Journal of Molecular Graphics and Modelling 40 (2013) 10–21

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

The role of loop closure propensity in the refolding of Rop protein probed by molecular dynamics simulations

Rashmi Tambe Shukla [a], Chetana Baliga [b,1], Yellamraju U. Sasidhar [a,*]

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

ARTICLE INFO

Article history:
Accepted 27 December 2012
Available online 4 January 2013

Keywords:
Dimeric-intermediate
Helix-turn-helix
Kinetic model
Loop closure
Protein folding

ABSTRACT

Rop protein is a homo-dimer of helix-turn-helix and has relatively slow folding and unfolding rates compared to other dimeric proteins of similar size. Fluorescence studies cited in literature suggest that mutation of turn residues D30-A31 to G30-G31 (Gly_2) increases its folding and unfolding rates considerably. A further increase in number of glycines in the turn region results in decrease of folding rates compared to Gly_2 mutant. To understand the effect of glycine mutation on folding/unfolding rates of Rop and the conformational nature of turn region involved in formation of early folding species, we performed molecular dynamics simulations of turn peptides, ^{25}KLNEL**DA**DEQ^{34} (**DA** peptide), ^{25}KLNELGGDEQ^{34} (**G_2** peptide), ^{25}KLNEL**GGG**DEQ^{35} (**G_3** peptide) and ^{25}KLNEL**GGG**EQ^{34} (**$G_{3'}$** peptide) from Rop at 300 K. Further **Wt-Rop** and mutant **G_2-Rop** monomers and dimers were also studied separately by molecular dynamics simulations. Our results show that glycine based peptides (G_n peptides) have a higher loop closure propensity compared to **DA**. Comparison of monomeric and dimeric Rop simulations suggests that dimeric Rop necessarily requires α_L conformation to be sampled at D30/G30 position in the turn region. Since glycine (at position 30) can readily adopt α_L conformation, G_n loop plays a dual role in both facilitating loop closure as well as facilitating reorganization/packing of helices required for structural adjustment during dimer formation in the folding of Rop. Based on our simulation results and available literature, we suggest a tentative kinetic model for Rop folding which allows us to estimate the contribution of loop closure propensity to the overall folding rates.

1. Introduction

The protein Rop (repressor of primer) has extremely slow folding and unfolding rates compared to other proteins of similar size despite the absence of rate limiting factors like complex topology, disulphide linkages, prolines and cofactors [1–4]. The structure of Rop has been elucidated by X-ray crystallography [5] and solution NMR [6]. It is a homo-dimer of helix-turn-helix monomers with all helices anti-parallel to each other (Fig. 1A) and has a molecular weight of around 15 kDa. It has a tight turn formed by residues D30 and A31 where D30 is found in α_L conformation. Stopped-flow fluorescence and stopped-flow circular dichroism (CD) studies show that Rop has biphasic kinetics when refolding to low concentrations of GnHCl with fast ($k_f \sim 29\ s^{-1}$) and slow ($k_f \sim 0.5\ s^{-1}$) kinetic phases. At higher GnHCl concentrations refolding is monophasic. Unfolding is slow and monophasic at all concentrations of denaturant with a $k_u \sim 10^{-7}\ s^{-1}$ at 25 °C [7–9]. Stopped-flow fluorescence studies in which D30-A31 in Rop turn are substituted with G30-G31 (Gly_2) linker showed 20-fold and 62-fold increase in rates of fast and slow phases respectively in the mutant compared to wild type (wt) protein [8]. A possible reason for this altered kinetics in G30-G31 mutant could be the stabilization of transition state during folding. The folding rates of loop G_n mutants (where $n = 2–9$) of Rop are in general higher than those of wt-Rop, suggesting an important role played by loop in folding. Further within G_n loop mutants the folding rate decreases with the increasing length of glycine linker indicating the effect of loop closure on folding [10,11]. Most of the core mutants of Rop have higher folding and unfolding rates and higher thermodynamic stability compared to wild type [7,12–14]. This indicates that both loop closure and hydrophobic core packing play a role during the folding of Rop. Real-time NMR study on Rop suggests that in the fast phase an intermediate is present which has both native and non-native characteristics [15].

From the refolding of Rop studied by Munson et al. [7] and Nagi et al. [8] it could be inferred that the fast phase of folding corresponds to the formation of a dimeric-intermediate with putative helical hairpin structures and exposed surface area. Further it was shown that the intermediate forms cooperatively

* Corresponding author. Tel.: +91 22 2576 7179; fax: +91 22 2576 7152.
E-mail address: sasidhar@chem.iitb.ac.in (Y.U. Sasidhar).
[1] Present address: Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India.

1093-3263/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.12.007

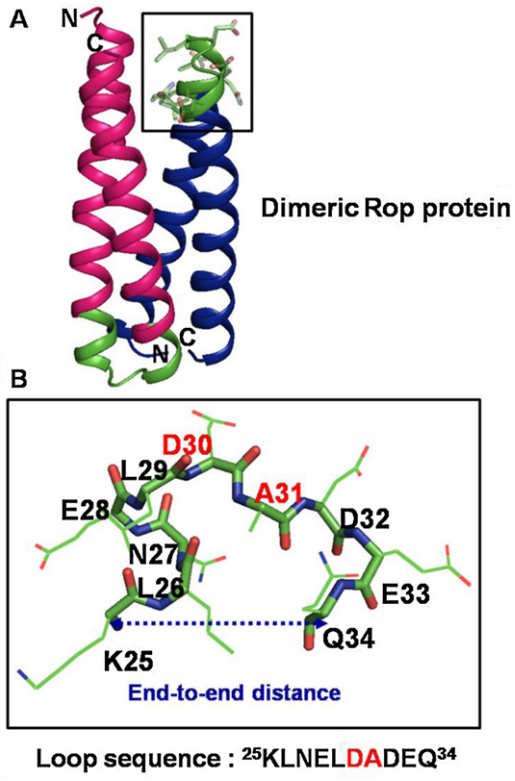


Fig. 1. The crystal structure of dimeric four-helix-bundle Rop protein (1ROP) is shown in (A) and the loop region is indicated in green color in both the helix-turn-helix monomers. An enlarged image of the loop region (considered in the peptide simulations) is shown in (B). Main chain is indicated in stick representation and side chains are shown in line representation. The end-to-end distance is the distance between C-α atoms of K25 (or K2 in peptide simulation) and Q34 (or Q11/Q12 in peptide simulations) residues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and is destabilized significantly with respect to wt protein [8]. Therefore one could envisage that in the intermediate state the monomer–monomer interactions are relatively weak and consolidation of loop and helical structures were to follow with the development of native structure. The exact conformational nature of the early folding species involving loop closure and formation of putative helices is unknown.

From the above discussion it is clear that loop region of Rop is involved in the folding of Rop since Gly mutation in this region affects both fast and slow kinetic phases of folding. Further there is evidence that hydrophobic core packing affects the folding. These observations raise questions about the nature of the role played by the loop region of Rop in its folding: will it contribute to the folding of the monomer? Will it play a role in the formation of dimeric-hairpin intermediate? Will it play a role in the packing of monomers to form the native dimer? Our objective in this paper is to address some of these questions. To this end we seek to understand if higher loop closure propensity could contribute to faster folding. Further, we also seek to understand how loop mutant(s) could affect the packing interaction between the monomers in dimeric state.

During early folding, only local interactions operate to a significant extent. Non-local tertiary interactions would not have developed during early folding times. Short peptides separated from the protein context, model the conformational features where only local interactions are dominant as would be the case in the unfolded state beginning to fold. Hence, we considered the wt loop peptide ^{25}*KLNEL***DA***DEQ*^{34} (**DA** peptide) (Fig. 1B) and its glycine mutants ^{25}*KLNEL***GG***DEQ*^{34} (**G_2** peptide), ^{25}*KLNEL***GGG***DEQ*^{35} (**G_3** peptide) and ^{25}*KLNEL***GGG***EQ*^{34} (**$G_{3'}$** peptide) from Rop. The peptides **DA**, **G_2**, **G_3** and **$G_{3'}$** were simulated for 500 ns each at 300 K. While experimental data is available for Rop protein involving **DA**, **G_2** and **G_3** loop sequences, but no such data is available for the protein involving **$G_{3'}$** loop sequence. **$G_{3'}$** sequence is studied, in addition to **G_3**, mainly to check for consistency of results in peptides where three glycines are present in the loop region. In this work we have also modeled monomeric and dimeric states of Rop to understand the conformational features of the end points of the folding. In particular, the conformational features of monomeric state would give us some hints about the factors required for dimer formation. This would also provide an understanding of how wt and mutant loop sequences affect the reorganization/packing of helices during Rop folding. Hence, we studied the dynamics of both **Wt-Rop** and **G_2-Rop** monomers and dimers separately. Comparison of protein dynamics of Rop in monomeric and dimeric states would allow us to infer how changes in loop sequence affects the monomer–monomer interactions in dimeric Rop. Thus, **Wt-Rop** and **G_2-Rop** monomers were simulated for 200 ns each at 300 K and **Wt-Rop** and **G_2-Rop** dimers were simulated for 60 ns each at 300 K.

In the peptide simulations, we find that loop formation is more probable in G_n loop peptides (**G_2**, **G_3** and **$G_{3'}$**) in comparison to **DA** peptide and loop formation in G_n peptides is largely characterized by formation of non-native persistent main chain hydrogen bonds. On the other hand, monomer and dimer simulations of **Wt-Rop** and **G_2-Rop** point to the role played by the loop in the packing of helices during the slow phase of folding of Rop. These simulations show that Gly_2 mutant loop can easily meet the conformational requirements (α_L conformation at position 30) in dimeric state, in comparison to the wt-loop.

2. Methods

2.1. Simulation systems

2.1.1. Peptide systems

The coordinates for loop peptides ^{25}KLNEL**DA**DEQ^{34} (designated as **DA** peptide), ^{25}KLNEL**GG**DEQ^{34} (designated as **G_2** peptide), ^{25}KLNEL**GGG**DEQ^{34} (designated as **G_3** peptide) and ^{25}KLNEL**GGG**EQ^{34} (designated as **$G_{3'}$** peptide) were taken from the crystal structure of Rop (PDB: 1ROP) [5]. **G_2**, **G_3** and **$G_{3'}$** (G_n) peptide coordinates were generated by mutating the corresponding residues in the **DA** peptide to glycines using DeepView software (http://www.expasy.org/spdbv/) [16]. The ends of the two peptides were protected using acetyl (CH_3—CO) group at the N-terminus and N-methyl (NH—CH_3) group at the C-terminus. All the four peptides were simulated with a starting polyproline II (PPII) conformation. Studies on short poly-alanine and poly-lysine peptides in water show that PP II is the conformation for unfolded state [17–20]. PP II conformation was generated by setting all the backbone dihedral angles in both the peptides as $\varphi = -76^\circ$ and $\psi = 149^\circ$ using DeepView software.

2.1.2. Protein systems

Coordinates of **Wt-Rop** and **G_2-Rop** monomers and **Wt-Rop** and **G_2-Rop** dimers were obtained from protein data bank (PDB: 1ROP). **G_2-Rop** monomer and dimer coordinates were prepared by mutating D30 and A31 residues to glycines using DeepView software as mentioned previously. The initial structure in all four protein simulations was the crystal structure of Rop monomer and dimer, respectively.

All peptide and protein simulations were performed in a truncated octahedron with an image distance (d) of 5.4 nm for peptide systems and 8.2 nm for protein monomer and dimer systems. The

image distance specifies both box vectors and angles between them [21]. Peptides **DA**, **G_2**, **G_3** and **$G_{3'}$** were solvated using 3931, 3932, 3925 and 4170 SPC water molecules [22]. **Wt-Rop** and **G_2-Rop** monomer systems were solvated with 13,914 and 13,921 SPC water molecules respectively and corresponding **Wt-Rop** and **G_2-Rop** dimers were solvated with 13,627 and 13,635 SPC water molecules, respectively. The total number of atoms in **DA**, **G_2**, **G_3** and **$G_{3'}$** peptide systems were 11,900, 11,897, 11,881 and 12,606 atoms, respectively. The system size for **Wt-Rop** and **G_2-Rop** monomer systems was 42,332 and 42,345 atoms, respectively and that for **Wt-Rop** and **G_2-Rop** dimers was 43,061 and 42,069 atoms, respectively. Net charges on **DA**, **G_2**, **G_3** and **$G_{3'}$** peptides were −3, −2, −2 and −1, respectively; while the net charges on **Wt-Rop** monomer, **G_2-Rop** monomer, **Wt-Rop** dimer and **G_2-Rop** dimer were −4, −3, −8 and −6, respectively. The charges in all the simulation systems were neutralized using counter ions (Na^+ ions) by replacing corresponding number of water molecules.

2.2. MD simulation parameters

MD simulations were performed using GROMACS software package version 4.0.4 [23] and united atom force field GROMOS96 ffG43a1 [24] on dual xeon quad-core processor based machines with Cent OS 4.3 operating system (http://www.centos.org/) installed on them. The electrostatic interactions were treated using Particle Mesh Ewald (PME) algorithm [25,26] with a coulomb cut-off of 1.1 nm, PME interpolation order of 4.0 and Fourier spacing of 0.12 nm. The van der Waal's interactions were treated using Lennard–Jones potential and switching function [21] with a cutoff of 1.1 nm and a switching distance of 0.8 nm. Peptide and protein systems were subjected to energy minimization after they were solvated with SPC water molecules and neutralized using counter ions. Energy minimization was carried out using steepest descent algorithm for around 400 steps with a force tolerance value of 100 kJ mol^{-1} nm^{-1}. The potential energy was lowered within the specified upper limit of 400 steps. To obtain convergence with respect to the tolerance value mentioned above, conjugate gradient method was subsequently used. Subsequent to energy minimization, position restrained MD was performed on peptide and protein systems at 300 K. The peptide coordinates were fixed and water was simulated around it to equilibrate the system and to remove any solvent holes. Position restrained MD was performed in four different steps of 50 ps each, with the force constant in each simulation step being 1000 kJ mol^{-1} nm^{-1}, 100 kJ mol^{-1} nm^{-1}, 10 kJ mol^{-1} nm^{-1} and 0 kJ mol^{-1} nm^{-1} [27]. After equilibration, the final productive MD was performed. For **DA**, **G_2**, **G_3** and **$G_{3'}$** peptides the final productive MD was performed at 300 K for 500 ns each. The protein simulations on **Wt-Rop** and **G_2-Rop** monomers were performed at 300 K for 200 ns each and on **Wt-Rop** and **G_2-Rop** dimers at 300 K for 60 ns each. LINCS algorithm [28] was used to constrain all bonds. A time step of 2 fs was used to integrate the equation of motion. The positions and the velocities of the atoms were collected at every 0.5 ps or 250 steps in the trajectory file. Peptide (or protein) and solvent (including ions) were coupled to separate temperature baths using Berendsen's thermostat [29] with a time constant of 0.1 ps at 300 K. Pressure coupling was also done using Berendsen algorithm with a time constant of 1.0 ps and a reference pressure of 1 bar was used. Analysis tools incorporated in GROMACS software package (version 4.0.4) were used to analyze the results of all the simulation systems. All the graphs and figures were made using Matlab (http://www.mathworks.com), Xmgrace (http://plasma-gate.weizmann.ac.il/Grace/) and PyMol (http://www.pymol.org/).

2.3. Convergence of conformational sampling in peptide and protein systems

The convergence in the peptide simulation trajectories of **DA**, **G_2**, **G_3** and **$G_{3'}$** was checked in terms of the following two quantities: (a) the number of intra-peptide hydrogen bonds with simulation time and (b) the frequencies of first two most populated clusters with respect to simulation time. Both these quantities seem to converge with time indicating conformational equilibrium for the four peptide simulations (Figure S1 of Supporting information). **Wt-Rop** and **G_2-Rop** monomer simulations equilibrate in the region from 100 to 200 ns based on the variation of root mean square deviation (RMSD) of backbone and the variation of radius of gyration (Rg) as a function of simulation time (Figure S2 of Supporting information). Hence all the analyses in these two monomer simulations were performed for the equilibrium period only. The **Wt-Rop** and **G_2-Rop** dimer systems also show equilibration in terms of RMSD and Rg with respect to simulation time (Figure S3 of Supporting information), except for a fluctuation in RMSD during 40–60 ns in **G_2-Rop** dimer trajectory which is mainly due to a dynamic N-terminus in monomer 1 of **G_2-Rop** dimer (Figure S3B of Supporting information). Hence for dimer simulations complete trajectory was used in all the analyses.

2.4. Criterion for loop formation

A four residue reverse turn is considered to be present if the end-to-end distance between C-α atoms of first and fourth residues is less than or equal to 0.7 nm or 7 Å [30–32]. We considered that for the peptides **DA**, **G_2**, **G_3** and **$G_{3'}$**, a loop conformation is present if the distance between C-α atoms of K2 and Q11/Q12 residues is less than or equal to 0.7 nm. This distance was designated as 'end-to-end distance'. A value of 0.7 nm or less for end-to-end distance in peptide simulations was considered as a measure of loop closure as it is reasonable in view of the distribution of end-to-end distances in loops found in proteins [33–36]. For monomer and dimer systems of **Wt-Rop** and **G_2-Rop**, distance between C-α atoms of K25 (or K81) and Q34 (or Q90) residues was designated as the 'loop end-to-end distance'.

2.5. Secondary structure analysis

Secondary structure analysis was performed using the DSSP program [37] incorporated in the GROMACS version 4.0.4. According to the DSSP algorithm, a coil conformation is the one with no hydrogen bonds and with absence of any curvature in the backbone. On the other hand, bends are regions with high curvature of backbone. It is the conformation at ith residue formed by $i-2$, $i-1$, i, $i+1$ and $i+2$ residues with a curvature angle at the backbone at ith position of at least 70° [37]. Bend has no hydrogen bonds. n-Turn is recognized with a single hydrogen bond. If an n-turn starts at ith residue, then there is a hydrogen bond between backbone CO (i) and NH ($i+n$) where value of $n = 3, 4, 5$.

2.6. Hydrogen bond analysis

The program g_hbond from GROMACS version 4.0.4 was used to calculate the variation of number of hydrogen bonds as a function of simulation time. We calculated total intra-peptide hydrogen bonds in peptide systems **DA**, **G_2**, **G_3** and **$G_{3'}$**. All possible donor–hydrogen–acceptor triplets were determined and the percentage of time a hydrogen bond existed between them during the simulation was calculated using perl, matlab and shell scripts. If a particular hydrogen bond existed for more than or equal to 10% of the total simulation time, it was designated as a '**persistent**' hydrogen bond. We further extended our analysis on these

persistent hydrogen bonds by calculating the percentage of the time they co-existed with each other. In the case of protein monomer and dimer simulations, intra-protein and protein-solvent hydrogen bonds were calculated for complete protein, only loop region (K25-Q34) and only helices. Hydrogen bond formation was determined by considering a cutoff of 3.5 Å for the donor–acceptor distance and a cutoff of 30° for the acceptor–donor–hydrogen angle.

2.7. Cluster analysis

Clustering was performed on peptide systems using g_cluster program and gromos algorithm [38] incorporated in GROMACS 4.0.4 package. As per this algorithm, the conformational pool in the peptide simulations was divided into different subsets with members in a subset having RMSD of around 0.3 nm (average pair-wise RMSDs for **DA**, **G_2**, **G_3** and **$G_{3'}$** peptides are 0.31 nm, 0.30 nm, 0.36 nm and 0.32 nm, respectively) with respect to the central conformation. Based on this cutoff, the cluster with maximum number of members was eliminated from the pool and the steps were repeated to generate remaining clusters. We have used the variation of the frequencies of cluster 1 and cluster 2 with respect to simulation time as a measure of convergence in peptide simulations (Figure S1E–H of Supporting information and also refer to Section 2.3). Cluster analysis was also performed on protein monomer and dimer systems using g_cluster program and gromos algorithm. The pair-wise RMSD distributions in monomer systems were used to compare the conformational variability.

3. Results

3.1. All four loop peptides sample compact conformations, but a more favorable loop closure is observed for G_n peptides (**G_2**, **G_3** and **$G_{3'}$**) in comparison to **DA** peptide

The peptides **DA**, **G_2**, **G_3** and **$G_{3'}$** become compact starting from an initial PP II conformation indicated with a drop in the Rg value from 1.0 nm to 0.5 nm in initial 5 ns of simulations (Figure S4 of Supporting information). As mentioned in Section 2.4, the end-to-end distance with a value of less than or equal to 0.7 nm in the peptide simulations was considered as a measure of loop closure. Normalized and cumulative frequency distribution plots for end-to-end distances for the peptide systems **DA**, **G_2**, **G_3** and **$G_{3'}$** are shown in Fig. 2A–D. The loop closure propensities in **DA**, **G_2**, **G_3** and **$G_{3'}$** peptides are 36%, 61%, 56% and 58%, respectively. Thus, all G_n mutants show a greater loop closure propensities with end-to-end distance less than or equal to 0.7 nm, in comparison to **DA** peptide. The average values for end-to-end distance in **DA**, **G_2**, **G_3** and **$G_{3'}$** peptide simulations are 0.8 nm, 0.7 nm, 0.7 nm and 0.7 nm, respectively (Fig. 2A–D). Mean values for the distribution of loop end-to-end distances in the equilibrium period (100–200 ns) of **Wt-Rop** monomer (0.8 nm) and **G_2-Rop** monomer (0.7 nm) simulations (Fig. 2E and F) are found to be compatible with the mean values of end-to-end distances sampled in **DA** and **G_2** peptide simulations (Fig. 2A and B). On the other hand, in case of dimer simulations greater values for loop end-to-end distance (~1.0 nm) is sampled (Fig. 2G–J) compared to those in peptide and monomer simulations.

The variation of secondary structure as a function of time and the conformations sampled at different times in the trajectories of **DA**, **G_2**, **G_3** and **$G_{3'}$** peptides is shown in Fig. 3. In all four simulations, peptides adopt compact conformations (also refer to variation of Rg in Figure S4 of Supporting information) but in the case of **G_2**, **G_3** and **$G_{3'}$** peptides, there is a greater occurrence of such conformations which could lead to loop closure (encircled conformations in Fig. 3) and such conformers have a number of backbone hydrogen bonds. Hydrogen bond analysis performed on **DA**, **G_2**, **G_3** and **$G_{3'}$**

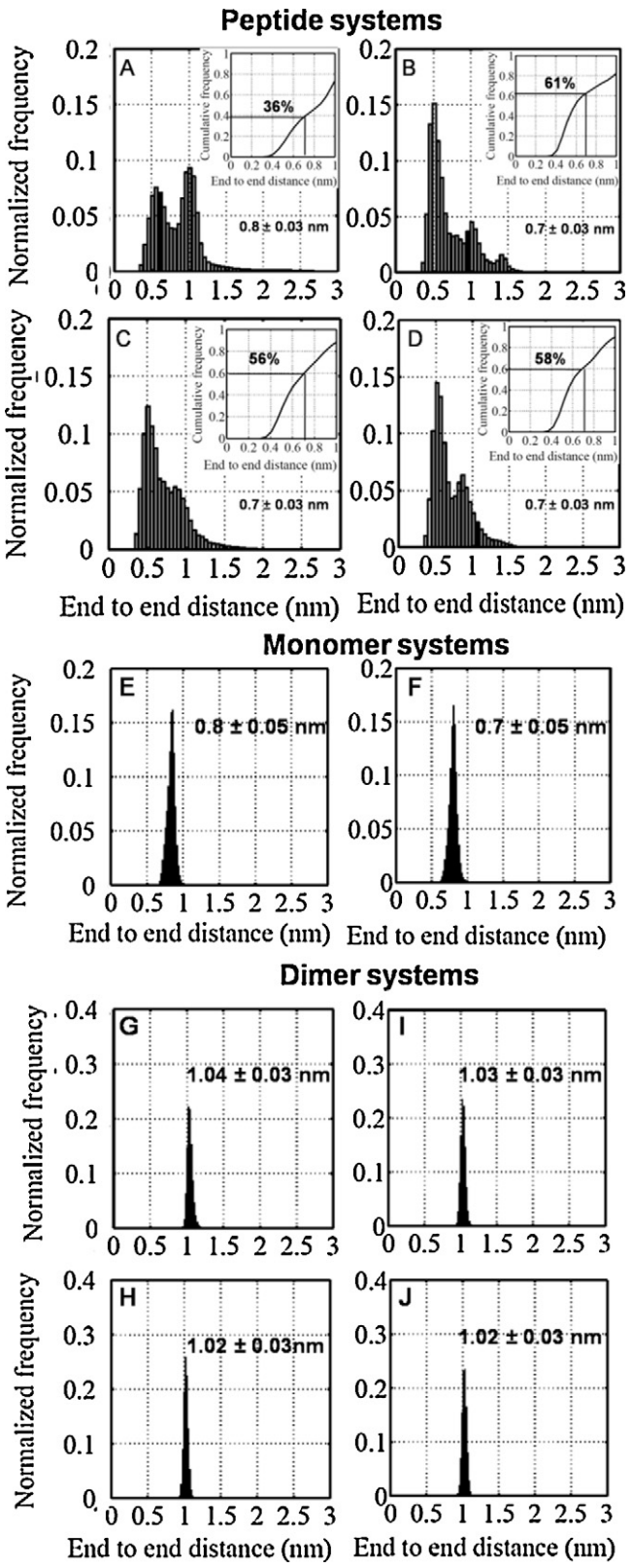


Fig. 2. Normalized frequency distribution and cumulative frequency plots (in inset) for the end-to-end distance (distance between C-α atoms of K2 and Q11/Q12 residues) for the peptides **DA**, **G_2**, **G_3** and **$G_{3'}$** are shown in (A), (B), (C) and (D), respectively. The distribution of loop end-to-end distance (distance between C-α atoms of K25 and Q34 residues) in monomer simulations of **Wt-Rop** and **G_2-Rop** is shown in (E) and (F), respectively. The distribution for loop end-to-end distance (distance between C-α atoms of K25 (or K81) and Q34 (or Q90) residues) for each monomer of **Wt-Rop** dimer is shown in (G) and (H) and for monomers of **G_2-Rop** dimer is shown in (I) and (J), respectively. The mean values and standard deviations for the end-to-end distance in each case are also indicated.

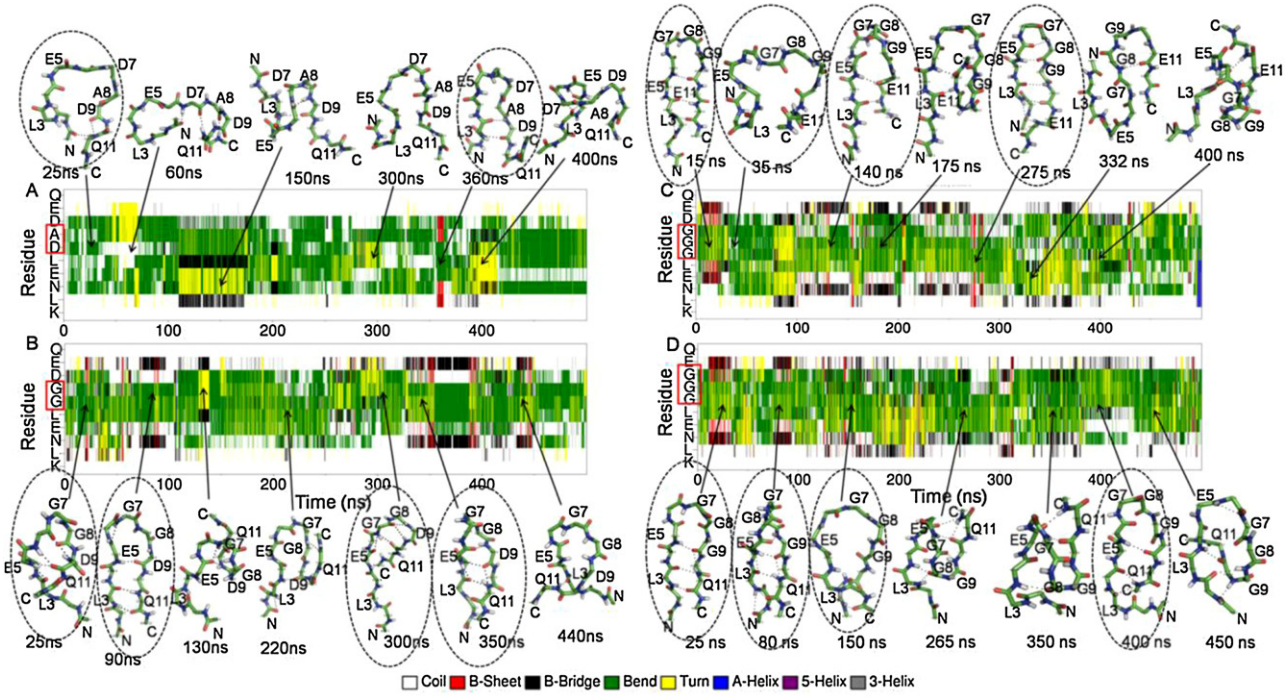


Fig. 3. Variation of secondary structure with time in the simulation for **DA** peptide is shown in (A), for $\mathbf{G_2}$ peptide is shown in (B), for $\mathbf{G_3}$ peptide is shown in (C) and for $\mathbf{G_3}$ peptide is shown in (D). The color codes for different secondary structural elements are shown in the legend at the bottom of the figure. Snapshots of conformations sampled at different times in **DA** and G_n mutant peptide trajectories are also shown. The backbone is shown in stick representation and backbone hydrogen bonds are shown in dashed lines. The N and C-termini of the peptides are indicated. The conformations which have ends of the peptides together are encircled by a dotted line.

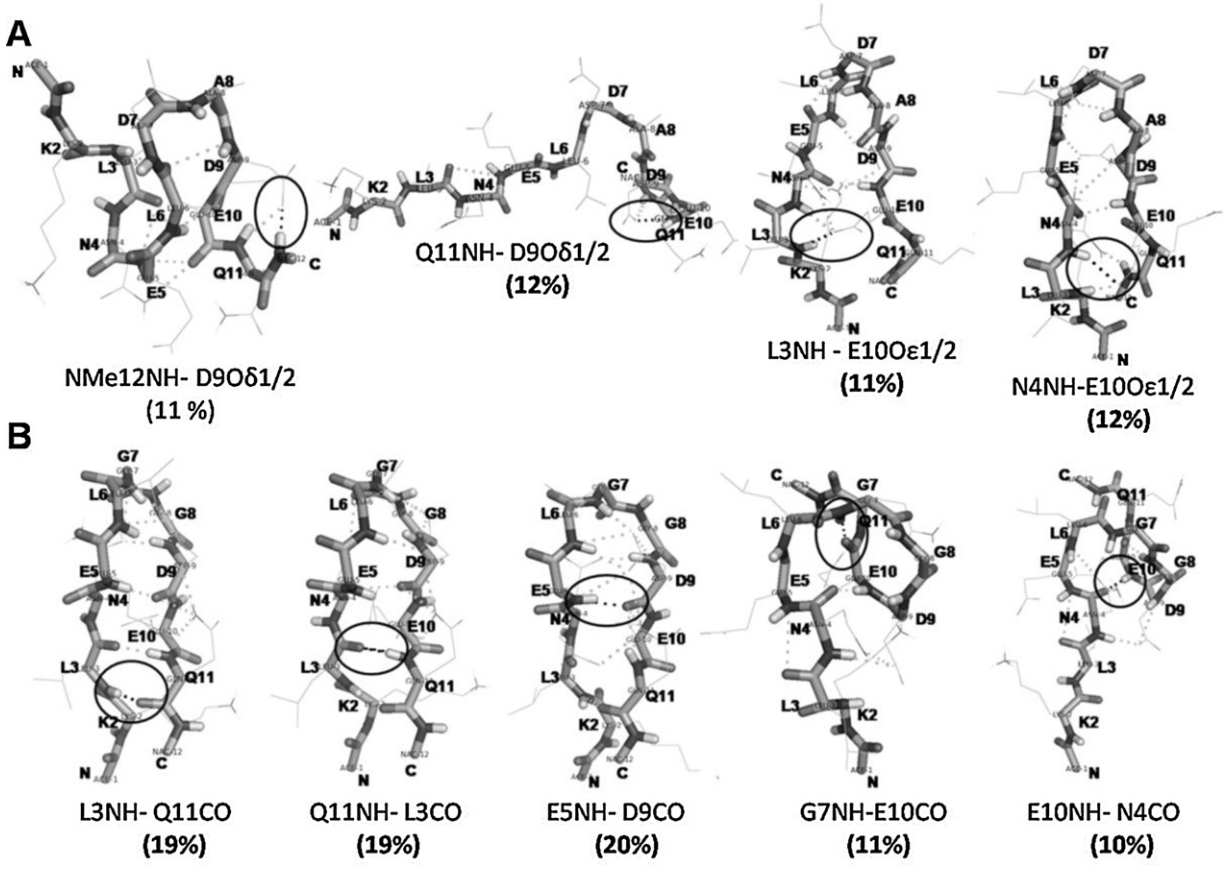


Fig. 4. Conformations with persistent hydrogen bonds formed during the simulations of **DA** peptide are shown in (A) and those in case of $\mathbf{G_2}$ peptide are shown in (B). The conformations shown are labeled with a particular hydrogen bond (encircled) which is sampled for >10% of the total simulation time. All conformations having a given hydrogen bond are clustered together and in each case the central conformation of the cluster is shown. The percentage occurrence of these persistent hydrogen bonds in **DA** and $\mathbf{G_2}$ peptides is indicated below each of the conformations.

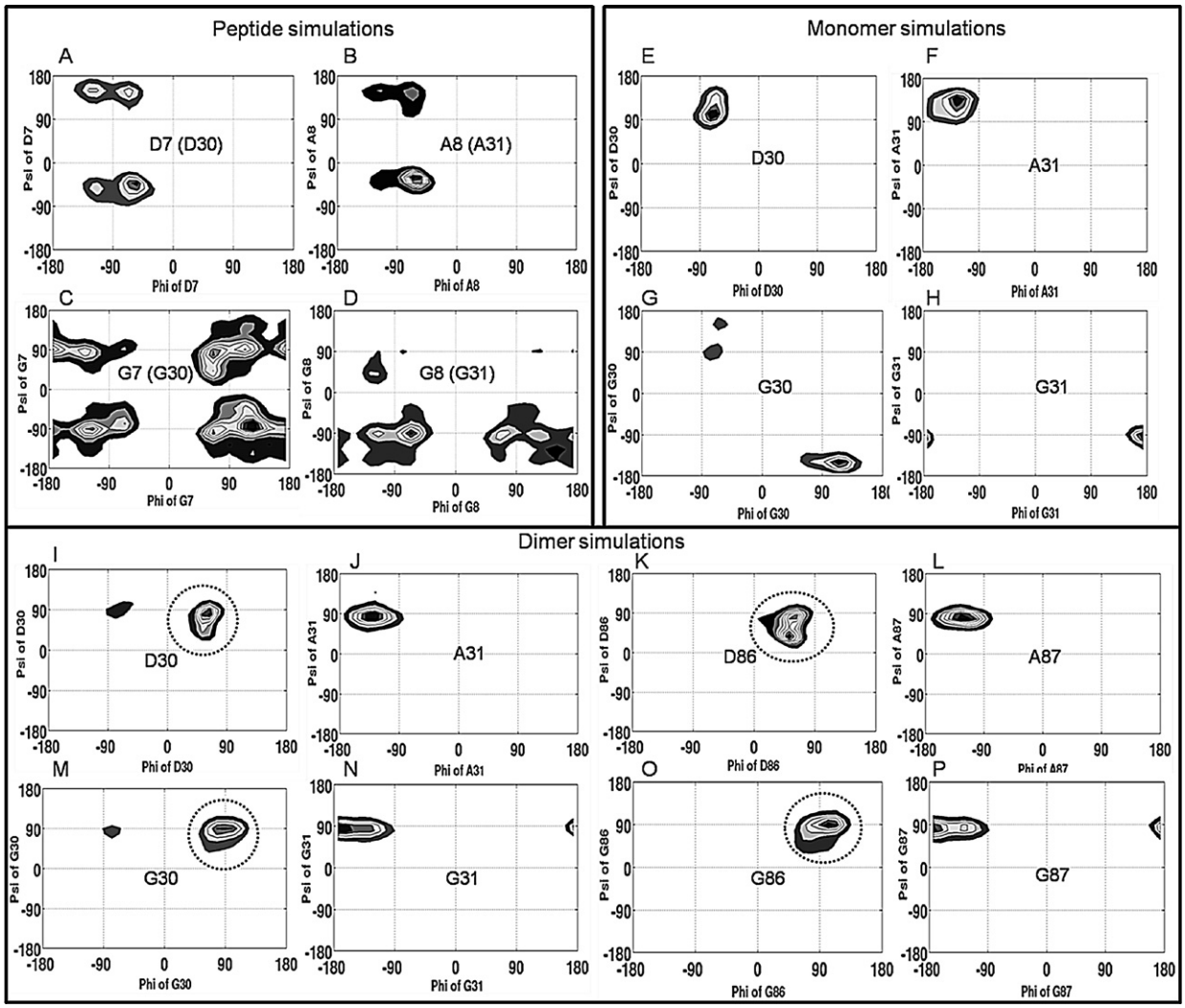


Fig. 5. Backbone dihedral angles (φ, ψ) sampled by the residues D7 and A8 in **DA** peptide simulation are shown in (A) and (B), by the residues G7 and G8 residues in $\mathbf{G_2}$ peptide are shown in (C) and (D). The backbone dihedral angles sampled by D30 and A31 residues in **Wt-Rop** monomer simulation are shown in (E) and (F) and by G30 and G31 residues in $\mathbf{G_2}$**-Rop** monomer are shown in (G) and (H). The backbone dihedral angles sampled by D30/D86 and A31/A87 residues in the two monomers of **Wt-Rop** dimer simulation are shown in (I)–(L) and by G30/G86 and G31/G87 residues in the two monomers of $\mathbf{G_2}$**-Rop** dimer simulation are shown in (M)–(P). Note that D30 and D86 residues in **Wt-Rop** dimer and G30 and G86 residues in $\mathbf{G_2}$**-Rop** dimer sample α_L conformation (encircled).

peptides indicate a number of persistent hydrogen bonds (existing for more than 10% of total simulation time) present in all the four peptides. A comprehensive list of the sampling of various hydrogen bonds in **DA**, $\mathbf{G_2}$, $\mathbf{G_3}$ and $\mathbf{G_{3'}}$ peptide simulations is presented in Tables S1, S2, S3 and S4 of Supporting information. We extracted the conformations sampled in **DA** and $\mathbf{G_2}$ peptide simulations with persistent hydrogen bonds and they are shown in Fig. 4. In the case of **DA** peptide presence of two of the persistent hydrogen bonds (L3:NH-E10:Oε1/2 and N4:NH-E10:Oε1/2) lead to sampling of conformations which could cause a loop closure (Fig. 4A). In contrast to this, in $\mathbf{G_2}$ peptide three of the persistent hydrogen bonds (L3:NH-Q11:CO, Q11:NH-L3:CO and E5:NH-D9:CO) could give rise to conformations with loop closure (Fig. 4B). The percentage occurrence of the persistent hydrogen bonds which are associated with loop closure, is also found to be higher in $\mathbf{G_2}$ peptide (19–20%) in comparison to that in **DA** peptide (11–12%) (Fig. 4A and B). Further we note that the three persistent main chain hydrogen bonds in $\mathbf{G_2}$ peptide, which could lead to loop closure, are also sampled simultaneously for 10–11% of the simulation time (Table S6 of Supporting information). Hence, these hydrogen bonds strengthen each other. $\mathbf{G_3}$ and $\mathbf{G_{3'}}$ peptides also sample certain persistent main-chain hydrogen bonds and these hydrogen bonds are also present in $\mathbf{G_2}$ peptide simulation (see Table 1).

3.2. Phi-psi sampling in peptide, protein monomer and dimer simulations suggests an important role of 'glycine' at position 30 in optimizing the dimeric structure

Fig. 5 shows the phi-psi sampling for D30 and A31 residues as part of **DA** peptide and as part of **Wt-Rop** monomer and dimer simulations. It also shows the phi-psi sampling for G30 and

Table 1
Comparison of percentage occurrence of persistent hydrogen bonds in $\mathbf{G_2}$, $\mathbf{G_3}$ and $\mathbf{G_{3'}}$ mutant peptides is shown.

$\mathbf{G_2}$ peptide (^{2}*KLNEL***GG***DEQ*^{11})	L3NH-Q11CO 19%	Q11NH-L3CO 19%	E5NH-D9CO 20%
$\mathbf{G_3}$ peptide (^{2}*KLNEL***GGG***DEQ*^{12})	L3NH-Q12CO 9%	Q12NH-L3CO 15%	E5NH-D10CO 15%
$\mathbf{G_{3'}}$ peptide (^{2}*KLNEL***GGG***EQ*^{11})	L3NH-Q11CO 10%	Q11NH-L3CO 10%	E5NH-G9CO 7%

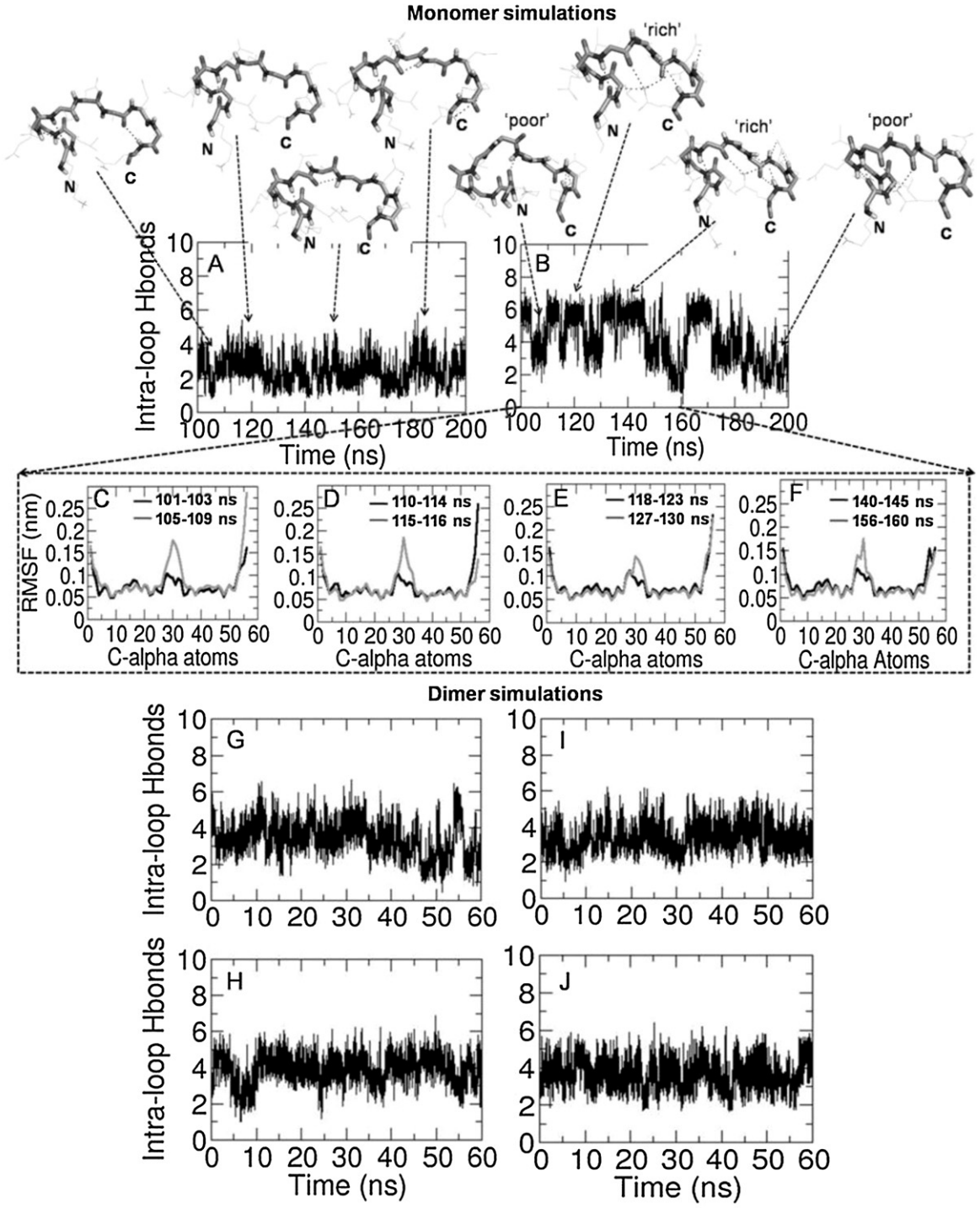


Fig. 6. Variation of number of intra-loop hydrogen bonds with simulation time during the equilibrium period (100–200 ns) is shown in (A) for **Wt-Rop** monomer and in (B) for **G_2-Rop** monomer simulation. Some of the loop conformations sampled in the two monomer simulations are also shown. The variation of C-α atom RMSF for **G_2-Rop** monomer when 'rich' hydrogen bonded loop is sampled (black) versus the case when 'poor' hydrogen bonded loop is sampled (gray) is shown in (C) for 101–109 ns; (D) for 110–116 ns; (E) for 118–130 ns and (F) for 140–160 ns. Note that poor hydrogen bonded loop has higher RMSF compared to rich hydrogen bonded loop. The variation of number of intra-loop hydrogen bonds for monomer1 and 2 is shown for the dimer simulation of **Wt-Rop** in (G) and (H) and for monomer 1 and 2 in **G_2-Rop** dimer simulation is shown in (I) and (J). It is interesting to note that a switching behavior between 'rich' and 'poor' hydrogen bonded regions is obserevd in case of **G_2-Rop** monomer but not in case of **G_2-Rop** dimer.

G31 residues in **G_2** peptide simulation and corresponding **G_2-Rop** monomer and dimer simulations. G30 as part of **G_2** peptide has more conformational freedom compared to D30 in **DA** peptide (Fig. 5A and C). In the monomer, D30 residue samples extended conformations (Fig. 5E); while in the dimer it samples mainly α_L conformation (Fig. 5I and K). This suggests that in the dimer the D30 residue is constrained to adopt an α_L conformation, which must be essential for proper re-organization/packing of helices. Further, it can be seen that G30/G86 residues in dimeric **G_2-Rop** also sample α_L conformation (Fig. 5M and O). Thus, a more flexible residue, like G30 in place of D30 as part of the dimer could attain α_L conformation with ease and hence facilitate dimerization, since

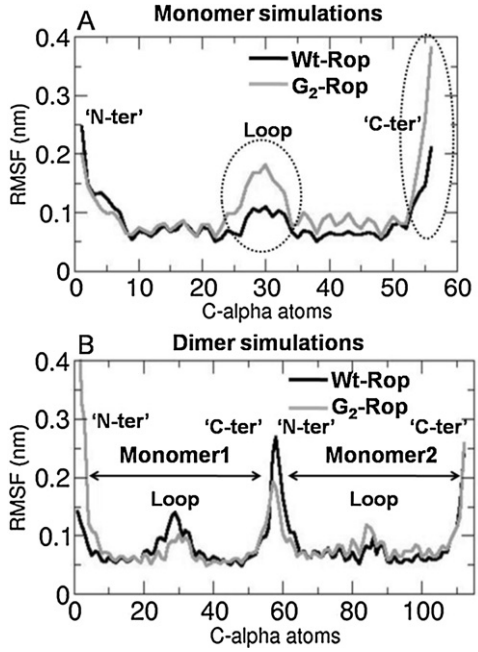


Fig. 7. Variation of C-α atom RMSF for all the residues in **Wt-Rop** monomer (black) and **G_2-Rop** monomer (gray) is shown in (A) and for **Wt-Rop** dimer (black) and **G_2-Rop** dimer (gray) simulations is shown in (B).

dimeric structure of Rop requires α_L conformation at position 30 in the turn [5,6].

3.3. A switching between 'poor' and 'rich' hydrogen bonded loop leads to a more dynamic loop region (25–34) in *G_2-Rop* monomer compared to *Wt-Rop* monomer

We calculated the number of intra-protein and protein-solvent hydrogen bonds in **Wt-Rop** and **G_2-Rop** monomer and dimer systems as described in Section 2.6. Table 2 shows the intra-protein and protein-solvent hydrogen bonds in protein monomer and dimer systems. The number of hydrogen bonds (both intra-protein and protein-solvent) is almost same in case of **Wt-Rop** and **G_2-Rop** *dimers*. On the other hand, **Wt-Rop** monomer has greater number of protein-solvent hydrogen bonds than **G_2-Rop** monomer. This difference is mainly accounted for, by the loop region (K25-Q34) of monomer where there is a difference of 9 protein-solvent hydrogen bonds (Table 2). Thus, the loop region of **G_2-Rop** monomer seems to be having greater number of intra-protein hydrogen bonds compared to the loop region of **Wt-Rop** monomer. The variation of intra-loop hydrogen bonds with time in both monomer and dimer simulations is shown in Fig. 6. It is interesting to note that, loop region in **G_2-Rop** monomer shows a switching between 'rich' hydrogen bonded loop (up to 6 or 7 hydrogen bonds) and 'poor' hydrogen bonded loop (with 2 or 3 hydrogen bonds) which is not observed in the case of **Wt-Rop** monomer (Fig. 6A and B). This switching behavior is also absent in case of **Wt-Rop** and **G_2-Rop** dimers (Fig. 6G–J), suggesting the relatively rigid nature of the loop region in the native dimeric state. Fig. 6C–F shows the C-α atom root mean square fluctuations (RMSF) for the 'rich' and 'poor' hydrogen bonded loop in **G_2-Rop** monomer trajectory and it is found that rich hydrogen bonded loop has a lower RMSF compared to the poor hydrogen bonded loop. Further the variation of overall C-α atom RMSF for the monomer and dimer simulations is shown in Fig. 7. The loop region in **G_2-Rop** monomer is more dynamic compared to the loop region of **Wt-Rop** monomer (Fig. 7A). In case of **G_2-Rop** monomer the C-terminus is also found to be more dynamic compared to C-terminus of **Wt-Rop** monomer (Fig. 7A). On the other hand, the loop region in the dimer simulations is flexible neither in **Wt-Rop** nor in the **G_2-Rop** dimer simulations (Fig. 7B). Thus, introduction of glycines in the turn region (as in the case of **G_2-Rop** monomer) has imparted a more dynamic nature to the loop region of the monomer. But once the monomer is packed in the dimeric state the loop becomes more rigid. These RMSF results suggest that presence of a relatively more dynamic monomer, as in **G_2-Rop** monomer simulations, could facilitate reorganization/packing within the dimeric-intermediate during native dimer formation.

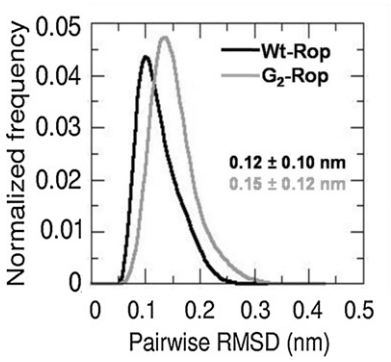


Fig. 8. Pair-wise RMSD distribution in case of monomer systems is shown for **Wt-Rop** monomer in black and **G_2-Rop** monomer in gray. The mean values and standard deviations for the pair-wise RMSD are indicated in each case.

3.4. Pair-wise RMSD distribution in protein monomer simulations suggest relatively greater conformational variability in *G_2-Rop* monomer compared to *Wt-Rop* monomer

Fig. 8 shows the pair-wise RMSD distribution for monomer systems. **G_2-Rop** monomer shows a relatively broader pair-wise RMSD distribution around a slightly higher value of RMSD suggesting greater conformational variability in **G_2-Rop** monomer compared to **Wt-Rop** monomer. This is in agreement with earlier results where a more dynamic loop and C-terminal region were observed for **G_2-Rop** monomer in comparison to a more rigid **Wt-Rop** monomer (Fig. 7A). Further, in the case of dimer systems, **G_2-Rop** monomer 1 shows a flexible N-terminus as evident from Fig. 7B; while monomer 2 of both the dimeric proteins has almost similar RMSF. In monomer 1 of **G_2-Rop** dimer this dynamic nature may be coming from sampling of higher RMSD value after 40 ns as shown in Figure S3B of Supporting information. These results suggest that during the rate limiting step of folding, where dimeric-intermediate forms and reorganization/packing of helices takes place, **G_2-Rop** monomer could enhance folding due to more dynamic nature of the individual monomers involved.

4. Discussion

4.1. The possible role of loop closure in the folding of dimeric helix-turn-helix Rop

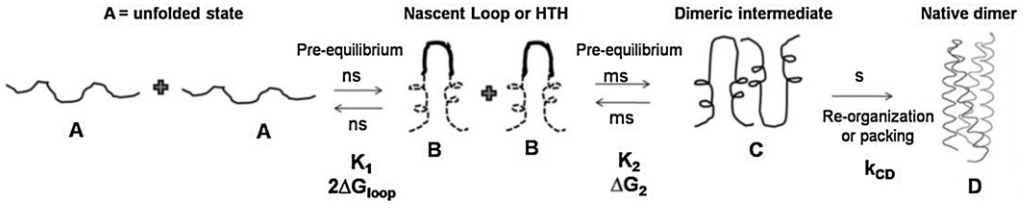
Several experimental studies are available that deal with the folding of helix-turn-helix or four-helix-bundle proteins. A spectroscopic study on a monomeric helical hairpin porcine PYY shows that a mutation in turn region particularly affects the folding transition state. It is thought that the hydrophobic core formation must be taking place after the transition state is formed [39]. Du and Gai [40] studied the folding kinetics of α-helical hairpin Z34C and its turn and hydrophobic core mutants using spectroscopic techniques. They find that the rate limiting step is governed by the formation of a reverse turn, while unfolding rate depends on strength of hydrophobic associations between helices. Phi value analysis on

Table 2
Lists the number of total, loop and helix hydrogen bonds for **Wt-Rop** monomer and dimer simulations and for **G_2-Rop** monomer and dimer simulations at 300 K. Intra-protein and protein-solvent hydrogen bonds are shown separately. For monomer simulations, these numbers are calculated over the equilibrium part of the trajectory as described in Section 2.

	Wt-Rop monomer	**Wt-Rop** dimer	**G_2-Rop** monomer	**G_2-Rop** dimer
Hydrogen bonds in helix-turn-helix				
Intra-protein	47 ± 2	51 ± 3 M1[a]	49 ± 2	50 ± 3 M1[a]
		51 ± 3 M2[a]		51 ± 3 M2[a]
		107 ± 4[b]		107 ± 4[b]
Protein-solvent	136 ± 6	51 ± 5 M1[a]	129 ± 7	50 ± 5 M1[a]
		50 ± 5 M2[a]		47 ± 5 M2[a]
		129 ± 40[b]		125 ± 38[b]
Hydrogen bonds in loop region (K25-Q34) of helix-turn-helix				
Intra-protein	2 ± 1	3 ± 1 M1[a]	4 ± 1	3 ± 1 M1[a]
		3 ± 1 M2[a]		3 ± 1 M2[a]
Protein-solvent	39 ± 3	11 ± 2 M1[a]	30 ± 3	9 ± 2 M1[a]
		10 ± 2 M2[a]		8 ± 2 M2[a]
Hydrogen bonds in helix1 and helix2 of helix-turn-helix				
Intra-protein	38 ± 2	40 ± 2 M1[a]	37 ± 2	39 ± 2 M1[a]
		40 ± 2 M2[a]		40 ± 2 M2[a]
Protein-solvent	96 ± 5	40 ± 4 M1[a]	98 ± 5	41 ± 5 M1[a]
		39 ± 4 M2[a]		39 ± 4 M2[a]

[a] M1 and M2 refers to the average and standard deviation values for the number of hydrogen bonds sampled by monomer1 and monomer2 separately in dimer simulations.
[b] These numbers refer to the dimer as a whole.

a three-helix-bundle villin headpiece HP-35 also emphasizes the role of turn hydrogen bonds. Elimination of turn hydrogen bonds in this three-helix-bundle protein by amide to ester mutations leads to slower folding rates suggesting that turn is involved in the formation of folding nucleus [41]. Turn or loop formation is also a crucial event during folding of β-hairpins. Studies on β-hairpin forming peptides suggest that turn formation is a rate-limiting step in the folding of a β-hairpin while hydrophobic interactions and hydrogen bond formation contribute toward stability of the hairpin and also affects its unfolding kinetics [42–44]. The double-funneled energy landscape theory for Rop folding also suggests that loop region of Rop is ordered in the folding transition state ensemble [45]. It is known from experimental fluorescence studies that the fast phase of Rop folding involves formation of a dimeric-intermediate. It is thought to be composed of an early kinetic intermediate which has both native and non-native characteristics which could be involving partially formed helical hairpins [7,15]. These observations correlate well with each other and are suggestive of the role played by loop closure in the folding of Rop. Thus it seems plausible to assume that loop closure is required for helix-turn-helix associ-



$$k_{CD} = \text{rate constant for the step } C \longrightarrow D$$

$$K_1 = \frac{[B]^2}{[A]^2} = \left(\frac{[B]}{[A]}\right)^2 = K_{loop}^2 = e^{\frac{-2\Delta G_{loop}}{RT}}$$

$$K_2 = \frac{[C]}{[B]^2}$$

	Experimental data [8]				MD simulation data	
System	Fast phase (s^{-1})	Fold increase	Slow phase (s^{-1})	Fold increase	K_1 obtained from kinetic model	Fold increase
Wt-Rop	29	1	0.5	1	0.32	1
G_2-Rop	580	20	31	62	2.43	7.6
G_3-Rop	190	6.5	14	28	1.65	5.2
G_3'-Rop	------	-----	-----	-----	2.0	6.2

$$V_{folding} = \frac{d}{dt}[D] = k_{CD}[C] = k_{CD} \times K_2[B]^2$$

$$V_{folding} = k_{CD} \times K_2 \times K_1[A]^2$$

$$V_{folding} = k_{CD} \times K_2 \times e^{\frac{-2\Delta G_{loop}}{RT}} \times [A]^2$$

$$V_{folding} = k_{CD} \times e^{\frac{-\Delta G_2}{RT}} \times e^{\frac{-2\Delta G_{loop}}{RT}} \times [A]^2$$

$$k_{folding} \propto e^{\frac{-2\Delta G_{loop}}{RT}} \text{ or } K_1$$

Fig. 9. Schematic diagram of a simple mechanism proposed for Rop folding on the basis of available information from literature and our simulation study to highlight the contribution of loop propensity to the overall folding rates of Rop. Dotted lines in species 'B' indicate nascent/unstable helical turns that are brought together as a consequence of loop formation. The interaction between the nascent/unstable helical units is expected to be weak. The inset table compares experimental 'fold increase' with computed 'fold increase' from current simulations. Note that computed K_1 is proportional to rate of folding from experiment.

ation. Based on such an assumption, a tentative kinetic scheme can be suggested that can correlate folding rates to loop propensity as shown in Fig. 9. From our peptide simulations it is clear that folding and unfolding of loop structures takes place over a timescale of nanoseconds. As per Nagi et al.'s experimental study using fluorescence [8], the formation of dimeric-intermediate occurs as a fast phase over a timescale of milliseconds. Further a slow phase involving reorganization/packing leading to native dimer occurs over a timescale of seconds [7–9]. As suggested, a possible folding initiation event in a helix-turn-helix folding could be formation of a bend or turn in the polypeptide chain followed by diffusion of nascent helices toward each other [46] leading to formation of a helical hairpin which is further stabilized by hydrophobic packing of side chains [47]. In view of the time scales indicated, pre-equilibrium for steps 1 and 2 of the model would prevail and the overall folding rate for the Rop dimer can be calculated as shown in Fig. 9. In this model, the loop propensity is modeled by the factor $e^{-2\Delta \mathrm{Gloop}/RT}$ in the folding rate of Rop. The purpose of this model is mainly to estimate the contribution of loop propensity to the overall folding rate. This contribution can be expressed quantitatively in terms of the equilibrium constant K_1 in step 1 which involves loop closure. The values for K_1 for different peptides **DA**, $\mathbf{G_2}$, $\mathbf{G_3}$ and $\mathbf{G_{3'}}$ in the kinetic model are obtained as a ratio of fraction folded (end-to-end distance less than or equal to 0.7 nm) to fraction unfolded (end-to-end distance greater than or equal to 0.7 nm) and these are shown in the inset table in Fig. 9 along with the 'fold increase'. The table in Fig. 9 also gives the folding rates for **Wt-Rop**, $\mathbf{G_2}$**-Rop** and $\mathbf{G_3}$**-Rop** from the stopped-flow fluorescence experiment [8] and the relative experimental fold increase. It is interesting to note that the experimental and simulation 'fold increase' values show an order of magnitude agreement. The contribution of loop propensity in re-organization/packing of helices is not captured by this model. It is quite plausible that loop propensity/flexibility will have a role in the reorganization/packing of the monomers. Inferences about the role of loop propensity in this final step, where reorganization/packing occurs, could be drawn from the simulations of monomers and dimers of **Wt-Rop** and $\mathbf{G_2}$**-Rop** as discussed in Sections 3.2, 3.3 and 3.4. During the rate limiting step of folding where reorganization/packing occurs, presence of more dynamic monomers, as observed in the case of $\mathbf{G_2}$**-Rop** monomer simulation, could enhance folding rates by facilitating association thus, more dynamic monomers in case of $\mathbf{G_2}$ mutant in comparison to wt protein could facilitate packing in the dimeric-intermediate state (species C in Fig. 9). Another factor which will influence proper packing in the dimer is the sampling of α_L conformation by D30 residue. G30/G86 residue in $\mathbf{G_2}$**-Rop** dimer could sample α_L conformation with much more ease compared to D30/D86 residue in **Wt-Rop** dimer. We note that certain turn mutants of Rop prevent dimerization of the helix-turn-helix units [48]. In this study Rop sequence was substituted for the dimerization domain of λ repressor and it was found that turn mutants 'VED', 'VPD' and 'YPD' at positions 30, 31 and 32 in Rop turn prevented dimerization. This observation suggests that presence of bulkier residues like Val and Tyr at position 30 in the Rop turn delay/prevent dimerization and hence delay folding due to the requirement for an α_L conformation.

4.2. *Early loop formation events in **Wt-Rop** and its glycine loop mutants may involve non-native hydrogen bonding*

Fluorescence resonance energy transfer and simulation studies on poly (Gly-Ser) peptides have shown that initial collapse in these polypeptides is driven by formation of intra-peptide backbone hydrogen bonds [49,50]. NMR relaxation dispersion study on a four-helix-bundle FF domain from HYPA/FBP11 and its mutant L24A also indicates presence of non-native interactions in the folding intermediate which are lost in the native state [51]. Our simulation study on **DA**, $\mathbf{G_2}$, $\mathbf{G_3}$ and $\mathbf{G_{3'}}$ peptides suggest sampling of conformations with loop closure, involving formation of non-native persistent hydrogen bonds. The persistent hydrogen bonds in case of the G_n peptides are mainly backbone hydrogen bonds and are sampled for greater percentage (10–20%) of the total simulation time compared to those in **DA** peptide (11–12%). The persistent backbone hydrogen bonds observed in case of G_n peptide simulations could be possible representation of early loop closure events during the folding of glycine loop mutants of Rop. Such hydrogen bond loop forming conformers are relatively less in **DA** peptide simulation suggesting a rationale for the observed slow folding rates in wt-Rop [8].

4.3. *A more dynamic or conformationally variable monomer in $\mathbf{G_2}$**-Rop** could facilitate reorganization/packing in dimeric-intermediate during folding*

The variation of RMSF for C-α atoms in the monomer simulations indicated a more flexible loop region and C-terminal region in $\mathbf{G_2}$**-Rop** monomer in comparison to **Wt-Rop** monomer. During the folding when monomer–monomer interactions are weak (as in the dimeric-intermediate state), a G30-G31 mutation in loop could lend more conformational flexibility for reorganization/packing in the rate limiting step and thus, could be responsible for enhanced folding rates in G_n mutants in comparison to **Wt-Rop**. We note that in our study the G30-G31 mutation in Rop monomer has both local and distant effects. The mutation of D30-A31 to G30-G31, as in the $\mathbf{G_2}$**-Rop** monomer, leads to a dynamic loop which also makes the C-terminal of the monomer more dynamic. The observation that mutation at one region affects a distant part in a protein is reminiscent of allosteric effects. Akin to this observation, in an NMR spectroscopic study on a four-helix-bundle acyl coenzyme A binding protein (ACBP) shows that a mutation in a loop affects association of far away helices suggesting cooperativity during folding [52]. Another important implication for the folding of dimeric helix-turn-helix Rop protein comes from the dynamic nature of the N-terminus observed in one of the monomers of $\mathbf{G_2}$**-Rop** dimer simulation. In comparison to this a relatively rigid conformation is observed for dimeric **Wt-Rop** (Fig. 7B). Such dynamic N-terminus in $\mathbf{G_2}$**-Rop** dimer could have implications for faster unfolding rates observed for Gly_2 mutant of Rop compared to wt-protein in the fluorescence study [8].

4.4. *Possible force field bias effects*

Replica exchange molecular dynamics (REMD) studies on amyloid forming peptides $A\beta_{16-22}$ [53] and H1 peptide from Syrian hamster prion protein [54,55] show that GROMOS 43a1 force field along with SPC water model stabilizes the formation of β-structure compared to other force fields. Cao and Wang [54] also find that GROMOS 43a1 force field gives a free energy change from unfolded state to β-hairpin, which is in good agreement with experiments on β peptides. For H1 peptide, GROMOS 43a1 force field is found to give better results in comparison to OPLS-AA force field [55]. Further, MD simulation study on a 16-mer β-hairpin forming peptide from Nrf2 protein yields a native-like β-hairpin starting from an extended conformation, only in case of GROMOS 43a1 force field in comparison to other nine force fields, suggesting no bias in this case [56].

These observations seem to suggest that success of a given force field in determining reliable conformational properties could also be sequence or system dependent. In the current simulation study the secondary structure plots in Fig. 3 show a dominance of bend conformations for *all* the four peptides (**DA**, $\mathbf{G_2}$, $\mathbf{G_3}$ and $\mathbf{G_{3'}}$). There is no α-helical or β-sheet conformational sampling in **DA** peptide simulation while, a small percentage of β-bridge and sometimes

β-hairpin sampling is observed in case of G_n (**G_2**, **G_3** and **$G_{3'}$**) peptides. Thus, while we cannot rule out force field bias effects, we cannot confirm the presence of bias also either toward α-helical or β-sheet conformations in the current peptide simulations.

5. Conclusions

We have considered peptide models of loop region of Rop protein to assess the role played by the loop region in the folding of Rop protein. The glycine mutant loop peptides (G_n) have higher loop closure propensities compared to wt-peptide and therefore could facilitate the early steps in the folding of G_n Rop mutants. Based on our simulation results of loop peptides (**DA** and G_n peptides), and assuming that loop closure is a prerequisite for folding, we have shown that loop closure propensity could contribute to overall folding rates of Rop by a factor $e^{-2\Delta G\text{loop}/RT}$. Comparing monomeric and dimeric **Wt-Rop** and **G_2-Rop** simulations, a more dynamic monomer is observed in case of **G_2-Rop**, which could facilitate the formation of dimeric-intermediate. We also find that dimeric form requires D30/G30 or D86/G86 residues to sample α_L conformation. This conformation is more readily attained by G_n mutant Rop protein, where glycine based loop could facilitate reorganization/packing of dimeric state better than wt-protein, thus contributing to faster folding of G_n Rop mutants. Thus loop region potentially plays a dual role in the folding of Rop, both as an initiator of folding and optimizer of monomer–monomer interactions in the dimeric state.

Acknowledgements

RTS thanks Indian Institute of Technology Bombay, Mumbai for teaching assistance. This work was supported by the research grant from the Board of Research in Nuclear Sciences (BRNS), Department of Atomic Energy, Mumbai [sanction number 2007/37/32/BRNS/1909] sanctioned to YUS.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2012.12.007.

References

[1] A. Rosengarth, J. Rösgen, H.J. Hinz, Slow unfolding and refolding kinetics of the mesophilic Rop wild-type protein in the transition range, European Journal of Biochemistry 264 (1999) 989–995.
[2] M.S. Gittelman, C.R. Matthews, Folding and stability of trp aporepressor from *Escherichia coli*, Biochemistry 29 (1990) 7011–7020.
[3] M.E. Milla, R.T. Sauer, P22 Arc repressor: folding kinetics of a single-domain, dimeric protein, Biochemistry 33 (1994) 1125–1133.
[4] H. Wendt, C. Berger, A. Baici, R.M. Thomas, H.R. Bosshard, Kinetics of folding of leucine zipper domains, Biochemistry 34 (1995) 4097–4107.
[5] D.W. Banner, M. Kokkinidis, D. Tsernoglou, Structure of the ColE1 rop protein at 1.7 Å resolution, Journal of Molecular Biology 196 (1987) 657–675.
[6] W. Eberle, A. Pastore, C. Sander, P. Rösch, The structure of ColE1 rop in solution, Journal of Biomolecular NMR 1 (1991) 71–82.
[7] M. Munson, K.S. Anderson, L. Regan, Speeding up protein folding: mutations that increase the rate at which Rop folds and unfolds by over four orders of magnitude, Folding and Design 2 (1997) 77–87.
[8] A.D. Nagi, K.S. Anderson, L. Regan, Using loop length variants to dissect the folding pathway of a four-helix-bundle protein, Journal of Molecular Biology 286 (1999) 257–265.
[9] S. Dalal, D. Canet, S.E. Kaiser, C.M. Dobson, L. Regan, Conservation of mechanism, variation of rate: folding kinetics of three homologous four-helix bundle proteins, Protein Engineering Design and Selection 21 (2008) 197–206.
[10] A.D. Nagi, L. Regan, An inverse correlation between loop length and stability in a four-helix-bundle protein, Folding and Design 2 (1997) 67–75.
[11] P.F. Predki, V. Agrawal, A.T. Brünger, L. Regan, Amino-acid substitutions in a surface turn modulate protein stability, Nature Structural Biology 3 (1996) 1055.
[12] M. Munson, R. O'Brien, J.M. Sturtevant, L. Regan, Redesigning the hydrophobic core of a four-helix-bundle protein, Protein Science 3 (1994) 2015–2022.
[13] M.S. Munson, S. Balasubramanian, K.G. Fleming, A.D. Nagi, R. O'Brien, J.M. Sturtevant, L. Regan, What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties, Protein Science 5 (1996) 1584–1593.
[14] S.B. Hari, C. Byeon, J.J. Lavinder, T.J. Magliery, Cysteine-free rop: a four-helix bundle core mutant has wild-type stability and structure but dramatically different unfolding kinetics, Protein Science 19 (2010) 670–679.
[15] N.A.J. van Nuland, C.M. Dobson, L. Regan, Characterization of folding the four-helix bundle protein Rop by real-time NMR, Protein Engineering Design and Selection 21 (2008) 165–170.
[16] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723.
[17] A.L. Rucker, T.P. Creamer, Polyproline II helical structure in protein unfolded states: lysine peptides revisited, Protein Science 11 (2002) 980–985.
[18] Z.S. Shi, C.A. Olson, G.D. Rose, R.L. Baldwin, N.R. Kallenbach, Polyproline II structure in a sequence of seven alanine residues, Proceedings of the National Academy of Sciences of the United States of America 99 (2002) 9190–9195.
[19] A. Rath, A.R. Davidson, C.M. Deber, The structure of unstructured regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition, Biopolymers 80 (2005) 179–185.
[20] J. Makowska, S. Rodziewicz-Motowidlo, K. Bagińska, J.A. Vila, A. Liwo, L. Chmurzyński, H.A. Scheraga, Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins, Proceedings of the National Academy of Sciences of the United States of America 103 (2006) 1744–1749.
[21] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen, H.J.C. Berendsen, Gromacs User Manual Version 4.0, AG Groningen, The Netherlands, 2005.
[22] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The missing term in effective pair potentials, Journal of Physical Chemistry 91 (1987) 6269–6271.
[23] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, Journal of Chemical Theory and Computation 4 (2008) 435–447.
[24] W.R.P. Scott, P.H. Hünenberger, I.G. Tironi, A.E. Mark, S.R. Billeter, J. Fennen, A.E. Torda, T. Huber, P. Krüger, W.F. van Gunsteren, The GROMOS biomolecular simulation program package, Journal of Physical Chemistry A 103 (1999) 3596–3607.
[25] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald: an $N\log(N)$ method for Ewald sums in large systems, Journal of Chemical Physics 98 (1993) 10089–10092.
[26] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth Particle Mesh Ewald method, Journal of Chemical Physics 103 (1995) 8577–8593.
[27] P.F.J. Fuchs, A.M.J.J. Bonvin, B. Bochicchio, A. Pepe, A.J.P. Alix, A.M. Tamburro, Kinetics and thermodynamics of type VIII β-turn formation: a CD, NMR, and microsecond explicit molecular dynamics study of the GDNP tetrapeptide, Biophysical Journal 90 (2006) 2745–2759.
[28] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, Journal of Computational Chemistry 18 (1997) 1463–1472.
[29] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, Journal of Chemical Physics 81 (1984) 3684–3690.
[30] P.N. Lewis, F.A. Momany, H.A. Scheraga, Chain reversals in proteins, Biochimica et Biophysica Acta 303 (1973) 211–229.
[31] P.Y. Chou, G.D. Fasman, β-Turns in proteins, Journal of Molecular Biology 115 (1977) 135–175.
[32] C.M. Wilmot, J.M. Thornton, Analysis and prediction of the different types of β-turn in proteins, Journal of Molecular Biology 203 (1988) 221–232.
[33] J.F. Leszczynski, G.D. Rose, Loops in globular proteins: a novel category of secondary structure, Science 234 (1986) 849–855.
[34] H.X. Zhou, Loops in proteins can be modeled as worm-like chains, Journal of Physical Chemistry B 105 (2001) 6763–6766.
[35] I.C. Yeh, G. Hummer, Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent, Journal of the American Chemical Society 124 (2002) 6563–6568.
[36] M.J. Feige, E. Paci, Rate of loop formation in peptides: a simulation study, Journal of Molecular Biology 382 (2008) 556–565.
[37] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[38] X. Daura, W.F. van Gunsteren, A.E. Mark, Folding–unfolding thermodynamics of a β-heptapeptide from equilibrium simulations, Proteins 34 (1999) 269–280.
[39] M.M. Waegele, F. Gai, Infrared study of the folding mechanism of a helical hairpin: porcine PYY, Biochemistry 49 (2010) 7659–7664.
[40] D. Du, F. Gai, Understanding the folding mechanism of an α-helical hairpin, Biochemistry 45 (2006) 13131–13139.
[41] M.R. Bunagan, J. Gao, J.W. Kelly, F. Gai, Probing the folding transition state structure of the villin headpiece subdomain via side chain and backbone mutagenesis, Journal of the American Chemical Society 131 (2009) 7470–7476.
[42] V. Muñoz, P.A. Thompson, J. Hofrichter, W.A. Eaton, Folding dynamics and mechanism of β-hairpin formation, Nature 390 (1997) 196–199.

[43] D. Du, Y. Zhu, C.Y. Huang, F. Gai, Understanding the key factors that control the rate of β-hairpin folding, Proceedings of the National Academy of Sciences of the United States of America 101 (2004) 15915–15920.
[44] S. Patel, Y.U. Sasidhar, A shorter peptide model from staphylococcal nuclease for the folding–unfolding equilibrium of a β-hairpin shows that unfolded state has significant contribution from compact conformational states, Journal of Structural Biology 164 (2008) 60–74.
[45] Y. Levy, S.S. Cho, T. Shen, J.N. Onuchic, P.G. Wolynes, Symmetry and frustration in protein energy landscape: a near degeneracy resolves the Rop dimer-folding mystery, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 2373–2378.
[46] M. Karplus, D.L. Weaver, Protein folding dynamics: the diffusion–collision model and experimental data, Protein Science 3 (1994) 650–668.
[47] D. Hoffmann, E.W. Knapp, Folding pathways of a helix-turn-helix model protein, Journal of Physical Chemistry B 101 (1997) 6734–6740.
[48] L. Castagnoli, C. Vetriani, G. Cesareni, Linking an easily detectable phenotype to the folding of a common structural motif. Selection of rare turn mutations that prevent the folding of Rop, Journal of Molecular Biology 237 (1994) 378–387.
[49] A. Möglich, K. Joder, T. Kiefhaber, End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation, Proceedings of the National Academy of Sciences of the United States of America 103 (2006) 12394–12399.
[50] I. Daidone, H. Neuweiler, S. Doose, M. Sauer, J.C. Smith, Hydrogen-bond driven loop-closure kinetics in unfolded polypeptide chains, PLOS Computational Biology 6 (2010) e1000645.
[51] D.M. Korzhnev, R.M. Vernon, T.L. Religa, A.L. Hansen, D. Baker, A.R. Fersht, L.E. Kay, Nonnative interactions in the FF domain folding pathway from an atomic resolution structure of a sparsely populated intermediate: an NMR relaxation dispersion study, Journal of the American Chemical Society 133 (2011) 10974–10982.
[52] S.W. Bruun, V. Iešmantavičius, J. Danielsson, F.M. Poulsen, Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein, Proceedings of the National Academy of Sciences of the United States of America 107 (2010) 13306–13311.
[53] P.H. Nguyen, M.S. Li, P. Derreumaux, Effects of all-atom force fields on amyloid oligomerization: replica exchange molecular dynamics simulations of the $A\beta_{16-22}$ dimer and trimer, Physical Chemistry Chemical Physics 13 (2011) 9778–9788.
[54] Z. Cao, J. Wang, A comparative study of two different force fields on structural and thermodynamics character of H1 peptide via molecular dynamics simulation, Journal of Biomolecular Structure and Dynamics 27 (2010) 651–661.
[55] Z. Cao, L. Liu, J. Wang, Why the OPLS-AA force field cannot produce the β-hairpin structure of H1 peptide in solution when comparing with the GROMOS 43A1 force field? Journal of Biomolecular Structure and Dynamics 29 (2011) 527–539.
[56] E.A. Cino, W.-Y. Choy, M. Karttunen, Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations, Journal of Chemical Theory and Computation 8 (2012) 2725–2740.