loop conformation is largely sampled by the YKGQP peptide, while extended conformation is largely sampled by the YKAQP peptide. We also explain why the YKGQP sequence samples type II turn conformation in these simulations, whereas the sequence as part of the hairpin peptide DTVKLMYKGQPMTFR samples type I' turn conformation both in the X-ray crystal structure and in our earlier simulations on the folding of the hairpin peptide. We discuss the implications of our results to the folding of the staphylococcal nuclease. Copyright © 2007 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: turn; loop; peptide; free-energy landscape; staphylococcal nuclease; GROMOS96; protein folding

INTRODUCTION

In globular proteins the main chain reverses its direction on the surface of proteins, giving rise to loop and turn structures. The presence of loops and turns allows the proteins to adopt compact globular conformations. Further, loops and turns are associated with protein functions [1–4]. Loops usually connect secondary structural elements like strands and helices. Sibanda and Thornton [5] have analyzed loop sequences and find that a majority of two-residue loop sequences possess Gly, which predominantly assumes α_L and γ_L conformations [6]. Kwasigroch *et al.* [7] analyzed the sequences of 243 proteins and made a loop bank consisting of loops of three to eight amino acids in length. They found that Gly is most frequently found in the loops and its occurrence in loops is 1.7 times more frequent than in the rest of the protein regions. They concluded that in short loops, particularly in tight turns, where the residue is required to take up α_L or γ_L conformation [6], Gly is preferred.

The diffusion collision model of protein folding suggests that microdomains or protein folding initiation sites diffuse towards each other and collide, leading to stabilization of microdomains [8,9]. Some of the protein folding theories also suggest an initial hydrophobic collapse, leading to compaction of the unfolded molecule [10–12]. So these models directly or indirectly require the formation of loop structures in order to bring together microdomains or hydrophobic residues into close proximity. Loop closure events are now considered to be important early folding events during the process of protein folding [13–15]. The question of whether loops play an active role in the process of protein folding or play a passive role is being addressed vigorously in the literature [16–20]. Gu *et al.* [19] investigated the role of turns in peptostreptococcal protein L. They specifically considered two β-turns of the protein. They found that mutation of the Gly residue in the first turn slowed the rate of folding nearly 10 fold. This mutation has little effect on unfolding. On the basis of their results, they suggest that the first β-turn is formed in the folding transition state. Nagi *et al.* [20] considered the folding of the four-helix bundle Rop protein. This is a dimeric protein with a helix–loop–helix monomer.

*Correspondence to: Yellamraju U. Sasidhar, School of Biosciences & Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India; e-mail: sasidhar@iitb.ac.in

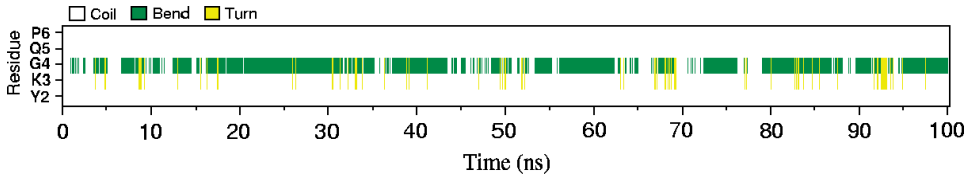

Figure 1 Evolution of secondary structure of the peptide YKGQP as a function of time in PPII simulation. Residues are shown on the *y*-axis and the secondary structure adopted by the residue at a given time is shown along the *x*-axis using color codes as shown in the figure. As can be seen, G4 samples the bend conformation most of the time.

They considered two residues (Asp-Ala) in the loop that connect the helices. When they replaced Asp-Ala with Gly-Gly they found significant accelerations of protein folding and unfolding. From their results they inferred the importance of loop closure events in protein folding, and the role played by Gly in accelerating the folding was evident.

Huang *et al.* [21] considered the two-strand sequences and the turn sequence connecting them from the *N*-terminal β-hairpin of ubiquitin. These three peptides are found to adopt a random coil conformation by CD and NMR experiments. End-to-end collision rates in these peptides were measured by fluorescence-based measurements. They find that the turn is the most flexible for all the three peptides and has higher end-to-end collision rate. They suggest that chain flexibility is an important factor during protein folding. From a mutational study of the turn region, they attribute the flexibility of the turn sequence TLTGK to the presence of Gly in it. Krieger *et al.* [22] showed that loop formation in short loops is accelerated by the presence of the Gly residue. Thus, there is sufficient evidence in the literature to suggest that loop formation is important during protein folding and the presence of Gly in loops seems to play an important role. While the structural role of Gly in loops is already well understood, its kinetic role in loop formation during protein folding is beginning to be probed and understood better now.

In this paper, we address the role played by the loop/turn sequence YKGQP in the refolding of staphylococcal nuclease. In particular, we also consider the role played by the Gly residue of the loop. Recently H/D exchange NMR experiments have been carried out on staphylococcal nuclease, which suggest the early formation of the β-hairpin region during its refolding [23,24]. In the crystal structure of the staphylococcal nuclease, β-hairpin is formed by strand 2 and 3 and connected by type I' reverse turn formed by the sequence YKGQP. Molecular dynamics (MD) study of the β-hairpin sequence DTVKLMYKGQPMTFR in explicit water reveals that starting from the completely unfolded conformations, the peptide folds into the hairpin conformation with both native- and nonnative-like hydrogen-bonding pattern [25]. These results correlate well with the experimental hydrogen exchange protection factors [23,24] available for the refolding of staphylococcal nuclease [25]. In most of the folding events observed in the folding of the hairpin peptide, loop formation involving the region YKGQP is an important event. In order to trace the origins of the loop propensity of the sequence YKGQP, we performed MD simulations of this peptide and a mutated peptide YKAQP in explicit water under constant volume and temperature (NVT) conditions at 300 K.

We find that the YKGQP peptide largely samples loop/turn conformations, whereas YKAQP samples largely extended conformations. The relative free energies of all the conformational forms were determined for both the peptides. We find that Gly plays an important role in determining the loop/turn propensity of the sequence YKGQP. We also explain why YKGQP sequence samples type II turn conformation in these simulations, whereas the sequence as part of hairpin peptide DTVKLMYKGQPMTFR samples type I' turn conformation in both X-ray crystal structure studies and in our earlier simulations on the folding of the hairpin peptide. Implications for the folding of staphylococcal nuclease are discussed.

METHODS

Materials

The coordinates of the sequence YKGQP were taken from the crystal structure of staphylococcal nuclease (PDB code 1SNC) [26]. The ends of the peptides are protected by the $-CH_3CO-$ group (Ac) at the *N*-terminus and the $-NHCH_3-$ group (NMe) at the *C*-terminus, resulting in Ac-Y2-K3-G4-Q5-P6-NMe. MD simulations in Ac-YKGQP-NMe were performed starting from the fully extended, polyproline II and native type I' turn conformations. A mutant peptide Ac-Y2-K3-A4-Q5-P6-NMe was also considered for simulation starting from the polyproline II conformation. In the fully extended conformation, all ϕ, ψ and side-chain dihedral angles were set to 180° except for Pro ϕ (−60°) and $\chi 1$ values, which were set at 60°. The peptide in polyproline II conformation was generated by setting all backbone dihedral angles of the peptide to polyproline II values ($\phi = -76°$ and $\psi = 149°$) except for Pro ϕ, which had a value of −60°.

The following abbreviations are used to indicate the simulations performed:

Native, simulation with initial type I' turn conformation which is the native conformation of the peptide in the crystal structure of staphylococcal nuclease;

DOI: 10.1002/psc

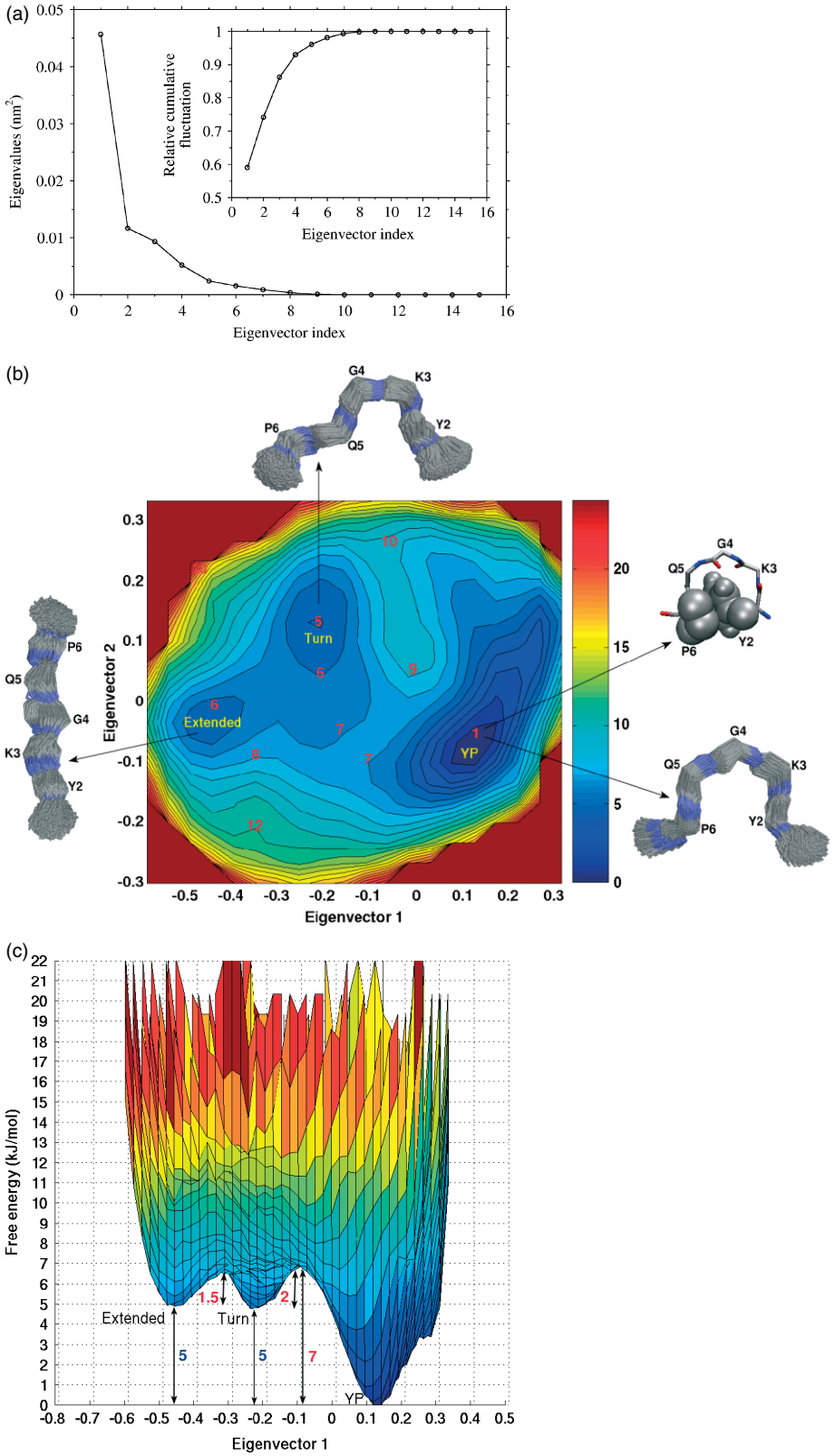

Figure 2 Free-energy landscape of the peptide YKGQP. (a) Eigenvalues obtained from the diagonalization of the C^{α} coordinates covariance matrix for the peptide YKGQP as a function of eigenvector index. The inset shows the relative cumulative positional fluctuation as a function of eigenvector index. (b) Free-energy landscape of the peptide YKGQP in the essential plane defined by eigenvectors 1 and 2. Numerical values labeling the contour levels in the essential plane represent free energies in kJ/mol. Clusters of conformations sampled at the minima labeled YP, Turn and Extended are shown. For the YP minimum, a typical conformation with side chains of Y2 and P6 in van der Waals representation is also shown. (c) Projection of the free-energy surface on eigenvector 1. Free energies (in units of kJ/mol) of the minima and barriers between them are indicated.

DOI: 10.1002/psc

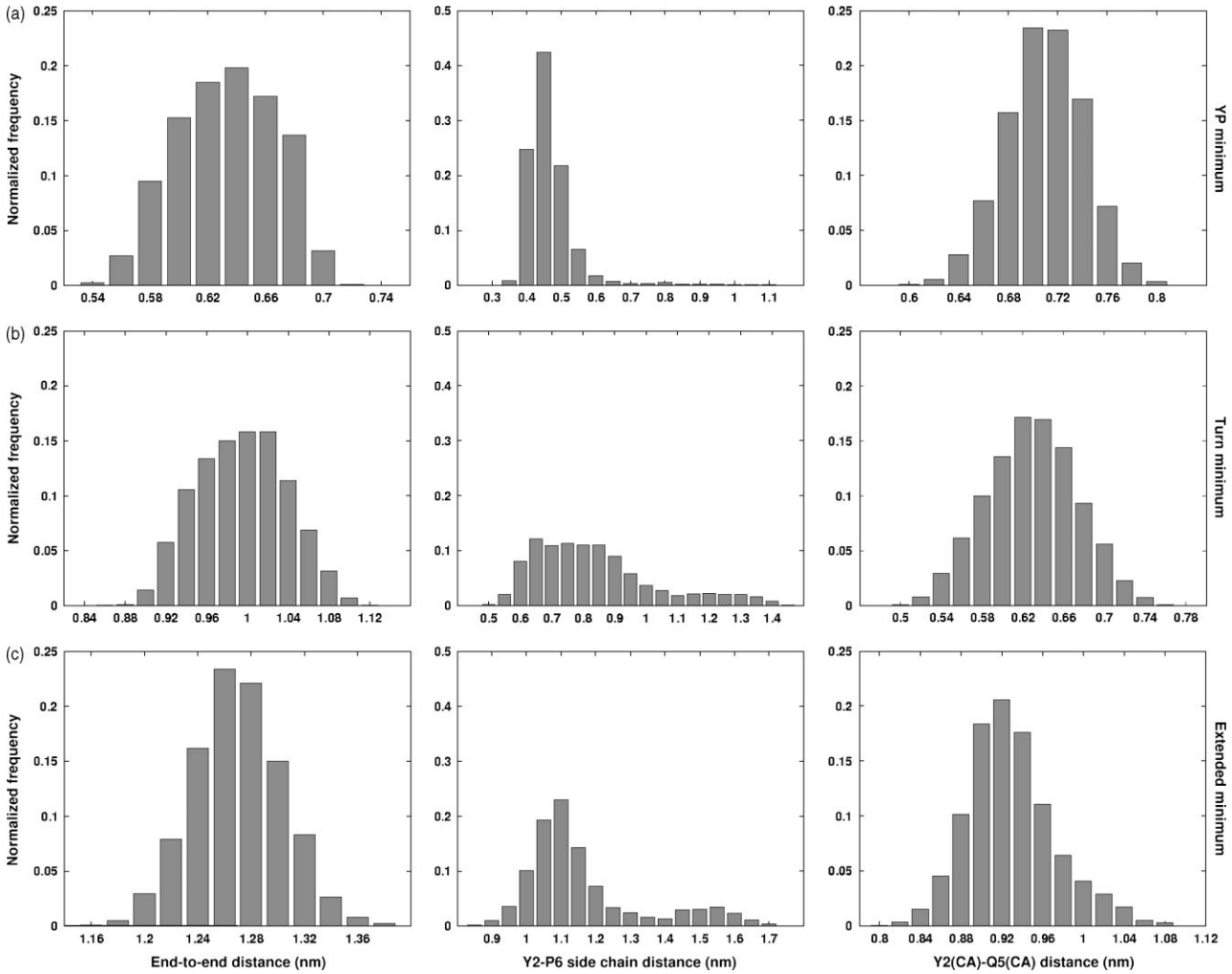

Figure 3 The distribution of end-to-end distances, Y2–P6 side-chain distances and Y2(C^{α})—Q5(C^{α}) distances for conformers sampled by the free-energy minima labeled 'YP' (a), 'Turn' (b) and 'Extended' (c) (see Figure 2(b) and (c) also). Note the shorter end-to-end distances and Y2(C^{α})—Q5(C^{α}) distances (<0.7 nm) in YP and Turn minima, suggesting loop/turn propensity.

PPII, simulation with initial polyproline II conformation of the peptide;

Ext, simulation with initial extended conformation of the peptide.

System for the Simulation

All the simulations were performed in cubic boxes with periodic boundary conditions. One simulation run is performed for each initial conformation with a total simulation time of 300 ns for the peptide YKGQP. The cubic box sides in Native, PPII and Ext simulations were 3.45, 3.84 and 4.02 nm, respectively, and contained 1310, 1692 and 2156 simple point charge (SPC) [27] based water molecules, respectively, solvating the peptide. The simulation for YKAQP was carried out in a cubic box of side 3.83 nm containing 1692 SPC water molecules. For YKAQP, the total simulation length was 100 ns. The total charge for both the peptides was +1 in electronic charge units. In all the simulations a negative counter ion, Cl^{-}, was added by replacing a water molecule to achieve electrical neutrality.

Molecular Dynamics

MD simulations were performed using the software GROMACS (version 3.2.1) [28] and GROMOS96 (ffG43a1) force field [29,30] on a dual xeon processor-based machines running Red-Hat Linux 8/Fedora core 3 (http://www.redhat.com). Except for polar hydrogens like -NH, -OH and aromatic ring hydrogens, the united atom approximation was used. The electrostatic interactions were treated by the Particle Mesh Ewald (PME) method [31,32] with a Coulomb cutoff of 1 nm, Fourier spacing of 0.12 nm and an interpolation order of 4. The van der Waals interactions were treated using the Lennard–Jones potential and a switching function with a cutoff distance of 1 nm and a switching distance of 0.9 nm.

The potential energy of the peptide in water was minimized using the steepest descent algorithm with a tolerance of 100 kJ/mol/nm, and convergence was obtained in all the cases. Subsequent to energy minimization, position-restrained MD was carried out for 50 ps using a reference temperature of 300 K. In this procedure, the atomic positions of the peptide were restrained and the water molecules were allowed to equilibrate around the peptide, to remove the solvent holes. Initial velocities required to start the procedure were generated

DOI: 10.1002/psc

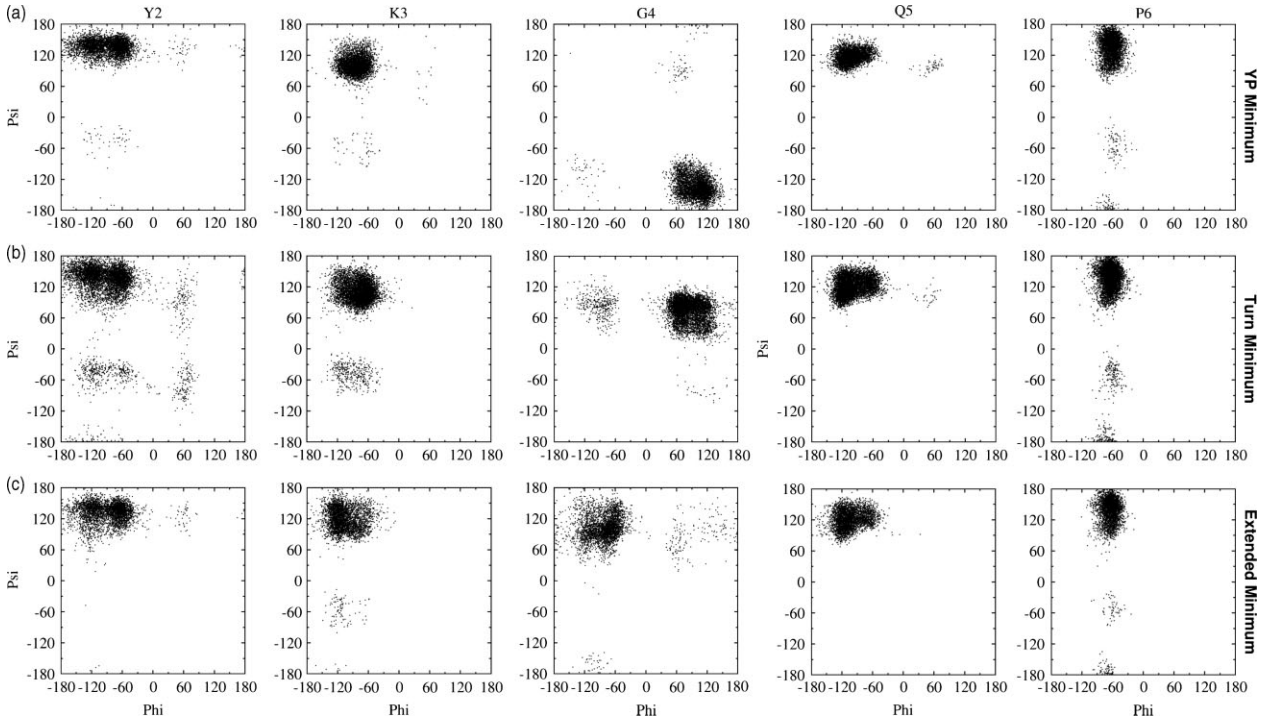

Figure 4 Ramachandran plot of the residues in the YKGQP peptide corresponding to free-energy minima labeled 'YP' (a), 'Turn' (b) and 'Extended' (c) (see Figure 2(b) and (c) also). Note that G4 samples ε and α_L regions [6] in (a) and (b), respectively. In the Turn minimum, a type II reverse-turn conformation is sampled involving residues K3 and G4 as the corner residues. In YP minimum, G4 samples predominantly the ε region [6].

conforming to the Maxwell velocity distribution at 300 K. Following these equilibration procedures, MD was initiated. A time step of 2 fs was used for integrating the equations of motion. The LINCS algorithm was used to constrain the bonds [33]. Coordinates were saved every 250 steps or 0.5 ps and velocities were saved every 10 000 steps or 20 ps. The peptide and the solvent were separately coupled to a Berendsen temperature bath [34] at 300 K using a time constant of 0.1 ps. All the simulations were done under NVT conditions.

The analysis tools provided by GROMACS software were used to analyze the data. The end-to-end distance refers to the distance between first and last residues C^{α} atoms. For secondary structure assignment, the DSSP program [35] was used, implemented in GROMACS. According to DSSP, a coil conformation is defined by a low-curvature region without any hydrogen bonds, and a bend conformation at ith residue is defined involving $i-2$, $i-1$, i, $i+1$, $i+2$ residues [35], where backbone changes direction by more than 70° at the ith residue. We followed the standard definitions of turn conformations [36]. Further, a turn is considered to be present when the distance between the C^{α} atoms of residues i and $i+3$ is less than 0.7 nm [37]. Matlab (http://www.mathworks.com) and Octave (http://www.octave.org) were also used for some of the analyses. Most of the graphs and figures were generated using Xmgrace (http://plasma-gate.weizmann.ac.il/Grace), Matlab and VMD [38].

Essential Dynamics Analysis

We carried out the essential dynamics analysis of the simulations on the peptide YKGQP and YKAQP following the methodology described by Amadei *et al.* [39] and de Groot *et al.* [40]. A covariance matrix of the positional fluctuations of C^{α} atoms was constructed from all the three simulations on YKGQP and diagonalized to obtain eigenvalues and eigenvectors. A similar analysis was carried out for YKAQP also. The eigenvalues were ordered in the descending order of their magnitude. The highest and second highest eigenvalues corresponding to first and second eigenvectors, account for 75 and 62% of the overall positional fluctuation in YKGQP and YKAQP, respectively, and therefore describe the most relevant conformational degrees of freedom. Thus, we used the first two eigenvectors to construct the essential plane for the calculation of the free-energy landscape.

Free Energy Calculation

The essential plane constructed from the first two eigenvectors was divided into 20×20 grid cells, resulting in 400 cells. Conformations sampled every 0.5 ps were projected on to this plane and the number of points in each cell was counted. Under equilibrium conditions, these counts will be proportional to the probabilities of the conformations sampled by the grid cells. Free energies are assigned to all the cells with respect to the reference cell containing the maximum number of points as described in the literature [41]. For example, the Helmholtz free energy of the cell 'i' is computed as

$$\Delta A_{\mathrm{ref}\longrightarrow \mathrm{i}} = -RT \ln N_i/N_{\mathrm{ref}}$$

where N_i and N_{ref} are the number of points in the ith and reference cell, respectively,

R is the ideal gas constant and T is the temperature (300 K).

DOI: 10.1002/psc

RESULTS

An Overview of the Conformations Sampled by YKGQP in Terms of Secondary Structure Assigned by DSSP

Figure 1 shows the evolution of the secondary structure as a function of time in PPII simulation. The green color at G4 indicates the presence of a bend conformation. Further, we observe that the turn conformation involving residue K3 and G4 are interspersed throughout, indicated by the yellow color at K3 and G4 positions. We find that the bend conformation is sampled 70% of the time in the simulation. Similarly, the turn conformation is sampled 5% of the time and coil conformation is observed for 25% of the time (Please see the methodology for DSSP definitions of bend, turn and coil conformations.) These observations establish in a general way that the sequence YKGQP has a turn/loop propensity. Similar results are also obtained in other simulations (Native and Ext) on the YKGQP peptide.

Free-energy Landscape of the Peptide YKGQP in the Essential Plane Defined by the First Two Eigenvectors Shows the Presence of Three Free-energy Minima

As explained in 'Methods' the simulations were carried out with three initial starting conformations (extended, PPII and native) for a total simulation length of 300 ns. Table 1 shows the gross conformational features of the YKGQP peptide in all the three simulations. Since the gross conformational features are similar, all the trajectories are concatenated to have better sampling density, and essential dynamics analysis was carried out. A similar practice has been adopted in the literature also [41]. Root mean square inner product (RMSIP) analysis [42,43] also showed that all the simulations have essentially converged. The RMSIP values for eigenvectors in essential plane are 0.90 (Native simulation), 0.97 (PPII simulation) and 0.98 (Ext simulation). Figure 2(a) shows a plot of the eigenvalue *versus* eigenvector index. The inset shows the relative cumulative fluctuation as a function of the eigenvector index. It is clear from the figure that first two eigenvectors account for about 75% of the overall positional fluctuations, and therefore the first two eigenvectors define the essential plane for conformational sampling. Free energies are computed in this essential plane as described in 'Methods'. Figure 2(b) shows the contour diagram of free energies in the essential plane. As can be seen, there are three free energy minima – one deep funnel-like minimum (labeled 'YP') and two shallow minima (labeled 'Turn' and 'Extended') (see later for the explanation of the labeling). Figure 2(c) shows the projection of the free-energy surface on eigenvector 1. This figure indicates the free energy of the minima and the barrier heights between them.

Table 1 Summary of gross conformational features of YKGQP in the three simulations (Native, PPII and Ext). Mean and standard deviation values of conformational parameters are given. RMSD is calculated with respect to backbone conformation of the peptide in the crystal structure of the staphylococcal nuclease. It can be seen that all conformational parameters in all the simulations are similar

Simulation	R_g (nm)	RMSD (nm)	End-to-end distance (nm)
Native	0.46 ± 0.06	0.25 ± 0.08	0.77 ± 0.23
PPII	0.46 ± 0.07	0.25 ± 0.09	0.81 ± 0.28
Ext	0.46 ± 0.05	0.23 ± 0.07	0.76 ± 0.22

The Free-energy Minimum Labeled 'YP' Corresponds to the Conformational Forms Dominated by Loop Conformation with Y2 and P6 Ring Association

Figure 3(a) summarizes the conformational features of the conformers sampled by the YP minimum (Figure 2(b)). The end-to-end distance distribution has a peak around 0.64 nm and has a range of 0.54–0.7 nm. This distribution indicates that the C^{α} atoms of Y2 and P6 are closer most of the time, suggesting the formation of a loop structure. The distribution of the $C^{\alpha}{}_{Y2}-C^{\alpha}{}_{Q5}$ distance shows peaks at 0.7 nm and less, suggesting the sampling of turn conformational forms, since a turn by definition has a distance of less than 0.7 nm between the C^{α} atom of first residue and the C^{α} atom of fourth residue [37]. Similarly, the distribution of the distance between the Y2 and the P6 ring is suggestive of association between the Y2 and the P6 rings in view of the short distances sampled. Figure 2(b) shows a typical conformation sampled by YP minimum. Thus, YP minimum samples loop conformational forms characterized by association between the Y2 ring and the P6 ring. These conformations largely correspond to bend conformation at G4 in the DSSP plot (Figure 1). Figure 4(a) shows the Ramachandran $\phi-\psi$ plot for all the residues of YKGQP in all the three minima. For YP minimum, it is readily seen that G4 samples most of the time in the fourth quadrant or ε region [6] normally not accessible to any other amino acid residues. This also underlines the importance of the G4 residue in determining the conformational forms sampled by the YP minimum.

The Free-energy Minimum Labeled 'Turn' Samples Turn Conformational Forms Involving Residues YKGQ

Figure 3(b) summarizes the conformational features of conformers sampled by the 'Turn' minimum (Figure 2(b)

DOI: 10.1002/psc

and (c)). The distribution of end-to-end distances (distance between C^{α}_{Y2} and C^{α}_{P6} atoms) has a peak at 1 nm, while the distribution of C^{α}_{Y2}–C^{α}_{Q5} distances has a peak around 0.62 nm, suggesting the sampling of a turn conformation involving the residue Y2-K3-G4-Q5 since a C^{α}_{i}–C^{α}_{i+3} distance less than 0.7 nm is considered to indicate a turn conformation [37]. Thus, this distribution has a large number of conformers with the C^{α}_{i}–C^{α}_{i+3} distance less than 0.7 nm. Figure 2(b) shows a cluster of conformations sampled by the 'Turn' minimum. Further, Figure 4(b) shows the Ramachandran plot for all the residues of YKGQP. From this plot it is clear that G5 samples largely the left-handed helical region and this region is energetically favorable for Gly in comparison to any other amino acid residue. The Ramachandran ϕ–ψ plots for K3 and G4 together imply that the turn conformation sampled is close to type II turn conformation in view of the dihedral angle requirement for type II turn conformation [36,44,45].

The Free-energy Minima Labeled 'Extended' Samples Extended Conformational Forms

From the end-to-end distance (distance between C^{α}_{Y2} and C^{α}_{P6} atoms) and the C^{α}_{Y2}–C^{α}_{Q5} distance distributions shown in Figure 3(c) and the Ramachandran plot shown in Figure 4(c), it is clear that the conformations sampled are essentially extended, and a cluster of conformations sampled by the 'Extended' minimum is shown in Figure 2(b).

DISCUSSION

Role of Glycine in Determining the Loop Propensity of YKGQP

As discussed in 'Results' the deep 'YP' minimum in the free-energy landscape (Figure 2(b) and (c)) is largely sampled by conformers with relatively shorter end-to-end distances (Figure 3(a)). These conformers are also characterized by the association of the Y2 and the P6 rings (Figure 2(b)) possibly due to hydrophobic interactions. The Ramachandran ϕ–ψ maps of the residues of YKGQP (Figure 4(a)) show that G4 samples largely in the fourth quadrant or ε region [6] which is not available to any other amino acids. Thus, Gly has a special role in determining the loop conformation sampled by the 'YP' minimum in the free-energy landscape. Similarly Gly plays a role in determining type II turn-like conformation sampled by the relatively shallow 'Turn' minimum in the free-energy landscape (Figures 2(b), (c) and 4(b)). In view of these observations, it can be inferred that Gly dictates the loop/turn conformational features of the sequence YKGQP.

Many studies on short peptides (four to five residues) reveal that the turn formation is due to local sequence propensity of the peptide sequence. In short peptides, the choice of the amino acid at specific positions is critical [46,47]. For example, Gly residue at the $i+2$ position is frequently observed in type II turn and type I' turn because this residue readily takes up the α_L conformation required for type II turn conformation [36,48,49]. In a recent MD simulation study on YPGDV peptide, it was found that the most populated cluster obtained from the 2 ns self-guided MD simulation featured a type II turn conformation involving Gly in the $i+2$ position adopting ϕ and ψ values unfavorable for other amino acid residues [50]. Scully and Hermans [51] also find that Gly is most stable in a type II turn conformation on the basis of free-energy calculations. In our peptide YKGQP, a Gly residue is present at the $i+2$ position, which enables the peptide to adopt a type II conformation more readily as discussed in 'Results'.

On Why YKGQP Adopts Type II Turn Instead of Type I' Turn as Observed in the Crystal Structure of Staphylococcal Nuclease and in the MD Simulations of the β-Hairpin Peptide DTVKLMYKGQPMTFR from Nuclease

In our simulations on YKGQP, we observed sampling of type II turn conformations involving YKGQ. However, in the crystal structure of the nuclease, YKGQ adopts type I' turn conformation connecting the β-strands of the hairpin peptide DTVKLMYKGQPMTFR. Similarly, when we carried out MD simulations of the folding of the hairpin peptide, we observed sampling of native hairpin conformation, with YKGQ adopting type I' turn conformation in some of the simulations [25]. The reason why we did not observe a type I' turn conformation for the peptide YKGQP is that in a short peptide only local interactions play an important role in determining the turn conformation. In contrast, in a β-hairpin, long-range interactions between the strands modulate the local interactions of the turn region and the right-handed twist in the β-strands could induce a type I' turn conformation in the turn region [52]. Similar observations were made earlier by Chothia [53] and Mattos *et al.* also [54]. Since in the YKGQP peptide the strand residues are missing, long-range interactions are absent and therefore the peptide's conformation is driven by local interactions only, leading to type II turn conformation as observed in our simulations. This is substantiated by the observation that the sequence YKGQ undergoes a transition from a type II to type I' turn conformation during folding of the hairpin peptide. We made this observation when we reanalyzed the trajectories of our earlier work (see Ref. 25 for methodology and results) on the folding of the β-hairpin peptide [25] (Figure 5). In some of the folding simulation of the β-hairpin peptide DTVKLMYKGQPMTFR, one of the early events in folding of the hairpin is the formation of a type II turn involving the YKGQ residues (Figure 5(a)). This conformation

DOI: 10.1002/psc

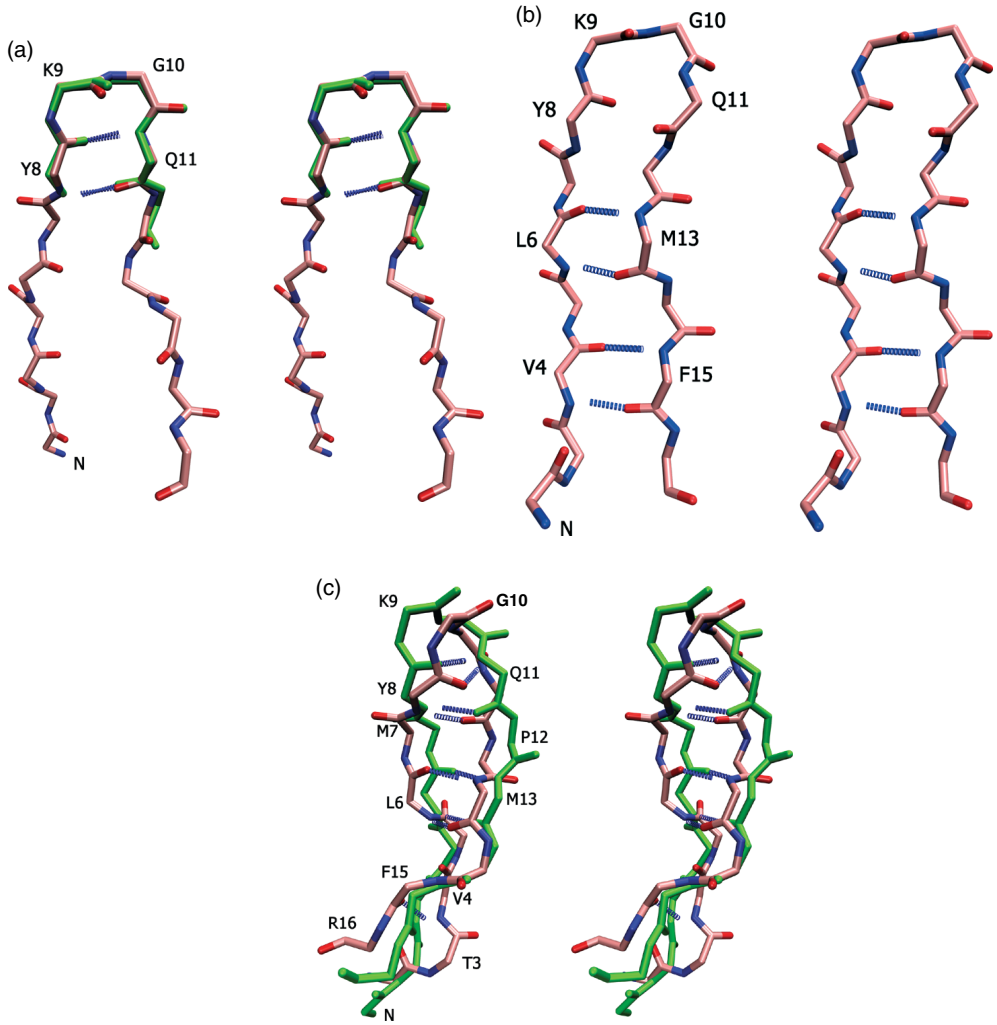

Figure 5 The folding of the β-hairpin peptide DTVKLMYGQPMTFR starting from polyproline II conformation in which the region YKGQ samples type II turn conformation at about 3.18 ns. In the present simulations of the peptide YKGQP, the sequence samples type II conformation. This is superposed on the sequence region YKGQP of the hairpin peptide (a). As can be seen, there is a good match between the turn conformation sampled during the folding of the hairpin peptide and the turn conformation sampled in the simulations of the smaller peptide YKGQP. During the folding of the bigger-hairpin peptide at a later time, 3.5 ns, the pleated structure of the hairpin is evident (b). Once the folding of the bigger hairpin peptide is complete, the hairpin adopts a right-handed twist and the turn region involving YKGQ adopts type I' turn conformation. This conformation is superposed on the structure of the hairpin from the crystal structure of the nuclease showing a good match (root mean square deviation = 0.174 nm) (c). Stereo diagrams showing (a), (b) and (c). In all the figures the conformation of the bigger peptide is shown in CPK color scheme. The superposed conformations are shown in green color.

superposes well with that obtained in the current simulations on YKGQP alone. Subsequently, with the alignment of strands and formation of hydrogen bonds between the strands, type II turn conformation is lost and the peptide samples a pleated sheet structure with an open YKGQ turn (Figure 5(b)). Gradually, the twist of the β-hairpin develops, and eventually type I' turn conformation is assumed by the YKGQ residues (Figure 5(c)). This twisted β-hairpin conformation with type I' conformation at the turn region persists during equilibrium. Thus, it appears that the right-handed twist in the β-hairpin conformation is necessary to stabilize type I' conformation of the turn. The presence of Gly in the $i+2$ position of the turn sequence YKGQ facilitates type I' turn formation.

Stability of the Loop Conformation

Loop conformation in the deep YP minimum is observed most of the time in comparison to other conformations in the YKGQP peptide (Figure 2). The stability of these conformations appears to be governed by the hydrophobic interaction between Y2 and P6 side chains facilitated by the flexibility of the Gly residue. Similar hydrophobic interactions are observed to stabilize a type I turn in the peptide YPGDV in MD simulations due to the hydrophobic interactions between Pro and

DOI: 10.1002/psc

Val [55]. In another simulation study on AYPYD peptide, the peptide was found to be sampling a type VI reverse-turn conformation, which was similar to the NMR-derived structure, and was stabilized by hydrophobic interactions between two Tyr and Pro interactions [56]. Mohanty *et al.* [57] studied a similar peptide SYPFDV and its variants (SYPYD and SYPFD) in water using MD simulation and observed a type VI turn. The preference for type VI turn conformation was attributed to the local propensity of *cis* Pro, to the electrostatic interactions between opposite charges of termini and to the hydrophobic interactions between Tyr/Phe and Pro residues [57]. Interaction between Tyr and Pro side chains is also observed in simulations on phakellistatin and antamanide peptides [58].

The Modest Free-energy Differences and Barriers Between the Minima in the Free-energy Landscape of the Peptide YKGQP

As can be seen from the Figure 2(c), the Turn and Extended minima have free energies of about 5 kJ/mol with respect to deep YP minimum. The barrier between the minima corresponding to turn and extended conformational form is about 1.5 kJ/mol (Figure 2(c)). From YP to Turn minima there is a barrier of about 7 kJ/mol and from Turn to YP minima there is a barrier of about 2 kJ/mol (Figure 2(c)). Thus, the free-energy differences and barrier are modest, and conformational forms can interconvert easily between the free-energy minima. Thus, it is relatively easy to sample turn/loop conformational forms for this sequence. Similar observations were also made in other simulation studies of small peptides. For example, Fuchs *et al.* [59] studied the conformational features of a designed peptide GDNP using MD simulations. The free-energy landscape of the peptide at 300 K showed two major conformations: extended and type VIII β-turn conformations. The extended conformation is sampled more frequently than the turn conformation with a free-energy difference of about 4.2 kJ/mol between them. On going from the extended to the turn conformation (along two different paths) the transition state energy barriers were ~13 and ~15 kJ/mol. They inferred from the small free-energy differences and the low barrier heights between the extended and turn conformations that one conformational form easily interconverts to the other at room temperature.

Yang *et al.* [52] determined the free energy of various turns and observed that the turn conformation is less stable than the extended conformation and the free-energy difference associated with various turn types vary between 1.6 and 7.7 kcal/mol depending on turn type and sequence. Similar observations were also made by Nakajima *et al.* [60]. In contrast, in our simulation we observed the conformation corresponding to YP minimum to be more stable than extended and turn conformations, although we observed small free-energy differences between the turn, extended and loop with Y2 and P6 contact conformations.

Glycine to Alanine Mutation Largely Reduces the 'Loop/Turn' Propensity of the Sequence YKAQP

In order to test the role of Gly in determining the conformational features of the sequence YKGQP, we have carried out G4 → A4 mutation resulting in the YKAQP peptide. MD simulation on this peptide was carried out for 100 ns as described in 'Methods'. Figure 6 compares the end-to-end (C^{α}_{Y2}–C^{α}_{P6}) and C^{α}_{Y2}–C^{α}_{Q5} distance distributions for both the peptide YKGQP and YKAQP. It is readily apparent that loop conformations with end-to-end distances less than 0.7 nm are largely suppressed in the peptide YKAQP. Similarly, turn conformational forms characterized by C^{α}_{Y2}–C^{α}_{Q5} distances less than 0.7 nm are significantly suppressed in the YKAQP peptide (Figure 7(a)). In the YKAQP peptide, no conformers are observed with YP association and turns involving YKAQ.

The Free-energy Landscape of YKAQP in Essential Plane Defined by the First Two Eigenvectors Shows the Presence of Two Free-energy Minima

Figure 7(a) shows the free-energy landscape computed for the peptide in the essential plane defined by the first two eigenvectors. They correspond to 62% of overall positional fluctuation and hence describe the most relevant conformational degrees of freedom. There are two free-energy minima. The deep minimum corresponds to extended conformational forms and the relatively shallow minimum corresponds to bend/turn conformational forms. As can be seen from Figure 7(b), the free-energy difference between the minima is modest. In comparison to YKGQP, in this case the dominant conformational form sampled is extended. We no longer see any minima for Y2- and P6-associated conformational forms. The only other conformational form sampled is a bend or a turn conformation, which is like a type VIII reverse-turn [48] conformation involving K3-A4-Q5-P6 residues. This is evident from the Ramachandran ϕ, ψ map for the residues A4 and Q5 in 'Bend/Turn' minimum (Figure 8). Thus, mutating G4 → A4 leads to the loss of the Y2- and P6-associated loop conformational form and the Y2-K3-G4-Q5 turn conformational form. This establishes that for preferential sampling of loop conformation Gly is essential.

Implications for the Folding of Hairpin Sequence from Staphylococcal Nuclease

In the MD simulations of the peptide YKGQP, Y2- and P6-associated conformational forms are preferentially

DOI: 10.1002/psc

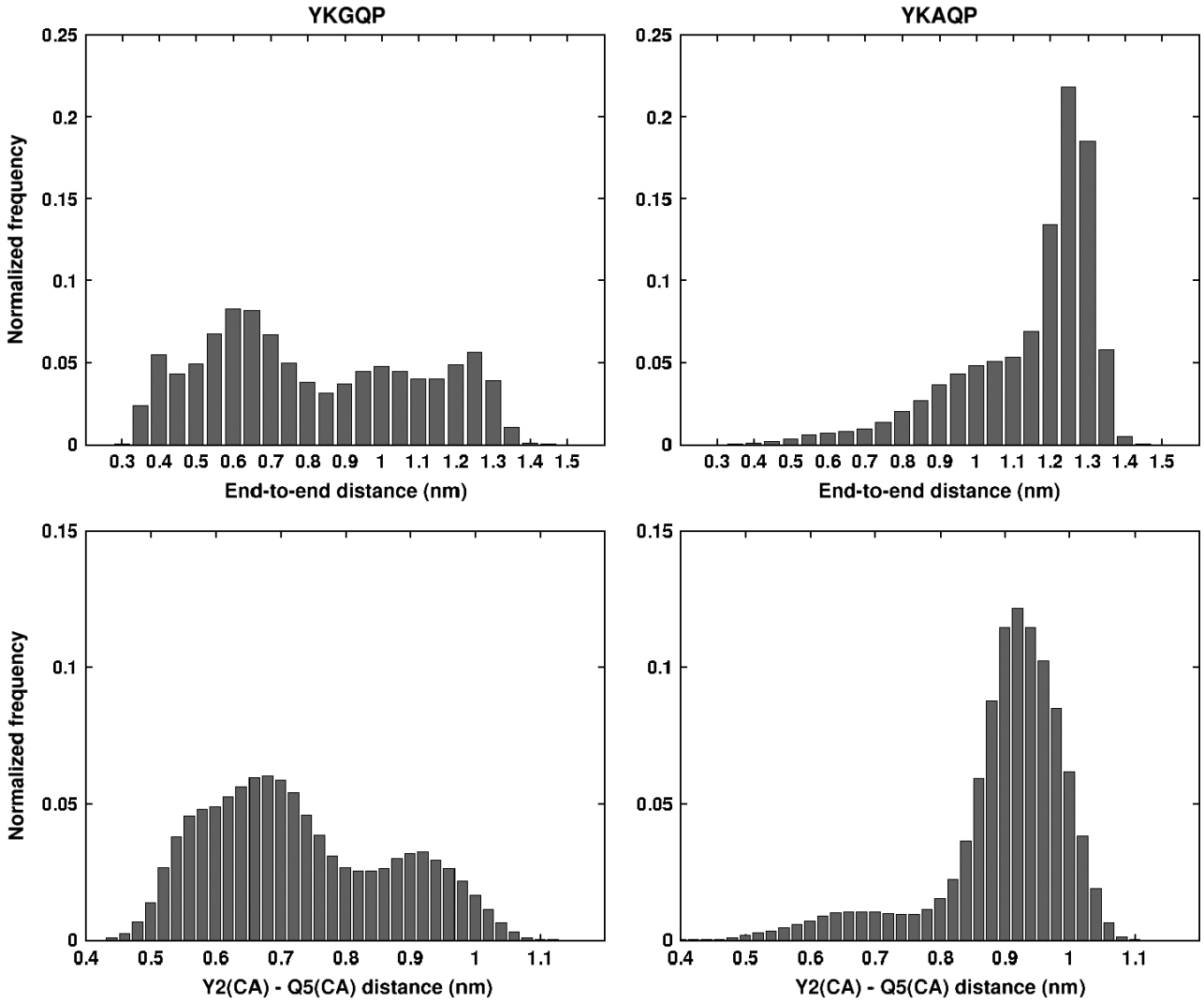

Figure 6 End-to-end (C^{α}_{Y2}–C^{α}_{P6}) and C^{α}_{Y2}–C^{α}_{Q5} distance distributions for YKGQP and YKAQP peptides in the simulations starting from polyproline II conformation. Note the suppression of shorter end-to-end distances and Y2(C^{α})—Q5(C^{α}) distances (<0.7 nm) in the YKAQP peptide, suggesting that YKGQP has greater loop/turn propensity.

sampled and can lead to the nucleation of hairpin formation as observed in the simulations of longer peptide DTVKLMYAGQPMTFR [25]. However, since type II turn conformation involving Y2-K3-G4-Q5 is sampled to a certain extent and since the free-energy difference between type II turn and Y2- and P6-associated conformational form is modest, the chance of turn conformation nucleating the hairpin formation is not negligible. There is some experimental ^{1}H NMR evidence for sampling of turn conformational forms for the sequence YKGQP [61]. In our earlier folding simulations of the hairpin peptide (DTVKLMYKGQPMTFR) [25], we found that the formation of contact between tyrosine and proline residues is an important event in the folding of the hairpin in two of the simulations. This correlates with the observation that Y2 and P6 residues establish contact in YKGQP also for a large part of the time as discussed earlier. Similarly, in another two simulations on the hairpin peptide DTVKLMYKGQPMTFR, formation of type II turn involving YKGQ is an important event during the folding of the hairpin (Figure 5). As discussed earlier, Gly in the $i+2$ position facilitates formation of a type II turn. Thus, the formation of Y2 and P6 contact or formation of type II turn involving YKGQP in hairpin folding is essentially a property of the sequence YKGQP. Since the formation of loop/turn involving YKGQ is important for the folding of the hairpin and since Gly plays a role in determining the formation of the loop/turn conformation, a mutant such as G4 → A4 should reduce the folding rate of the hairpin. Further, since it is shown both from simulation [25] and experiment [23,24] that the formation of hairpin is an early event in the refolding of nuclease, the G4 → A4 mutation should affect the folding of nuclease also. Thus, loop propensity of YKGQP driven by Gly is important for the folding of β-hairpin and therefore for the folding of nuclease.

DOI: 10.1002/psc

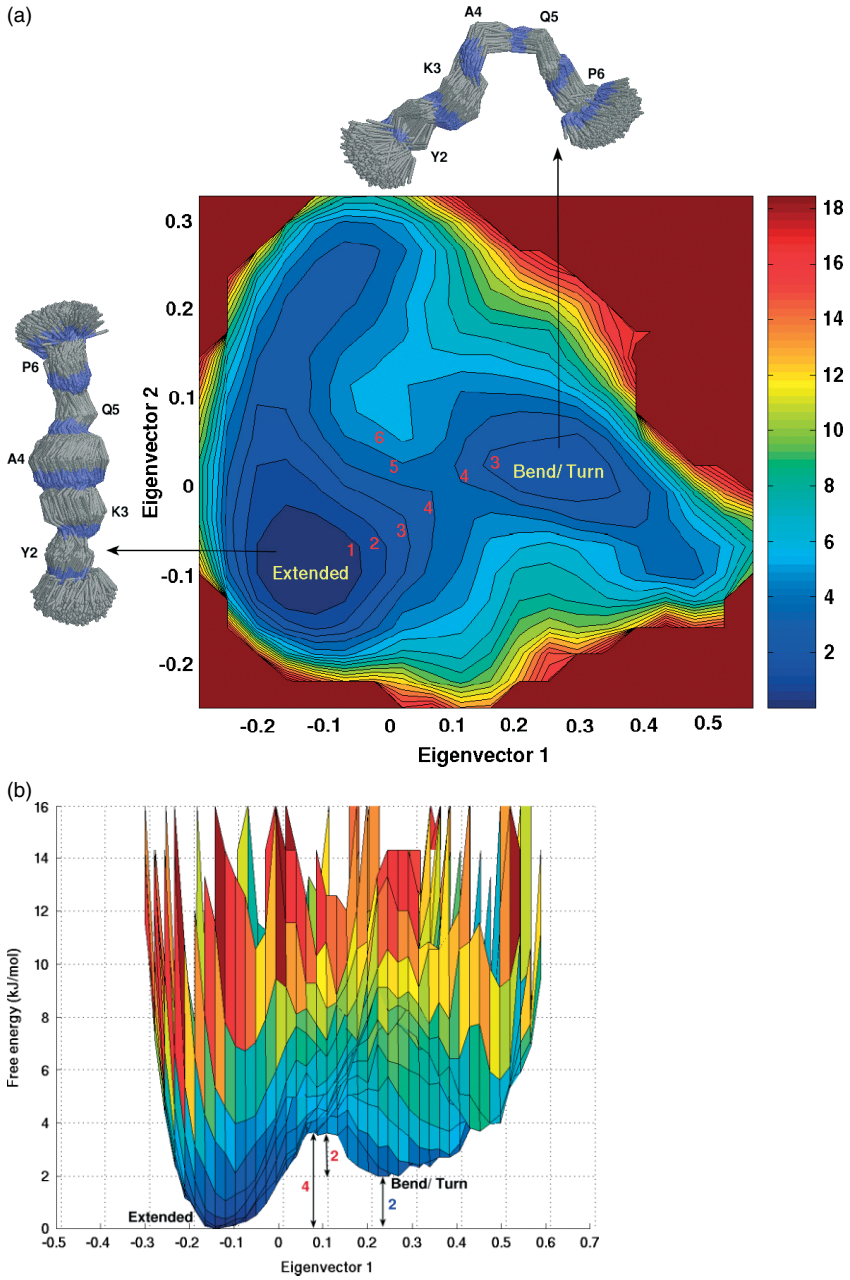

Figure 7 Free-energy landscape of the peptide YKAQP. (a) Free-energy landscape of the peptide YKAQP in the essential plane is defined by eigenvectors 1 and 2. Numerical values labeling the contour levels in the essential plane represent free energies in kJ/mol. Clusters of conformations sampled at the minima labeled 'Bend/Turn', Extended are shown. (b) Projection of the free-energy surface on eigenvector 1 for the peptide YKAQP. Free energies (in units of kJ/mol) of the minima and barriers between them are indicated.

General Implications for Protein Folding

From the free-energy landscape views of both the peptides YKGQP and YKAQP, it is clear that the minima in the free energies are sampled by well-defined conformational forms. The peptides, despite being small in length, do not sample conformations randomly. Instead, they sample 'folded' turn/loop structures or 'unfolded' extended structures in equilibrium with each other. Thus, there are conformational biases built in even in small peptides and these biases can be present when these small peptides are part of the parent proteins and the parent proteins are in an 'unfolded' state beginning to fold. Such biases ultimately lead the unfolded protein to a folded state.

CONCLUSIONS

The free-energy landscape of the peptide YKGQP shows minima corresponding to loop, turn and extended

DOI: 10.1002/psc

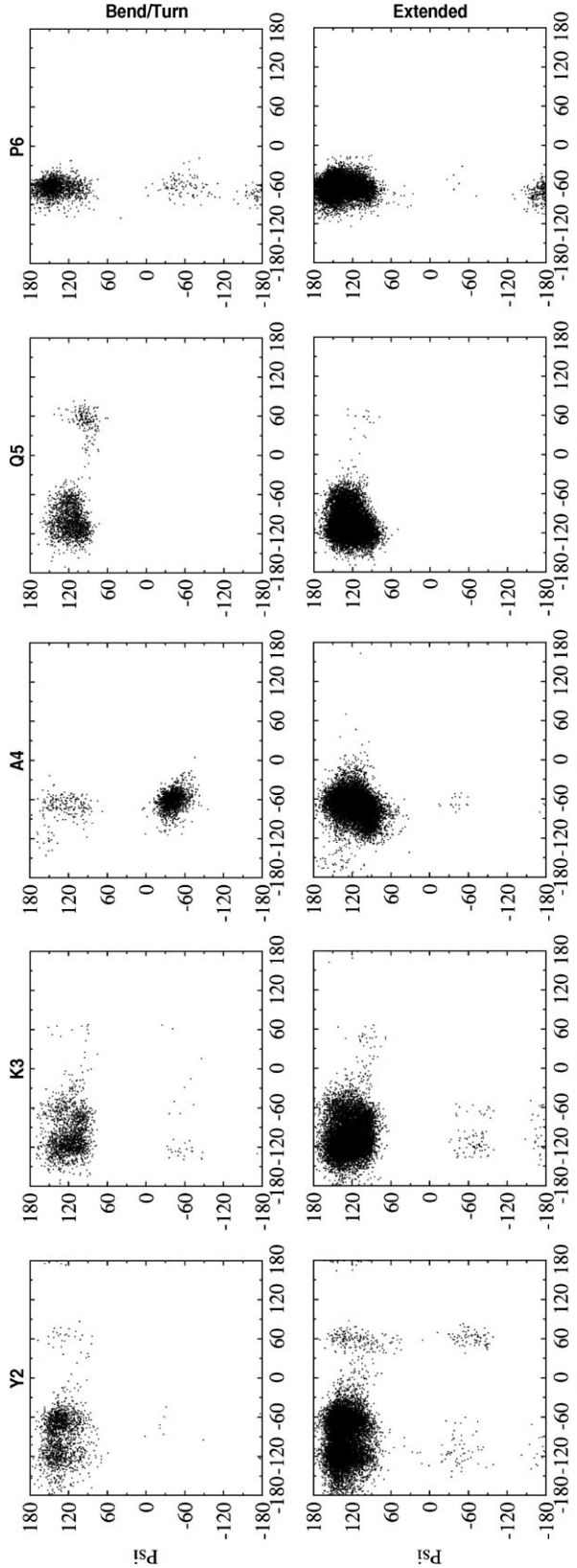

Figure 8 Ramachandran plot of the residues in the YKAQP peptide corresponding to free-energy minima labeled 'Bend/Turn' and 'Extended' (see Figure 7 also). Compare this figure to Figure 4. Bend/Turn minima samples type VIII turn $(\phi, \psi)_{i+1}$: (−60°, −30°) and $(\phi, \psi)_{i+2}$: (−120°, 120°) conformation with A4 and Q5 as corner residues.

DOI: 10.1002/psc

conformational forms. We find that Gly plays an important role in determining the conformational features of turn and loop conformers and thus the loop propensity of the sequence. This loop propensity is important for directing the folding of β-hairpin. Further, it is clear that the type II conformational preference of the peptide YKGQP is modulated to type I' conformation when the sequence is part of the hairpin peptide owing to interactions between flanking strands.

Acknowledgements

We thank BRNS for funding the research through a research grant (sanction number 2001/37/18/BRNS/796) sanctioned to YUS and a JRF fellowship to SP. We thank Prof. P. V. Balaji, School of Biosciences and Engineering, IIT Bombay, for many helpful discussions.

REFERENCES

1. Rose GD, Gierasch LM, Smith JA. Turns in peptides and proteins. *Adv. Protein Chem.* 1985; **37**: 1–109.
2. Fuller-Schaefer CA, Kadner RJ. Multiple extracellular loops contribute to substrate binding and transport by the Escherichia coli cobalamin transporter BtuB. *J. Bacteriol.* 2005; **187**: 1732–1739.
3. Helms LR, Wetzel R. Destabilizing loop swaps in the CDRs of an immunoglobulin VL domain. *Protein Sci.* 1995; **4**: 2073–2081.
4. Hynes TR, Kautz RA, Goodman MA, Gill JF, Fox RO. Transfer of a beta-turn structure to a new protein context. *Nature* 1989; **339**: 73–76.
5. Sibanda BL, Thornton JM. Conformation of beta hairpins in protein structures: classification and diversity in homologous structures. *Methods Enzymol.* 1991; **202**: 59–82.
6. Sibanda BL, Blundell TL, Thornton JM. Conformation of beta-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol.* 1989; **206**: 759–777.
7. Kwasigroch JM, Chomilier J, Mornon JP. A global taxonomy of loops in globular proteins. *J. Mol. Biol.* 1996; **259**: 855–872.
8. Karplus M, Weaver DL. Protein-folding dynamics. *Nature* 1976; **260**: 404–406.
9. Karplus M, Weaver DL. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.* 1994; **3**: 650–668.
10. Dill KA. Theory for the folding and stability of globular proteins. *Biochemistry* 1985; **24**: 1501–1509.
11. Dill KA. Dominant forces in protein folding. *Biochemistry* 1990; **29**: 7133–7155.
12. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding–a perspective from simple exact models. *Protein Sci.* 1995; **4**: 561–602.
13. Eaton WA, Munoz V, Thompson PA, Henry ER, Hofrichter J. Kinetics and dynamics of loops, a-helixes, b-hairpins, and fast-folding proteins. *Acc. Chem. Res.* 1998; **31**: 745–753.
14. Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, Garcia AE. Peptide folding simulations. *Curr. Opin. Struct. Biol.* 2003; **13**: 168–174.
15. Kubelka J, Hofrichter J, Eaton WA. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* 2004; **14**: 76–88.
16. Rotondi KS, Gierasch LM. Role of local sequence in the folding of cellular retinoic abinding protein I: structural propensities of reverse turns. *Biochemistry* 2003; **42**: 7976–7985.
17. McCallister EL, Alm E, Baker D. Critical role of beta-hairpin formation in protein G folding. *Nat. Struct. Biol.* 2000; **7**: 669–673.
18. Nauli S, Kuhlman B, Baker D. Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol.* 2001; **8**: 602–605.
19. Gu H, Kim D, Baker D. Contrasting roles for symmetrically disposed beta-turns in the folding of a small protein. *J. Mol. Biol.* 1997; **274**: 588–596.
20. Nagi AD, Anderson KS, Regan L. Using loop length variants to dissect the folding pathway of a four-helix-bundle protein. *J. Mol. Biol.* 1999; **286**: 257–265.
21. Huang F, Hudgins RR, Nau WM. Primary and secondary structure dependence of peptide flexibility assessed by fluorescence-based measurement of end-to-end collision rates. *J. Am. Chem. Soc.* 2004; **126**: 16665–16675.
22. Krieger F, Moglich A, Kiefhaber T. Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J. Am. Chem. Soc.* 2005; **127**: 3346–3352.
23. Jacobs MD, Fox RO. Staphylococcal nuclease folding intermediate characterized by hydrogen exchange and NMR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 1994; **91**: 449–453.
24. Walkenhorst WF, Edwards JA, Markley JL, Roder H. Early formation of a beta hairpin during folding of staphylococcal nuclease H124L as detected by pulsed hydrogen exchange. *Protein Sci.* 2002; **11**: 82–91.
25. Patel S, Sista P, Balaji PV, Sasidhar YU. Beta-Hairpins with native-like and non-native hydrogen bonding patterns could form during the refolding of staphylococcal nuclease. *J. Mol. Graphics Modell.* 2006; **25**: 103–115.
26. Loll PJ, Lattman EE. The crystal structure of the ternary complex of staphylococcal nuclease, Ca2+, and the inhibitor pdTp, refined at 1.65 A. *Proteins* 1989; **5**: 183–201.
27. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In *Inter Molecular Forces*, Pullman B (ed.). Reidel Publishing Company Dordrecht: The Netherlands, 1981; 331–342.
28. Lindahl E, Hess B, van der Spoel D. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* 2001; **7**: 306–317.
29. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. vdf Hochschulverlag AG an der ETH Zürich: Groningen, 1996.
30. Scott WRP, Huenenberger P, Tironi I, Mark A, Billeter S, Fennen J, Torda A, Huber T, Krueger P, van Gunsteren W. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* 1999; **103**: 3596–3607.
31. Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L. A smooth particle mesh Ewald method. *J. Chem. Phys.* 1995; **103**: 8577–8593.
32. Darden T, York D, Pedersen L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 1993; **98**: 10089–10092.
33. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 1997; **18**: 1463–1472.
34. Berendsen H, Postma J, DiNola A, Haak J. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 1984; **81**: 3684–3690.
35. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983; **22**: 2577–2637.
36. Wilmot CM, Thornton JM. Analysis and prediction of the different types of beta-turn in proteins. *J. Mol. Biol.* 1988; **203**: 221–232.
37. Lewis PN, Momany FA, Scheraga HA. Chain reversals in proteins. *Biochim. Biophys. Acta* 1973; **303**: 211–229.
38. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J. Mol. Graph.* 1996; **14**: , 33–38 27–28.
39. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. *Proteins* 1993; **17**: 412–425.

40. de Groot BL, Amadei A, Scheek RM, van Nuland NA, Berendsen HJ. An extended sampling of the configurational space of HPr from E. coli. *Proteins* 1996; **26**: 314–322.
41. Daidone I, Amadei A, Di Nola A. Thermodynamic and kinetic characterization of a beta-hairpin peptide in solution: an extended phase space sampling by molecular dynamics simulations in explicit water. *Proteins* 2005; **59**: 510–518.
42. Amadei A, Ceruso MA, Di Nola A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins* 1999; **36**: 419–424.
43. Merlino A, Vitagliano L, Ceruso MA, Di Nola A, Mazzarella L. Global and local motions in ribonuclease A: a molecular dynamics study. *Biopolymers* 2002; **65**: 274–283.
44. Chou PY, Fasman GD. Beta-turns in proteins. *J. Mol. Biol.* 1977; **115**: 135–175.
45. Sibanda BL, Thornton JM. Beta-hairpin families in globular proteins. *Nature* 1985; **316**: 170–174.
46. Rotondi KS, Gierasch LM. Local sequence information in cellular retinoic acid-binding protein I: specific residue roles in b-turns. *Biopolymers* 2003; **71**: 638–651.
47. Dyson HJ, Rance M, Houghten RA, Lerner RA, Wright PE. Folding of immunogenic peptide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* 1988; **201**: 161–200.
48. Hutchinson EG, Thornton JM. A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* 1994; **3**: 2207–2216.
49. Gunasekaran K, Gomathi L, Ramakrishnan C, Chandrasekhar J, Balaram P. Conformational interconversions in peptide beta-turns: analysis of turns in proteins and computational estimates of barriers. *J. Mol. Biol.* 1998; **284**: 1505–1516.
50. Wu X, Wang S. Folding studies of a linear pentamer peptide adopting a reverse turn conformation in aqueous solution through molecular dynamics simulation. *J. Phys. Chem. B* 2000; **104**: 8023–8034.
51. Scully J, Hermans J. Backbone flexibility and stability of reverse turn conformation in a model system. *J. Mol. Biol.* 1994; **235**: 682–694.
52. Yang AS, Hitz B, Honig B. Free energy determinants of secondary structure formation: III. beta-turns and their role in protein folding. *J. Mol. Biol.* 1996; **259**: 873–882.
53. Chothia C. Conformation of twisted beta-pleated sheets in proteins. *J. Mol. Biol.* 1973; **75**: 295–302.
54. Mattos C, Petsko GA, Karplus M. Analysis of two-residue turns in proteins. *J. Mol. Biol.* 1994; **238**: 733–747.
55. Tobias DJ, Mertz JE, Brooks CL. 3r Nanosecond time scale folding dynamics of a pentapeptide in water. *Biochemistry* 1991; **30**: 6054–6058.
56. Demchuk E, Bashford D, Case DA. Dynamics of a type VI reverse turn in a linear peptide in aqueous solution. *Fold. Des.* 1997; **2**: 35–46.
57. Mohanty D, Elber R, Thirumalai D, Beglov D, Roux B. Kinetics of peptide folding: computer simulations of SYPFDV and peptide variants in water. *J. Mol. Biol.* 1997; **272**: 423–442.
58. Galzitskaya O, Caflisch A. Solution conformation of phakellistatin 8 investigated by molecular dynamics simulations. *J. Mol. Graphics Modell.* 1999; **17**: 19–27.
59. Fuchs PF, Bonvin AM, Bochicchio B, Pepe A, Alix AJ, Tamburro AM. Kinetics and thermodynamics of type VIII beta-turn formation: a CD, NMR, and microsecond explicit molecular dynamics study of the GDNP tetrapeptide. *Biophys. J.* 2006; **90**: 2745–2759.
60. Nakajima N, Higo J, Kidera A, Nakamura H. Free energy landscapes of peptides by enhanced conformational sampling. *J. Mol. Biol.* 2000; **296**: 197–216.
61. Ramakrishna V, Sasidhar YU. A pentapeptide model for an early folding step in the refolding of staphylococcal nuclease: the role of its turn propensity. *Biopolymers* 1997; **41**: 181–191.

DOI: 10.1002/psc