β-hairpin forms by rolling up from C-terminal: Topological guidance of early folding dynamics

Søren Enemark*, Nicholas A. Kurniawan*† & Raj Rajagopalan‡

Singapore-MIT Alliance, National University of Singapore, 4 Engineering Drive 3, E4-14-10, Singapore 117576.

Correspondence and requests for materials should be addressed to R.R. (raj.private.email@gmail.com; raj@skolkovotech.ru)

* These authors contributed equally to this work.

† Current address: FOM Institute AMOLF, Science Park 102, 1098XG Amsterdam, The Netherlands.

‡ Current address: Skolkovo Institute of Science and Technology, Moscow, The Russian Federation.

That protein folding is a non-random, guided process has been known even prior to Levinthal's paradox; yet, guided searches, attendant mechanisms and their relation to primary sequence remain obscure. Using extensive molecular dynamics simulations of a β-hairpin with key sequence features similar to those of >13,000 β-hairpins in full proteins, we provide significant insights on the entire pre-folding dynamics at single-residue levels and describe a single, highly coordinated roll-up folding mechanism, with clearly identifiable stages, directing structural progression toward native state. Additional simulations of single-site mutants illustrate the role of three key residues in facilitating this roll-up mechanism. Given the many β-hairpins in full proteins with similar residue arrangements and since β-hairpins are believed to act as nucleation sites in early-stage folding dynamics of full proteins, the topologically guided mechanism seen here may represent one of Nature's strategies for reducing early-stage folding complexity.

D ynamics of relatively open structures of polypeptides holds clues to the optimal search Nature has coded in the primary sequences as the polypeptide seeks its native structure. One would expect a directed, albeit statistical, dynamics in the vast conformational space far from the folding funnel for resolving Levinthal's paradox and to explain the significant reduction of the folding time observed in practice. What mechanisms drive such an optimal search and how the primary sequences aid such mechanisms are thus key questions in protein folding.

Although how proteins reach native states is of paramount importance, there is little or only limited consensus on the underlying mechanisms, even for secondary structures. Early studies interpreted measurements on folding kinetics as sequential transitions from denatured to native state, possibly through intermediate state(s)[1,2]. However, experimental and computational advances have revealed heterogeneous intermediates with different folding rates, leading to the hypothesis that folding can be achieved through multiple pathways[3–5]. Similarly, the concept of folding funnels on free energy landscapes was introduced to describe parallel folding processes[6–8]. The funnel theory is however still debated, and the existence of individual unrelated pathways has been challenged[9,10]. Furthermore, although these energy-landscape-oriented models are useful abstract descriptors of the statistical mechanics behind the process, *per se* they provide no physical insights on the underlying structural transitions toward the native state.

Studying single-domain proteins or polypeptides (SDPs) can shed important light on the above issue not only due to the much reduced structural complexity of SDPs, but also because the secondary structures formed by SDPs play crucial roles in promoting the formation of higher-order structures in large proteins. A prime example is the case of β-hairpins, which are believed to act as nucleation sites during early stages of the folding of full proteins[11–13]. The focus on β-hairpins has so far been on the final stages of folding, *e.g.*, hairpin formation and side-chain rearrangements[14,15]. In contrast, analyses of open conformational states continue to pose a major experimental and computational challenge[16] since such states occupy a large and sparsely populated portion of the conformational space and are structurally diverse. As a result, dynamics further away from the 'folding funnel' remains largely unexplored and poorly understood. Nevertheless, as the early dynamics and collapse are associated with an entropic loss that is likely to contribute significantly to the total folding free energy, the critical role they play in SDPs is expected to influence folding dynamics and timescales of larger proteins as well[17].

To delineate the entire β-hairpin folding process and thereby identify details about mechanisms that can reduce the complexity of the early-stage dynamics of full proteins, we have carried out extensive conventional molecular dynamics (MD) simulations totaling >10 µs on a model polypeptide, namely, a β-hairpin named chignolin (GYDPETGTWG; Figure 1a)[18]. We have chosen chignolin in view of the fact that more than 13,000
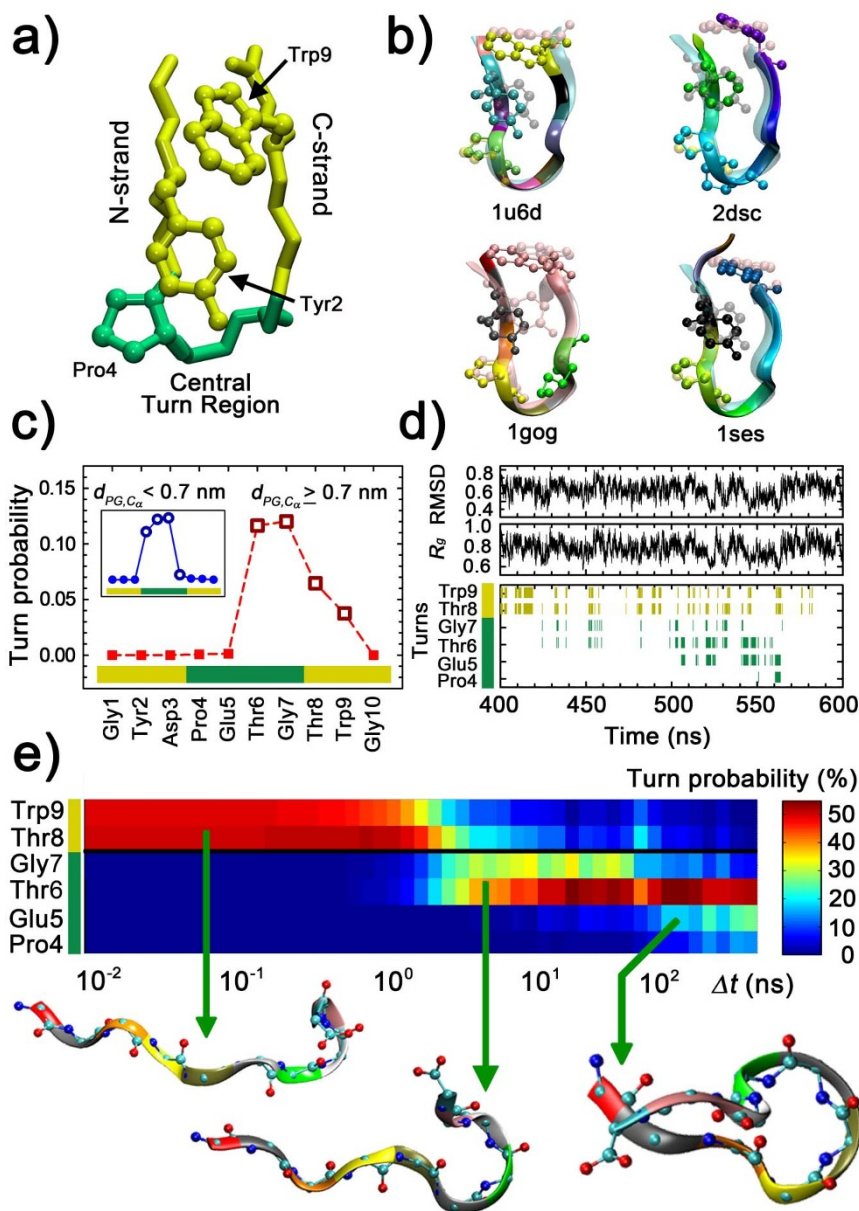
Figure 1 | Dynamics of formation of residue-level turn structures and native turn in chignolin. a) Folded chignolin, showing the native turn (NT) at the central turn region (emerald green) and the amino- and carboxyl-terminal strands (N-strand and C-strand, respectively; lime green). b) Four examples (coloured, solid structures) taken from the more than 13,000 β-hairpins which can be found in full proteins and which show features similar to chignolin (gray, transparent structures). c) Probability of residue-level turn structure formation when NT has not formed. NT is considered to form when the Cα-Cα distance between the corner residues Pro4 and Gly7, $d_{PG,C\alpha}$, is less than 0.7 nm[35]. The corresponding probabilities when NT has formed are shown in the inset. Probabilities higher than 1% are indicated by open symbols. In both cases, the residues in the N-strand (Gly1-Asp3) never form turns. d) Sample trajectory, showing the root-mean-squared-distance, $RMSD$, as described previously[19] (top panels), the radius of gyration R_g (middle panels), and the occurrences of turns (bottom panels), as identified using DSSP[34], in NT and the C-strand residues. While $RMSD$ and R_g reflect the fluctuations in the overall structural features, the turn occurrences show more coordinated evolution. e) Turn propagation from the C-strand to the NT region in the "roll-up mechanism". The probability of turn conformation is plotted as a function of time lapsed from a specific, pre-determined starting conformation, in this case conformations with turns at Thr8 and Trp9. The dark line marks the boundary of the C-strand from the NT region, Gly7-Pro4. Typical examples of backbone structures in the roll-up process are also shown.

β-hairpins with key primary structures similar to those of chignolin are found in full proteins (see discussion later as well as Figure 1b). Using a systematic statistical examination of conformational progression from extended conformations toward the native structure, we demonstrate that, surprisingly, a single, simple mechanism (which we call the roll-up mechanism) describes the entire pre-folding process. We present both structural and temporal details associated with individual steps in this roll-up mechanism and, in addition, demonstrate that the folding process is guided by a few key residues.

The results thus link the formation of a fundamental higher-order structural element, namely, a β-hairpin, to some of the important residues and their positions in the polypeptide's primary sequence and, moreover, reveal a mechanism that is behind the directed search for the native state. Since, as noted earlier, a large number of β-hairpins share the key features of chignolin at play here and in view of the potential role of β-hairpins as nucleation sites for the folding of full proteins, the kind of topological guidance illustrated by our results may not only be common to many polypeptides but also represent at

least one mechanism by which Nature initiates ordered and efficient structure formation in full proteins.

Results

A central point in previous β-hairpin studies has been to uncover the mechanisms that drive the folding process and, in particular, to examine whether folding follows a zipper-like mechanism or is driven by a hydrophobic collapse with subsequent side-chain-level rearrangements. For chignolin, we have recently shown that formation of the native turn (NT) region (Pro4-Gly7) drives folding and marks the beginning of the downhill zipping stage of the folding cascade[19]. Other recent studies of different β-hairpins have also reported turn-directed mechanisms, providing support for a general turn-centric folding scenario[13,15]. In light of these advances, we scrutinize the largely unexplored dynamics of open or relatively open structures, which precedes the abovementioned native-turn formation, as a logical next step towards understanding the entire conformational search and folding of β-hairpins as well as the key early structure-forming events in full proteins. Since standard parameters, such as radii of gyration, R_g, and root-mean-squared-distances, $RMSD$, provide only coarse descriptions of the overall structural features, we choose instead the occurrence of turn structure at individual residue level to extract finer details on the conformational space and dynamics and progression of structural evolution. As a first result, such residue-specific analyses reveal that (residue-level) turn formation is not limited to the NT residues and that outside the NT region turn formation occurs, interestingly, only in the C-strand (Figure 1c). Moreover, turn formation at Thr8 and Trp9 appears to play a central role in the temporal progression of turn structures when the peptide is relatively extended, and a clear trend emerges in the occurrence of turns leading up to turn formation closer to the NT region, although this progression is generally masked by intermittency (Figure 1d); note that, in comparison, such a turn progression is completely obscured if one restricts oneself to analyses in terms of $RMSD$ and R_g alone (Figure 1d).

To obtain a clearer picture of the dynamics, we employ a segmented trajectory analysis on the large number of events captured in our extensive simulations. From the entire trajectories, conformations are selected on the basis of structural features (or reaction coordinates) of interest and are considered as starting points in the analysis, thus effectively segmenting the trajectories into shorter simulations, all starting with a conformation of interest. The dynamics subsequent to this particular conformation is then analyzed by averaging the parameters of interest as functions of elapsed time.

The usefulness and efficacy of this method in unveiling β-hairpin dynamics are illustrated in Figure 1e, with turns in Thr8 and Trp9 as starting point. The turn-formation probability at each residue after a time lapse of Δt clearly demonstrates that turns at Thr8 and Trp9 lead to gradual turn propagation from the C-terminal to the central turn region. This progression, or *roll-up mechanism*, which is initially aided by Pro4-Trp9 interactions (see, also, Figures 2g–h), does not necessarily lead to NT formation; in fact, many of the roll-up attempts stop mid-way and succumb to turn dissolution or reverse propagation (*i.e.*, a *roll-back*) toward the C-terminal (Figure 1d). Indeed, prolonged Pro4-Trp9 interactions interfere with the formation of native Tyr2-Trp9 interactions and thus slow down NT formation and subsequent folding events (i.e., eventual establishment of native hydrogen bonds and packing of the hydrophobic core). However, on average, turn formation in the C-strand does eventually propagate to Pro4 through Gly7, and the transition from an extended state to a collapsed state with the native turn exhibits a smooth and directed (statistical) march.

Figure 1e also reveals important timescales for key events in the roll-up. For example, given that turns exist at Thr8 and Trp9 only, propagation to turns at Thr6 and Gly7 typically occurs on the order of a few ns, while further propagation to Glu5 and Pro4 occurs about

2 orders of magnitude more slowly. As folding is driven by NT formation[19], the timescale of turn propagation to Pro4 thus provides a lower bound for the overall folding time, estimated to be ~1 μs[17,20].

It is noteworthy that turn propagation occurs only from the C-strand; in fact, the N-strand residues never form turns (Figure 1c). Observations show that Tyr2 makes frequent hydrophobic contacts with Pro4; this distraction of Tyr2 by Pro4 and the attendant increased rigidity of the N-strand reduce the conformational options for the polypeptide, offer less hindrance to Pro4-Trp9 interactions and aid the roll-up along the C-strand, as we discuss later. The role of Tyr2-Pro4 interaction on the rigidity is confirmed by simulations of mutant chignolin-Y2A, which show that the replacement of Tyr by Ala increases the flexibility of the N-strand. As seen from Figure 2a-b, Tyr2 in chignolin exists primarily in extended (β-sheet) conformation, whereas Ala2 in Y2A also exists as less-extended α-helix. As a result of backbone flexibility in Y2A, the N-strand varies in end-to-end length, fluctuating between extended and bend conformations, while the same distance in chignolin shows a mono-modal distribution corresponding only to extended N-strand conformations (Figure 2c–d). Clearly, Tyr2-Pro4 interactions in chignolin help confine the available conformational space by restricting the mobility of Tyr2 and, consequently, the N-strand.

A similar conclusion on the role of Tyr2-Pro4 interactions can be made from simulations of chignolin-P4A (Pro4 replaced by Ala). In fact, replacing Pro4 has a more severe effect than that of replacing Tyr2 due to proline's intrinsic conformational restrictions. P4A substitution allows formation of turns anywhere along the polypeptide (Figure 2e–f), while turns are restricted to the C-strand in non-mutated chignolin. These results suggest that the roll-up mechanism, with its turn-propagation from the C-strand (and not from the N-strand) to the NT region, is heavily influenced by both the rigidity of proline's backbone and the Tyr2-Pro4 interactions.

Proline also increases the likelihood of a stable turn-propagation. In the initial steps in the roll-up, interactions between the side chains of Pro4 and Trp9 (Figure 2g), brought together by the formation of turns at Thr8 and Trp9, reduce the risk of roll-back. To test this hypothesis, we carried out simulations of chignolin-W9A (Trp9 replaced by Ala). Turn propagation is clearly less stable in W9A (Figure 2h) than in chignolin (Figure 1e), as evident from the diversity of turn conformations as time progresses. Hence, Pro4-Trp9 interactions stabilize the roll-up mechanism by impeding roll-back and providing a stable framework for turn propagation toward the NT region.

Discussion

Our results show that the pre-folding dynamics of chignolin is highly ordered and proceeds according to a single, simple roll-up mechanism with clearly identifiable stages. This roll-up begins with formation of a non-native turn microdomain at the C-strand and continues with turn propagation toward the central region, where NT forms. As alluded to earlier, during this process, interactions between specific residues, *i.e.*, between Tyr2 and Pro4 and between Pro4 and Trp9, respectively help reduce the available phase space and consolidate turn propagation. The statistical analysis we introduce here, using examinations of trajectory segments with the same initial structure, is able to identify statistically significant (dynamic) structural evolution as the polypeptide searches for the native turns. Moreover, the analysis also provides relevant characteristic timescales.

It might be tempting to characterize the observed non-native turn microdomains as semi-stable states (somewhat analogous to the misfolded states with mismatched cross-strand hydrophobic contacts that have previously been reported for chignolin at later stages of folding[19,21]). It is important to note, however, that the roll-up mechanism we report in the present paper is simply an efficient search mechanism employed by the polypeptide at early stages, prior to the onset of folding (represented by the previously described folding
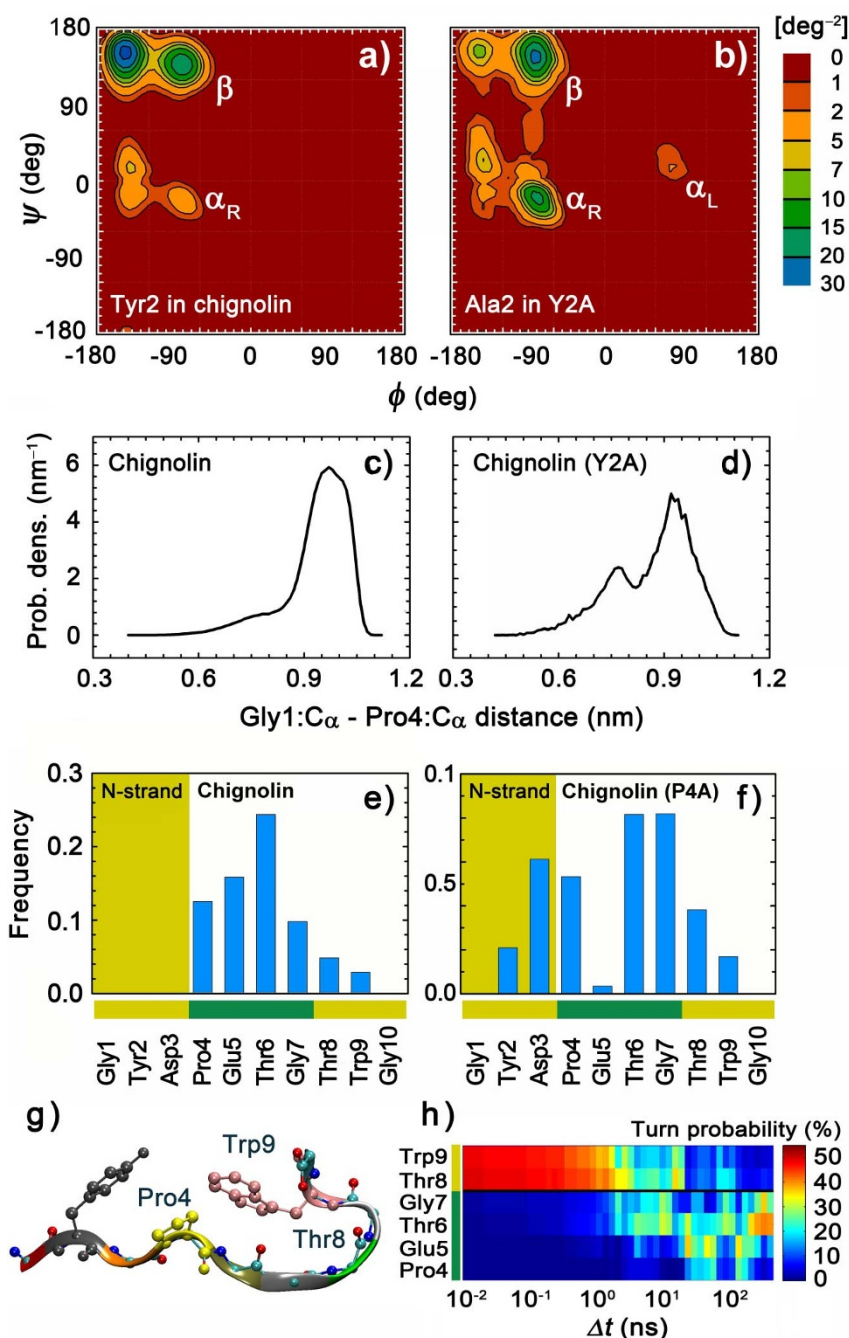
Figure 2 | Effect of single-site mutations on conformational features. a-b) Ramachandran plots of the dihedral angles of (a) Tyr2 (in non-mutated chignolin) and (b) Ala2 (in chignolin-Y2A). Interactions between Tyr2 side chain and Pro4 ring forces Tyr2 to assume an extended conformation with magnitudes of backbone dihedral angles Φ and Ψ both close to $\pm 180°$. In the absence of side-chain interactions a helical conformation emerges at $(\Phi, \Psi) \sim (-80°, -20°)$. c–d) Probability density distribution of the distance between Gly1 and Pro4 Cα-atoms in (c) chignolin and (d) chignolin-Y2A. The distance remains relatively constant at ~ 0.96 nm in chignolin, corresponding to extended conformation for the N-strand. In contrast, without the side-chain interactions the N-strand can also assume a bent conformation, as indicated by the second peak at ~ 0.76 nm. e–f) Frequency of the occurrence of turn structure in each residue in (e) chignolin and (f) chignolin-P4A. Rigidity of Pro4 limits turn formation to the NT region and the C-strand of chignolin, whereas replacement of Pro4 with Ala allows turn formation also in the N-strand (Gly1-Asp3). g) Typical structure with turns at Thr8 and Trp9. The formation of turn structures at Thr8 and Trp9 is facilitated by interactions between Pro4 and Trp9. h) Lack of consistent turn propagation in the absence of Pro4-Trp9 interactions. In the same manner as in Figure 1e, the probability of turn conformation for the ten residues in chignolin-W9A is plotted as a function of lapsed time, again for given conformations with turns at Thr8 and Ala9 as starting point.

steps[19]). The roll-up also describes the *statistical* progression of the structural changes temporally; that is, the roll-up mechanism and the corresponding underlying structures do not represent a single meta-stable state. Conclusions about the existence of semi-stable states on the basis of any particular residue-level turn conformation (or a collection of residue-level turns) should therefore be made with caution. Moreover, if any non-native semi-stable states exist during the early stages, they could have an impact on the timescale of subsequent folding events (e.g., native-turn formation, establishment of native hydrogen bonds and side-chain rearrangements; see ref. 17 for

a discussion of the various timescales involved), but they would have no implications to the roll-up process itself, unless the metastable states interfere with the formation of native-turn structure. Such is not the case for the class of polypeptides represented by chignolin (with respect to key residues). It is also worth noting here that studies on GB1, another β-hairpin, have suggested that a reptation-like movement of one strand with respect to the other can take place in the late folding stages[22,23] (although the underlying mechanism and timescale have not been explored). Again, the early-stage roll-up mechanism we report is different from the above reptation-like motion, which has been reported for transition between a folding intermediate and the folded state at the final stages of folding.

Finally, we ask whether the above roll-up mechanism is likely to be relevant to polypeptides other than chignolin. Hints available in the literature, which are retrospectively illuminating, suggest that the answer is 'yes'. For example, occurrence of specific non-native turns before formation of the native β-hairpin turn has been observed recently in Peptide 1[14], although it remains to be seen if there is a similar continual turn propagation as observed by us. In addition, the similarity of folding times (~ 1 μs) and of where Tyr, Pro, and Trp, which, we found, have a critical impact on the roll-up, are located in chignolin and Peptide 1 could imply that the folding of Peptide 1 is guided by the same overall, simple mechanism. The roles of the residues are also supported by results of another study[24], which found that the N-strand of a modified version of the inhibitor tendamistat β-hairpin, containing Tyr1 and Pro4, was constrained to β-strand conformation, much in line with our observation above for chignolin. Indeed, the side-chain interactions involved in the roll-up may be more general in nature and may consequently be instigated by other amino acid residues with similar characteristics. As alluded to earlier, a search for β-hairpin segments in large proteins containing two aromatic residues separated by proline returns more than 13,000 such β-hairpins (based on a search for 7-to-15-residues-long amino acid sequences of the following type: A1-nX-P-mX-A2, where A1 and A2 are aromatic residues Tyr, Trp, or Phe, X denotes any type of residue, and n and m denote the number of spacer residues between the aromatic residues and Pro, P; see examples in Figure 1b). This large number (with similar positions for the key hydrophobic residues) suggests that roll-up may be a general *topologically guided* mechanism by which the initial stages in the search for the folding funnel are facilitated by interactions between specific residues in the primary sequence.

Formation of small folding microdomains that can grow, collide, and assemble to form larger and more stable native-like structures has been observed experimentally and proposed in various structure-oriented folding models, such as the diffusion-collision model[25,26], the foldon model[27], and the zipping-assembly model[28]. While these models are useful in interpreting experimental and computational results, they only provide insights on large-scale structural transitions, ignoring residue-level interactions that govern the initial formation and dynamics of microdomains. In contrast, our results provide direct evidence for significant involvement of non-native, Pro4-Trp9 interaction and topological guidance in the folding and intra-chain turn propagation as well as the corresponding kinetics. In effect, our work suggests that, although diverse microscopic pathways down the folding funnel can be discerned, macroscopically the complete folding course can be well described by a single, simple mechanism. The frequent roll-up/roll-back dynamics suggests a relatively flat energy landscape far from the folding basin. On the other hand, even where the polypeptide assumes extended conformations, the local landscape must still have non-zero gradient, which ultimately causes the roll-up mechanism. The existence of a well-defined, topologically guided pathway may suggest that such a pathway bridging the denatured and native states represents one of Nature's ways of reducing complexity and enhancing efficiency of folding. Importantly, recent experimental advances in techniques such as 2D IR, UVRR, NMR, and single-molecule FRET offer structural information at unprecedented spatiotemporal resolutions[29,30]. Since the early stages of protein folding are still poorly understood, such technical improvements may provide complementary avenues for elucidating to what extent topological guidance plays a role in the intricate folding dynamics.

Methods

Extensive MD simulations totaling more than 10 μs of simulated time were carried out for chignolin and each of the three single-site chignolin mutants, Y2A, P4A and W9A, in which Tyr2, Pro4, and Trp9 had been replaced with alanine, respectively. All simulations were carried out starting from fully-extended all-trans conformations in a $4 \times 4 \times 4$ nm³ box with about 2,200 water molecules as well as two Na⁺ ions, counterbalancing the charges of Asp3 and Glu5 to provide a neutral system net charge. Short energy minimizations, which settled high-energy interactions, preceded MD simulations at 298 K and 1 bar and atom positions were sampled every 10 ps. Both energy minimizations and MD simulations were done using GROMACS 4.0.3[31]. Water and peptide bond lengths were constrained using SETTLE and LINCS-P, allowing a time step of 2 fs[31]. Electrostatic interactions were resolved with the particle mesh Ewald (PME) method, using a real space cut-off of 0.9 nm with a maximum distance between the PME grid points of 0.12 nm in x, y, and z direction and a fourth-order polynomial for interpolating the charge values at the grid points. The OPLS-AA force field[32] was employed together with the TIP4P water model based on their excellent ability to provide stability of the native state of chignolin, as noted in a recent comparison of force fields[33]. Identification of residue-level turn structure was done with DSSP[34].

1. Kim, P. S. & Baldwin, R. L. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459–489 (1982).
2. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.* **4**, 10–19 (1997).
3. Radford, S. E., Dobson, C. M. & Evans, P. A. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **358**, 302–307 (1992).
4. Krantz, B. A. & Sosnick, T. R. Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. *Nat. Struct. Mol. Biol.* **8**, 1042–1047 (2001).
5. Udgaonkar, J. B. Multiple routes and structural heterogeneity in protein folding. *Annu. Rev. Biophys.* **37**, 489–510 (2008).
6. Leopold, P. E., Montal, M. & Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8721–8725 (1992).
7. Wolynes, P., Onuchic, J. & Thirumalai, D. Navigating the folding routes. *Science* **267**, 1619–1620 (1995).
8. Karplus, M. Behind the folding funnel diagram. *Nat. Chem. Biol.* **7**, 401–404 (2011).
9. Bédard, S., Krishna, M. M. G., Mayne, L. & Englander, S. W. Protein folding: Independent unrelated pathways or predetermined pathway with optional errors. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7182–7187 (2008).
10. Krishna, M. M. G. & Englander, S. W. A unified mechanism for protein folding: Predetermined pathways with optional errors. *Protein Sci.* **16**, 449–464 (2007).
11. Blanco, F. J., Rivas, G. & Serrano, L. A short linear peptide that folds into a native stable β-hairpin in aqueous solution. *Nat. Struct. Mol. Biol.* **1**, 584–590 (1994).
12. McCallister, E. L., Alm, E. & Baker, D. Critical role of β-hairpin formation in protein G folding. *Nat. Struct. Mol. Biol.* **7**, 669–673 (2000).
13. Lewandowska, A., Oldziej, S., Liwo, A. & Scheraga, H. A. β-hairpin-forming peptides; models of early stages of protein folding. *Biophys. Chem.* **151**, 1–9 (2010).
14. Thukral, L., Smith, J. C. & Daidone, I. Common folding mechanism of a β-hairpin peptide via non-native turn formation revealed by unbiased molecular dynamics simulations. *J. Am. Chem. Soc.* **131**, 18147–18152 (2009).
15. Best, R. B. & Mittal, J. Microscopic events in β-hairpin folding from alternative unfolded ensembles. *Proc. Natl. Acad. Sci. USA* **108**, 11087–11092 (2011).
16. Sosnick, T. R. & Barrick, D. The folding of single domain proteins - have we reached a consensus? *Curr. Opin. Struct. Biol.* **21**, 12–24 (2011).
17. Kurniawan, N. A., Enemark, S. & Rajagopalan, R. Crowding alters the folding kinetics of a β-hairpin by modulating the stability of intermediates. *J. Am. Chem. Soc.* **134**, 10200–10208 (2012).
18. Honda, S., Yamasaki, K., Sawada, Y. & Morii, H. 10 residue folded peptide designed by segment statistics. *Structure* **12**, 1507–1518 (2004).
19. Enemark, S. & Rajagopalan, R. Turn-directed folding dynamics of β-hairpin-forming de novo decapeptide chignolin. *Phys. Chem. Chem. Phys.* **14**, 12442–1250 (2012).
20. van der Spoel, D. & Seibert, M. M. Protein folding kinetics and thermodynamics from atomistic simulations. *Phys. Rev. Lett.* **96**, 238102 (2006).
21. Satoh, D., Shimizu, K., Nakamura, S. & Terada, T. Folding free-energy landscape of a 10-residue mini-protein, chignolin. *FEBS Lett.* **580**, 3422–3426 (2006).

22. Wei, G., Mousseau, N. & Derreumaux, P. Complex folding pathways in simple β-hairpin. *Proteins* **56**, 464–474 (2004).

23. Bonomi, M., Branduardi, D., Gervasio, F. L. & Parrinello, M. The unfolded ensemble and folding mechanism of the C-terminal GB1 β-hairpin. *J. Am. Chem. Soc.* **130**, 13938–13944 (2008).

24. Wu, X. & Brooks, B. R. β-hairpin folding mechanism of a nine-residue peptide revealed from molecular dynamics simulations in explicit water. *Biophys. J.* **86**, 1946–1958 (2004).

25. Karplus, M. & Weaver, D. L. Protein-folding dynamics. *Nature* **260**, 404–406 (1976).

26. Karplus, M. & Weaver, D. L. Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Sci.* **3**, 650–668 (1994).

27. Maity, H., Maity, M., Krishna, M. M. G., Mayne, L. & Englander, S. W. Protein folding: The stepwise assembly of foldon units. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4741–4746 (2005).

28. Ozkan, S. B., Wu, G. A., Chodera, J. D. & Dill, K. A. Protein folding by zipping and assembly. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11987–11992 (2007).

29. Kim, Y. S. & Hochstrasser, R. M. Applications of 2D IR spectroscopy to peptides, proteins, and hydrogen-bond dynamics. *J. Phys. Chem. B* **113**, 8231–8251 (2009).

30. Oladepo, S. A., Xiong, K., Hong, Z. & Asher, S. A. Elucidating Peptide and Protein Structure and Dynamics: UV Resonance Raman Spectroscopy. *J. Phys. Chem. Lett.* **2**, 334–344 (2011).

31. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable Molecular Simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).

32. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **105**, 6474–6487 (2001).

33. Matthes, D. & de Groot, B. L. Secondary structure propensities in peptide folding simulations: a systematic comparison of molecular mechanics interaction schemes. *Biophys. J.* **97**, 599–608 (2009).

34. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

35. Venkatachalam, C. M. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* **6**, 1425–1436 (1968).

Author contributions

Additional information