Supporting Information

Maisuradze et al. 10.1073/pnas.1518298112

SI Materials and Methods

Cloning of the FBP28 WW Variants. Point mutations of the variants were introduced by the PCR using the appropriate complementary primers containing the mutation. Constructs were confirmed by DNA sequencing.

Protein Expression and Purification and NMR Assignments. All proteins were expressed as fused to a cleavable N-terminal tag as described in ref. 1. Cells were lysed using a Vibra-Cell VCX 750 Ultrasonic Processor (Sonics & Materials) in a phosphate buffer. The soluble proteins were purified by nickel-affinity chromatography as recommended by the manufacturer (HiTrap Chelating HP Column; GE Healthcare Life Science). The His6-GST fusion protein (polyHistidine-GST protein tag) was removed by protease digestion overnight and purified further by size-exclusion chromatography using HiLoadTM Superdex 30 Prepgrade Columns (GE Healthcare) in either a phosphate or Tris buffer. The final product was analyzed by MALDI-TOF MS. NMR data were recorded at 285 K on a Bruker Avance III 600-MHz Spectrometer equipped with a *z*-pulse field gradient unit and a triple-resonance (1H, 13C, and 15N) probe head. Intramolecular proton distance and side-chain torsion angles restraints were assigned in 2D homonuclear total correlation spectroscopy and NOESY experiments. NMRPipe (2) was used for spectra processing. CARA software (3) was used for spectra analysis and assignment. Spectra used for the calculation were integrated using the batch integration method of the XEASY package.

Structure Determination and Refinement. Crystallography and NMR System (www.mrc-lmb.cam.ac.uk) 1.1 (4) was used for structure calculation. Unambiguously assigned intramolecular distance restraints were derived from NOESY experiments. Torsion angles as well as secondary structure-based hydrogen bond patterns were obtained from the analysis of D_2O exchange and NOESY data acquired at short mixing times. The calculation protocol (hybrid distance geometry-simulated annealing protocol) consists of two iterations of 1 and 200 structures using 100,000 cooling steps. Water refinement of all calculated structures was done with an in-house Aria (5) modified protocol, which uses all experimental restraints during refinement. PROCHECK-NMR (6) was used for structure quality analysis of the lowest-energy structures. Images were generated using UCSF Chimera (7).

Melting Temperatures. Melting temperature determination experiments were carried out using Nano Differential Scanning Calorimetry (TA Instruments) scanning from 20 °C to 100 °C with 1 °C per minute increases. Experiments were performed in either phosphate or acetate buffer. Data analysis was performed using the supplied software. We thank TA Instruments for facilitating the access to this instrument.

UNRES Force Field. In the UNRES force field (8–17), a polypeptide chain is represented as a sequence of α-carbon atoms (C^α), shown in Fig. S4, where θ_i is the backbone virtual bond angle, γ_i is the backbone virtual bond dihedral angle, and α_i and β_i are the angles defining the location of the united side-chain center of residue *i*. The backbone consists of a sequential alternation of α-carbon atoms and $C^\alpha\cdots C^\alpha$ virtual bonds. United peptide groups are at the centers of the $C^\alpha\cdots C^\alpha$ virtual bonds. Side chains are represented by ellipsoids with their centers of mass at the *SC*s, which are connected to the backbone by the virtual bonds $C^\alpha\cdots SC$. The united peptide groups and the united side chains are the interaction sites, and the α-carbon atoms are simply geometric points to define the geometry of the chain and are not interaction sites.

The UNRES force field is based on the potential of mean force (10), also termed the restricted free-energy function of a system, in which the all-atom energy function of a polypeptide chain plus the surrounding solvent are neglected by averaging over the degrees of freedom. The effective energy function of the simplified system is further decomposed (10) into factors coming from interactions within and between given numbers of united interaction sites, which is expressed by Eq. **S1** (15):

$$\begin{aligned}U = {} & U_o(T) + w_{SC}\sum_{i<j} U_{SC_iSC_j} + w_{SC_ip_j}\sum_{i\neq j} U_{SC_ip_j} + w_{pp}^{VDW}\sum_{i<j-1} U_{p_ip_j}^{VDW} \\ & + w_{pp}^{el} f_2(T)\sum_{i<j-1} U_{p_ip_j}^{el} + w_{tor} f_2(T)\sum_i U_{tor}(\gamma_i) + w_{tord} f_3(T) \\ & \times \sum_i U_{tord}(\gamma_i,\gamma_{i+1}) + w_b\sum_i U_b(\theta_i) + w_{rot}\sum_i U_{rot}\left(\alpha_{SC_i}\beta_{SC_i}\right) \\ & + w_{corr}^{(3)} f_3(T) U_{corr}^{(3)} + w_{corr}^{(4)} f_4(T) U_{corr}^{(4)} + w_{corr}^{(5)} f_5(T) U_{corr}^{(5)} \\ & + w_{corr}^{(6)} f_6(T) U_{corr}^{(6)} + w_{turn}^{(3)} f_3(T) U_{turn}^{(3)} + w_{turn}^{(4)} f_4(T) U_{turn}^{(4)} \\ & + w_{turn}^{(6)} f_6(T) U_{turn}^{(6)} + w_{bond}\sum_{i=1}^{nbond} U_{bond}(d_i)\end{aligned} \quad \textbf{[S1]}$$

with

$$f_n(T) = \frac{\ln[\exp(1)+\exp(-1)]}{\ln\left\{\exp\left[(T/T_0)^{n-1}\right]+\exp\left[-(T/T_0)^{n-1}\right]\right\}},\ T_0 = 300\ \text{K}. \quad \textbf{[S2]}$$

The *w* variables are the weights of the multiplying energy terms and can be determined only by optimization of the energy function (17). The terms $U_{SC_iSC_j}$, $U_{SC_ip_j}$, $U_{p_ip_j}^{VDW}$, and $U_{p_ip_j}^{el}$ represent the mean free energy of the hydrophobic (hydrophilic) interactions between the side chain, the excluded volume potential of the side chain–peptide group interactions, and the energy of interactions between backbone peptide groups (which is split into the van der Waals and mean field dipole–dipole electrostatic contributions), respectively. U_{tor}, U_{tord}, U_b, and U_{rot} are the virtual bond dihedral angle potential, two consecutive virtual bond double-torsion potential, virtual bond angle-bending potential, and side chain–rotamer potential. The term $U_{corr}^{(m)}$ corresponds to the correlation or multibody contributions from the coupling between backbone-local and backbone-electrostatic interactions, and the term $U_{turn}^{(m)}$ is the correlation contribution involving *m* consecutive peptide groups. The term $U_{bond}(d_i)$ is the simple harmonic potential, where d_i is the length of the *i*th virtual bond. In this work, we used the force field calibrated with 1ENH and 1E0L.

PCA. A detailed description of the PCA method is available in our previous papers (17–21) and an earlier reference (22); therefore, only a brief outline of the approach is presented here. PCA, a covariance matrix-based mathematical technique, is an effective method for extracting important motions from MD simulations. In PCA, the Cartesian or internal coordinate space is rotated to a new space with new coordinates (PCs), a few of which are sufficient to describe a large part of the fluctuations of a protein. Here, structural fluctuations of the UNRES θ- and γ-angles (MSFs) can be decomposed into collective modes by PCA. The

modes have "frequencies" and directions corresponding to the eigenvalues and eigenvectors of the covariance matrix. The mode k with the largest eigenvalue (λ^k) corresponds to the mode that contributes the most to the structural fluctuations of the protein. The contribution of each angle (θ_i and γ_i) to mode k is called the influence, ν_i^k (21, 23, 24).

Structural Properties of the FBP28 WW Domain Mutants. High-resolution NMR spectroscopy data were used to characterize the structures of the FBP28 WW domain variants in atomic detail.

For residue E27, the mutation E27Y affects the conformation of the second turn of the domain with respect to the WT. Although E27 (side chain) has few contacts and is oriented toward the solvent in WT, in the mutant, Y27 is tightly packed between the turn and the aromatic ring of W8, most probably because of their difference in size and hydrophobicity. The orientation seems to be further stabilized by a Pi-Pi stacking between Y27 and W8 side chains (Fig. S1*D*).

Residue T29 was mutated to Asp and Tyr. D29 was especially prone to aggregation, which was detected in homonuclear 2D spectra, where the presence of broad peaks suggested the formation of multimers at several experimental conditions. The D29 carboxyl group orientation mimics that of the WT T29 hydroxyl, and contacts can be observed with W8 and Y20 aromatic groups (Fig. S1*E*).

In the 2D spectra of T29Y, we observed several duplicated peaks for residues Y20, W30, and E31 all situated in the vicinity of Y29. Addition of 10% (vol/vol) DMSO to the buffer and an increase of the temperature up to 295 K resulted in sharper peaks, and duplicated signals collapsed to one set, representing a single state. The T29Y mutation noticeably rearranges the surrounding residues, causing a sharper turn at the end of the third β-sheet (Fig. S1*F*) by interacting with E31 and P33, and also, reorients the side chain of E27. Moreover, two possible orientations for the Y29 ring satisfactorily agree with experimental data.

FEPs Along θ_i- and γ_i-Angles. In FEPs, the first angle along the sequence is θ_2, and the last is θ_{N-1}, in which N is the total number of residues. The first dihedral angle along the sequence is γ_2, and the last is γ_{N-2} (17, 21, 25).

It should be noted that the FEPs [$\mu(\theta) = -k_B T \ln P(\theta)$ and $\mu(\gamma) = -k_B T \ln P(\gamma)$, where P, T, and k_B are the probability distribution function, the absolute temperature, and the Boltzmann constant, respectively] presented here are effective FEPs, because they are computed from a nonequilibrium probability density and depend on the time duration and the initial conditions of the trajectory. The effective FEP differs from the actual FEP, which is an equilibrium thermodynamic property and should be computed from the entire sets of trajectories (folding and nonfolding). Because of the dependence of the effective FEP on the time duration of the trajectory and the initial conditions, we used the effective FEP to analyze the MD trajectories in detail and extract the reasons why a protein folds or does not fold in a single MD trajectory (17, 21) and why protein A folds with or without a kinetic trap (26). In this work, the effective FEPs were used to explain why L26D and L26W fold through different folding scenarios.

1. Macias MJ, Gervais V, Civera C, Oschkinat H (2000) Structural analysis of WW domains and design of a WW prototype. *Nat Struct Biol* 7(5):375–379.
2. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293.
3. Bartels C, Xia TH, Billeter M, Güntert P, Wüthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* 6(1):1–10.
4. Brünger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54(Pt 5): 905–921.
5. Nilges M, Macias MJ, O'Donoghue SI, Oschkinat H (1997) Automated NOESY interpretation with ambiguous distance restraints: The refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol* 269(3): 408–422.
6. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8(4):477–486.
7. Pettersen EF, et al. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
8. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA (1993) Prediction of protein conformation on the basis of a search for compact structures: Test on avian pancreatic polypeptide. *Protein Sci* 2(10):1715–1731.
9. Liwo A, et al. (1997) A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J Comput Chem* 18(7):849–873.
10. Liwo A, Czaplewski C, Pillardy J, Scheraga HA (2001) Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J Chem Phys* 115(5):2323–2347.
11. Oldziej S, Kozlowska U, Liwo A, Scheraga HA (2003) Determination of the potentials of mean force for rotation about Cα-Cα virtual bonds in polypeptides from the ab initio energy surfaces of terminally blocked glycine, alanine, and proline. *J Phys Chem A* 107(40):8035–8046.
12. Liwo A, Oldziej S, Czaplewski C, Kozlowska U, Scheraga HA (2004) Parametrization of backbone-electrostatic and multibody contributions to the UNRES force field for protein-structure prediction from ab initio energy surfaces of model systems. *J Phys Chem B* 108(27):9421–9438.
13. Oldziej S, Liwo A, Czaplewski C, Pillardy J, Scheraga HA (2004) Optimization of the UNRES force field by hierarchical design of the potential-energy landscape. 2. Off-lattice tests of the method with single proteins. *J Phys Chem B* 108(43):16934–16949.
14. Oldziej S, et al. (2004) Optimization of the UNRES force field by hierarchical design of the potential-energy landscape. 3. Use of many proteins in optimization. *J Phys Chem B* 108(43):16950–16959.
15. Liwo A, et al. (2007) Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. *J Phys Chem B* 111(1):260–285.
16. Kozlowska U, Liwo A, Scheraga HA (2007) Determination of virtual-bond-angle potentials of mean force for coarse-grained simulations of protein structure and folding from ab initio energy surfaces of terminally-blocked glycine, alanine, and proline. *J Phys Condens Matter* 19(28):285203.
17. Maisuradze GG, Senet P, Czaplewski C, Liwo A, Scheraga HA (2010) Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. *J Phys Chem A* 114(13):4471–4485.
18. Maisuradze GG, Liwo A, Scheraga HA (2009) Principal component analysis for protein folding dynamics. *J Mol Biol* 385(1):312–329.
19. Maisuradze GG, Liwo A, Scheraga HA (2009) How adequate are one- and two-dimensional free energy landscapes for protein folding dynamics? *Phys Rev Lett* 102(23):238102.
20. Maisuradze GG, Liwo A, Scheraga HA (2010) Relation between free energy landscapes of proteins and dynamics. *J Chem Theory Comput* 6(2):583–595.
21. Maisuradze GG, Liwo A, Senet P, Scheraga HA (2013) Local vs global motions in protein folding. *J Chem Theory Comput* 9(7):2907–2921.
22. Jolliffe IT (2002) *Principal Component Analysis* (Springer, New York).
23. Altis A, Nguyen PH, Hegger R, Stock G (2007) Dihedral angle principal component analysis of molecular dynamics simulations. *J Chem Phys* 126(24):244111.
24. Cote Y, Senet P, Delarue P, Maisuradze GG, Scheraga HA (2012) Anomalous diffusion and dynamical correlation between the side chains and the main chain of proteins in their native state. *Proc Natl Acad Sci USA* 109(26):10346–10351.
25. Senet P, Maisuradze GG, Foulie C, Delarue P, Scheraga HA (2008) How main-chains of proteins explore the free-energy landscape in native states. *Proc Natl Acad Sci USA* 105(50):19708–19713.
26. Kachlishvili K, et al. (2014) Accounting for a mirror-image conformation as a subtle effect in protein folding. *Proc Natl Acad Sci USA* 111(23):8458–8463.

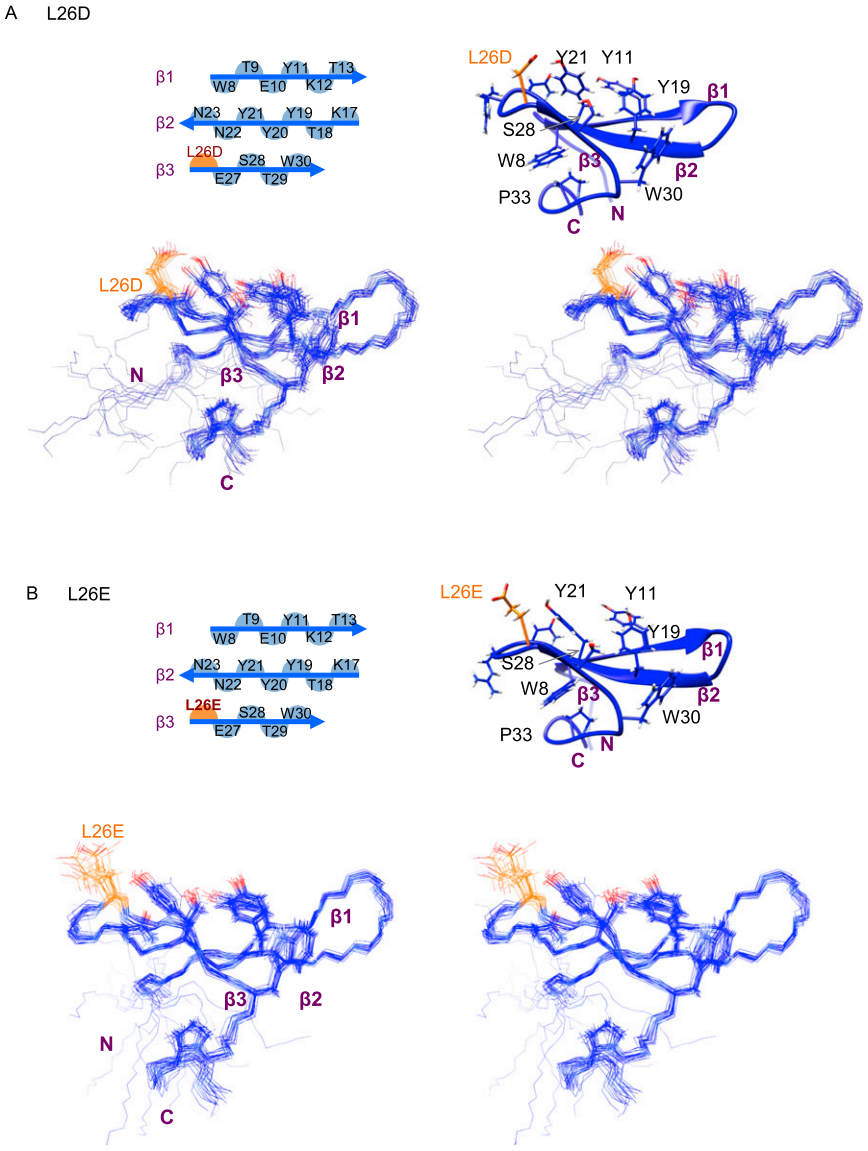

Fig. S1. (Continued)

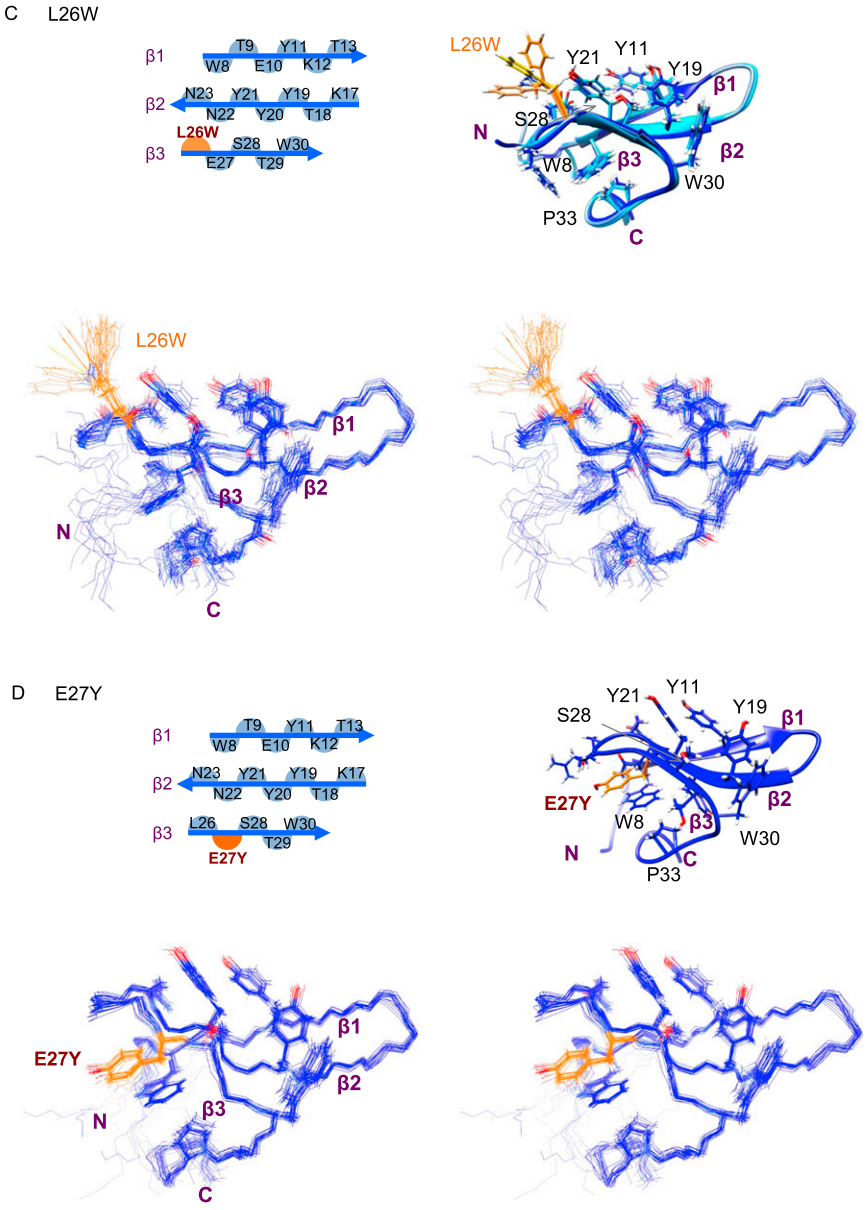

Fig. S1. (Continued)

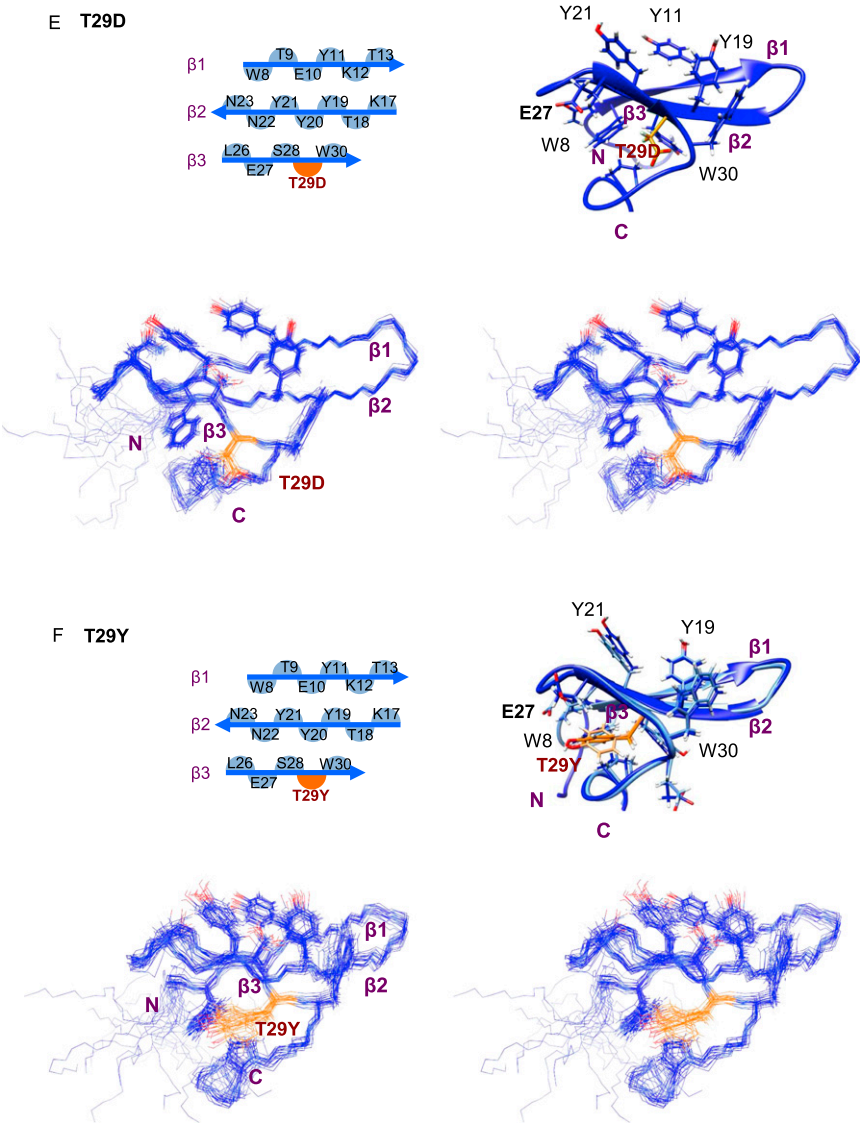

Fig. S1. Structures of six FBP28 WW mutants. (*A*) L26D mutant. (*Upper Left*) A schematic representation of the WW domain. Residues located on the three strands are represented as blue semicircles, with the position mutated highlighted in orange. (*Upper Right*) A cartoon representation of the lowest-energy structure. The N and C termini and the three strands are labeled in dark violet. A few selected side chains and the specific mutation are highlighted and labeled. (*Lower*) A superimposed stereoview representation of the family of structures deposited in the Protein Data Bank. Similar representations for L26E, L26W, E27Y, T29D, and T29Y are displayed in *B*–*F*.

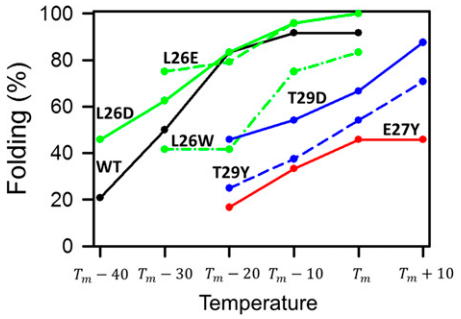

Fig. S2. Percentage of folding trajectories vs. temperature for all mutants and WT. Black line corresponds to WT, green line corresponds to L26D, green dashed line corresponds to L26E, green dash–dot line corresponds to L26W, red line corresponds to E27Y, blue line corresponds to T29D, and blue dashed line corresponds to T29Y.

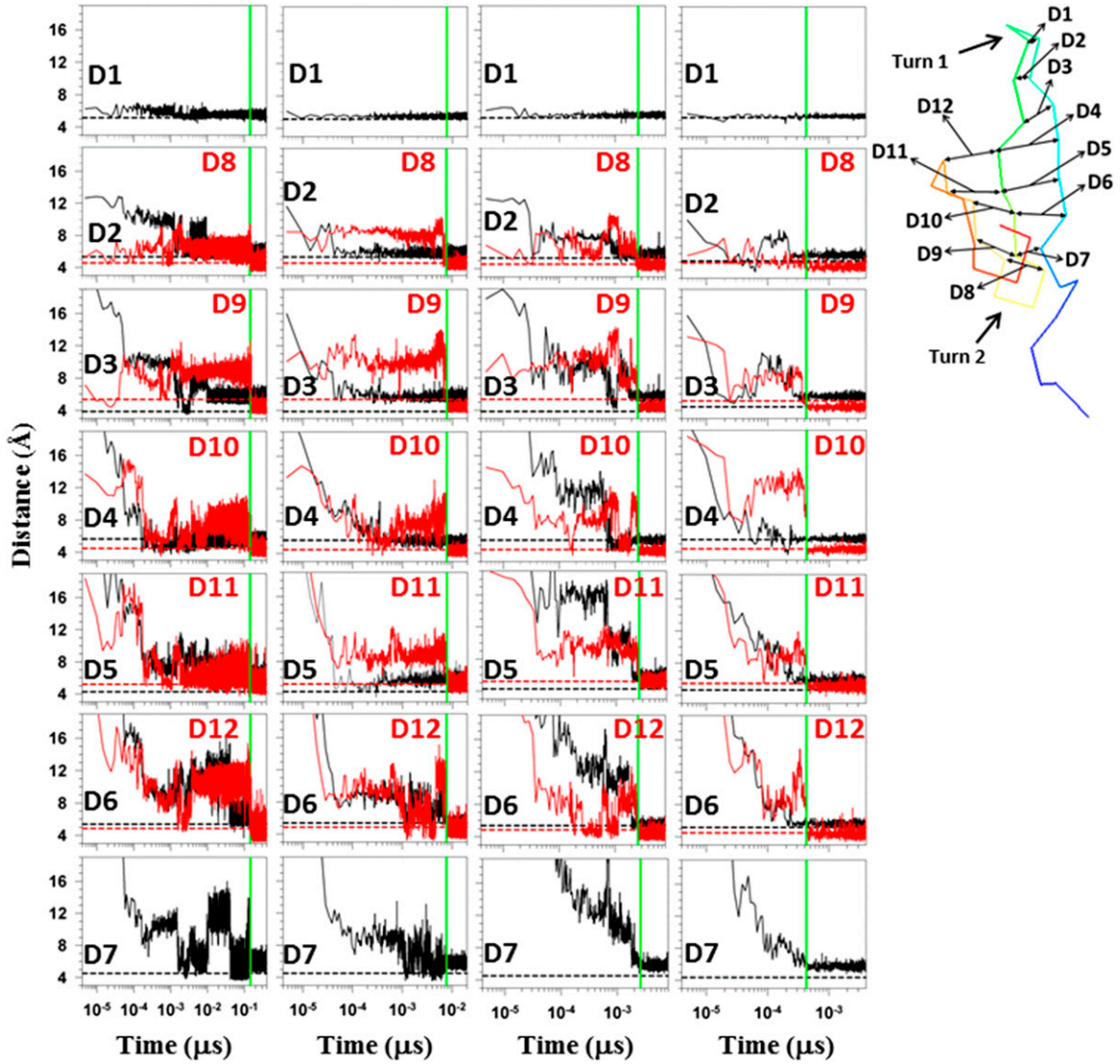

Fig. S3. Distances between C^{α}s of selected pairs of residues of hairpin 1 (D1 → Ala14 and Gly16, D2 → Thr13 and Lys17, D3 → Lys12 and Thr18, D4 → Tyr11 and Tyr19, D5 → Glu10 and Tyr20, D6 → Thr9 and Tyr21, and D7 → Trp8 and Asn22) and hairpin 2 (D8 → Asn23 and Asp26, D9 → Asn22 and Glu27, D10 → Tyr21 and Ser28, D11 → Tyr20 and Thr29, and D12 → Tyr19 and Trp30) as a function of time for L26D mutant (columns 1–3) and L26W mutant (column 4 and D8 → Asn23 and Trp26).

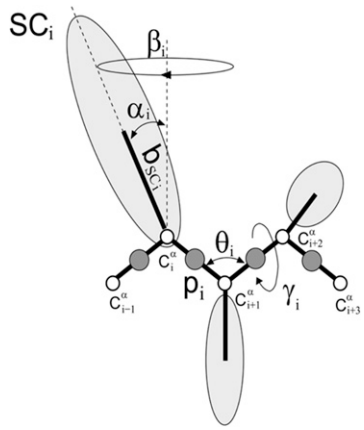

Fig. S4. Illustration of internal coordinates pertaining to the *i*th residue used in Eq. **S1**: backbone virtual bond valence angles (θ_i), backbone virtual bond dihedral angle (γ_i), side-chain virtual bond length (b_{SC_i}), and the angles-α_i and -β_{SC_i} defining the position of the *i*th side chain with respect to the local coordinate frame defined by C^{α}_{i-1}, C^{α}_{i}, and C^{α}_{i+1}. All peptide groups are assumed to be in the planar *trans* configuration with an equilibrium virtual bond length of 3.8 Å (1).

1. Liwo A, et al. (1997) A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J Comput Chem* 18(7):849–873.

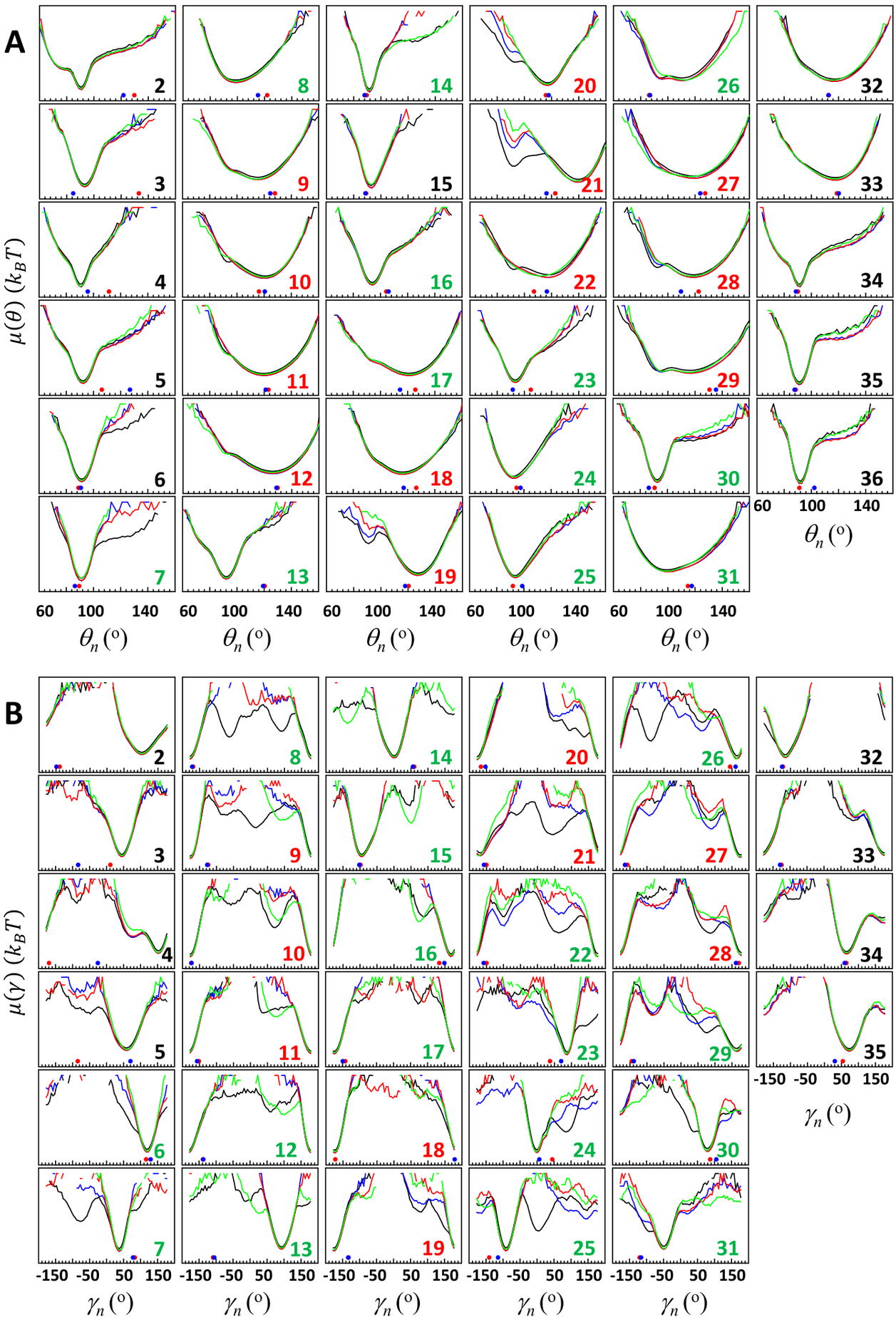

Fig. S5. FEPs, μ(θ), and μ(γ) along the (*A*) θ- and (*B*) γ-angles for three-state, two-state, and downhill folding MD trajectories of L26D and L26W mutants. Black, blue, and red curves correspond to FEPs computed over the entire three-state, two-state, and downhill folding MD trajectories, respectively, for L26D mutant. Green curves correspond to FEPs computed over the entire downhill folding MD trajectory for L26W mutant. The black numbers pertain to FEPs along the θ- and γ-angles that include only residues of the turns and ends, the red numbers pertain to FEPs along the θ- and γ-angles that include only residues of β-strands, and the green numbers pertain to FEPs along θ- and γ-angles that include the residues from both turn and β-strands or from both ends and β-strands. The NMR-derived structural data [small red (for L26D) and blue (for L26W) circles at the bottom of each panel] are computed from the first model.

Table S1. Calculated and experimental melting temperatures of WT FBP28 WW domain and its mutants

Name	T_m (K; calculated)	T_m (K; DSC)
WT	339	339
L26D	344	334
L26E	335	334
L26W	338	328
E27Y	326	328
T29D	332	329
T29Y	325	326

DSC, differential scanning calorimetry.

Table S2. Percentages of folding trajectories with different folding pathways at different temperatures for all six mutants and WT

System and temperature (K)	Hairpin 1 (%)	Hairpin 2 (%)	Hairpins 1 and 2 (%)
WT			
300	100	0	0
310	100	0	0
320	85	15	0
330	75	25	0
340	82	14	4
L26D			
305	90	10	0
315	67	20	13
325	73	27	0
335	70	30	0
345	51	49	0
L26E			
310	82	18	0
320	78	22	0
330	86	14	0
340	65	35	0
L26W			
310	60	20	20
320	57	43	0
330	83	17	0
340	55	40	5
E27Y			
305	75	25	0
315	100	0	0
325	100	0	0
335	80	20	0
T29D			
310	82	18	0
320	92	8	0
330	94	6	0
340	95	5	0
T29Y			
305	75	25	0
315	100	0	0
325	60	40	0
335	70	30	0

Table S3. Structural statistics corresponding to FBP28 WW derivatives

Structural statistics	L26D (L453D)*	L26E (L453E)*	L26W (L453W)*	E27Y (E454Y)*	T29D (T456D)*	T29Y (T456Y)*
Protein Data Bank	2n4r	2n4s	2n4t	2n4u	2n4v	2n4w
BioMagResBank	25678	25679	25680	25681	25682	25683
Restraints used for structure calculation[†]						
Sequential ($\|i - j\| = 1$)	159	172	148	205	173	194
Medium range ($1 < \|i - j\| \leq 4$)	84	87	96	117	77	108
Long range ($4 < \|i - j\|$)	223	257	204	274	235	229
Dihedrals	57	56	50	48	51	69
Hydrogen bonds	10	10	10	10	10	10
rmsd (Å) from experimental						
NOE ($\times 10^{-3}$)	4.46 ± 0.7	6.89 ± 0.3	3.03 ± 0.4	7.52 ± 0.3	6.89 ± 0.5	1.01 ± 0.2
Bonds ($\times 10^{-3}$)	4.83 ± 0.3	6.15 ± 0.2	3.94 ± 0.2	6.78 ± 0.2	6.18 ± 0.3	3.30 ± 0.1
Angles (°; $\times 10^{-1}$)	6.5 ± 0.3	7.4 ± 0.3	5.4 ± 0.2	8.3 ± 0.3	7.3 ± 0.2	4.4 ± 0.2
Coordinate precision (Å; for 20 best structures)[‡]						
Backbone from residue 6 (32 total)	0.76	0.52	0.65	0.54	0.56	0.81
CNS potential energy (kcal/mol)[§]						
Total energy	−1,429 ± 19	−1,338 ± 20	−1,479 ± 24	−1,237 ± 24	−1,375 ± 39	−1,519 ± 20
Electrostatic	−1,641 ± 23	−1,607 ± 23	−1,609 ± 29	−1,584 ± 33	−1,633 ± 46	−1,603 ± 26
van der Waals	−113.1 ± 7	−77.82 ± 9	−142.5 ± 11	−74.94 ± 8	−96.29 ± 8	−155.2 ± 10
Bonds	14.01 ± 1.6	22.77 ± 1.2	9.54 ± 1.2	28.18 ± 1.9	23.06 ± 2.8	6.64 ± 0.44
Angles	68.66 ± 5	89.61 ± 4	47.48 ± 3	115.6 ± 8	88.77 ± 5	32.25 ± 3
Structural quality (% residues found in characteristic regions of Ramachandran plot for 20 best structures)						
In most favored regions of the plot	84.5	85.5	83.6	84.2	83.5	90.2
In additionally allowed regions	13.0	13.2	14.7	12.4	15.0	8.8

*The numbers of the residues in deposited Protein Data Bank and NMR assignment files are corrected by 427 to match with the numbers of the FBP28 protein sequence deposited in the National Center for Biotechnology Information (accession no. AAB80727). Here, the mutated residues are indicated with both nomenclatures.

[†]All restraints are derived from experimental data. Sequential and medium- and long-range distances correspond to NOEs assigned between neighbor residues, across the elements of secondary structure, or between residues separated by more than four residues in the protein sequence respectively. Dihedral restraints are obtained from the measurements of J couplings, and hydrogen bonds were estimated measuring D_2O exchange rates. Stereospecific assignments were obtained from the analysis of NOESY experiments acquired at short mixing times.

[‡]The coordinate precision of the NMR ensemble is defined as the average rmsd between the 20 final structures and their mean coordinates.

[§]Crystallography and NMR System (CNS) is distributed with a force field that includes Lennard–Jones and electrostatic terms for the determination of NMR structures. For the structural determination, we used a full nonbonded representation during the water refinement, including Lennard–Jones, van der Waals, and electrostatic interactions from the Optimized Potentials for Liquid Simulations – XPLOR (OPLSX) force field with minor modifications.