High-resolution mapping of the folding transition state of a WW domain

Kapil Dave[†,¤], **Marcus Jäger**[§,¤], **Houbi Nguyen**[†], **Jeffery W. Kelly**[§], and **Martin Gruebele**[†,¶,*]

[†]Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA

[§]Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road BCC255, La Jolla, CA 92037, USA

[¶]Department of Physics and Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA

Abstract

Fast-folding WW domains are among the best-characterized systems for comparing experiments and simulations of protein folding. Recent microsecond-resolution experiments and long (totaling milliseconds) single-trajectory modeling have shown that even mechanistic changes due to mutation in folding kinetics can now be analyzed. Thus, a comprehensive set of experimental data would be helpful to benchmark the predictions made by simulations. Here, we use T-jump relaxation in conjunction with protein engineering and report Φ values as indicators for folding transition state structure for 65 side chain, 7 backbone hydrogen bond and 6 loop 1 deletion and/or insertion mutants of the 34-residue hPin1 WW domain. 45 cross-validated consensus mutants could be identified that provide structural constraints for transition state structure within all substructures of the WW domain fold (hydrophobic core, loop 1, loop 2, β sheet). We probe the robustness of the two hydrophobic clusters in the folding transition state, discuss how local backbone disorder in the native state can lead to non-classical Φ_M values ($\Phi_M > 1$) in the rate-determining loop 1 substructure, and conclusively identify mutations and positions along the sequence that perturb the folding mechanism from loop 1-limited towards loop 2-limited folding.
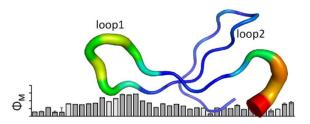
Graphical abstract

[*]Correspondence: M. Gruebele, Tel.: 217-333-1624; mgruebel@illinois.edu.
[¤]These authors contributed equally to the work

Introduction

WW domains are β sheet modular protein domains of 30–65 residues in length that modulate specific interactions with proline-rich protein ligands. WW domains have proven to be an excellent model for ultrafast folding experiments, for mechanistic experimental studies on the folding of a simple β sheet structure, and for benchmarking computational folding scenarios [1–3].

The best characterized natural WW domains to date are the hPin1 WW domain from human peptidyl-prolyl *cis-trans* isomerase Pin1 [3], and the FBP28 WW domain from formin-binding protein 28 [4], with limited data available for a third WW domain, the hYAP65 WW domain from human Yes-Kinase associated kinase [5]. Mutational Φ_M value analysis suggest that formation of loop 1 in WW domains is mostly rate limiting (Φ_M values > 0.80) [6].

In FBP28 WW and hYap65 WW, the N-terminal loop 1 sequence folds into a 5-residue type-I G-bulge turn, the statistically preferred conformation among WW domains. The longer, intrinsically disordered 6-residue loop 1 in hPin1 WW appears to have been selected for function. Its unusual loop conformation (type II-turn intercalated in a 6-residue loop) may position the side chains of residues S16 and R17 for optimal ligand binding [7]. Replacing the hPin1 loop 1 with the turn of FBP28 WW to make the FiP WW domain increases stability by up to 7 kJ/mole and speeds up folding from ~ 80 μs to ~ 13 μs, but compromises function [7]. A similar frustration of folding by function has also been observed in other cases, such as frataxin [8].

For WW domains with their loop 1 substructure optimized for folding thermodynamics and kinetics, formation of loop 2 becomes competitive as the rate-limiting step for folding. Indeed, further optimization of the loop 2 sequence in FiP (FiP N30G/A31T/Q33T, FiP-GTT hereafter) produced a WW domain with a folding relaxation time of ~ 4 μs, approaching the speed limit for folding [9].

Here we report an in-depth study of temperature jump kinetics for 78 mutants of the hPin1 WW domain (Table 1) that also includes data from two more limited, previous Φ value analyses [6, 7, 10, 11]. 45 mutants were amenable for Φ_M value analysis, providing energetic constraints for structural mapping of the folding transition state of hPin1 WW. Multiple side

chain substitutions at some key sequence positions (e.g. within the hydrophobic cores or loop 2) allow us to calculate error-weighted average Φ_M values that are more likely to be a robust representation of transition state *vs.* native state free energy changes than single (e.g. Ala) substitutions. We also identify substitutions that are not suitable for Φ_M value analysis, and discuss the reasons. This approach has been used by Davidson and co-workers to investigate 'conservatism' of substitutions at several sites of the SH3 domain [12].

Although wild type hPin1 WW and its variants fold more slowly than the redesigned loop 1 variant FiP, their folding rates are still in the microsecond range that is now within the reach of fast folding simulations. As computation of folding in the 50–500 μs range becomes feasible, we believe that the data presented in this study will prove to be a rich resource for detailed comparisons, providing constraints on mechanisms and rate changes deduced from molecular dynamics simulations, which are still debated in the literature [9, 13–15].

RESULTS AND DISCUSSION

After a brief review of hPin1 WW structure and native state interactions (Fig. 1, section 1), we begin our discussion of the results in section 2 with the mutational phi-value (Φ_M) analysis, focusing on which mutants are likely to be reliable reporters for transition state structure (Fig. 2). Next, a temperature-dependent phi-value (Φ_T) analysis is used in section 3 to identify mutations that perturb the folding mechanism and whose perturbing effect escapes detection by inspection of the mutational Φ_M values only (Fig. 3). The consensus set of 39 non-perturbing mutants with reliable Φ_M values is employed in section 4 to analyze the transition state structure of hPin1 WW (Figs 4–7). Section 5 looks at various loop 1 insertion and deletion variants within the rate-limiting loop 1 substructure (Fig. 8). A hypothetical "hybrid" Φ_M map for the ultrafast folding hPin1 WW variant FiP (Fig. 9) to benchmark recent molecular dynamics simulations concludes the paper.

1. Overview of hPin1 WW structure and native state interactions

Two types of interactions help stabilize and specify the three-stranded β sheet structure of the hPin1 WW domain. The first type is mediated by the side chains of conserved hydrophobic residues that form two segregated hydrophobic clusters, one on each side of the β sheet (Fig. 1a). The second type of interaction involves a network of 10 backbone-backbone and 4 backbone-side chain hydrogen bonds (Fig. 1b).

Hydrophobic cluster 1 is formed by the side chains of residues L7, P8, W11, Y24 and P37. The N-terminal Trp (W11 in hPin1 WW) and the C-terminal Pro (P37 in hPin1 WW) are absolutely conserved in WW domains. Mutation of residues W11, Y24 and P37 to Ala or Leu in hPin1 WW results in partially unfolded, or fully unfolded protein, even at low temperature (4° C) (Fig. 1c and [10]). As hydrophobic cluster 1 does not contribute to ligand binding, these medium-long range side chain interactions appear to have evolved to maximize thermodynamic stability of hPin1 WW, rather than its biological function.

Hydrophobic core 2 lies on the ligand-binding face of the three-stranded β sheet, and is formed by the side chains of residues R14, Y23 and F25 (Fig. 1a). These residues are only moderately conserved in WW domains, presumably because hydrophobic core 2 contributes

to ligand binding. Ala mutations of residues 14, 23 and 25 in hPin1 WW, although severely destabilizing the native state ($\Delta G_f \sim 9$ kJ/mole) (Fig. 1c), allow folding into the native state structure under the most favorable folding conditions (4 °C).

Using amide-to-ester mutagenesis, we showed that the degree of destabilization of the native state upon eliminating a backbone hydrogen bond is strongly context-dependent [16]. Hydrogen bonds near the two loop substructures are less influential than hydrogen bonds that are protected within a hydrophobic core. The side chain amino group of N26 (β strand 2) forms a hydrogen bond with the backbone carbonyl group of P9 and to the indole ring of W11, thus linking β strands 1 and 2 of the three-stranded β sheet. Like the hydrophobic core 1 residues (W11, Y24 and P37 in hPin1 WW), the Asn in strand 2 (N26 in hPin1 WW) is highly conserved among WW domains and N26A or N26L mutations unfold hPin1 WW (Fig. 1c) [10].

2. Φ_M-value analysis

The mutational $\Delta\Delta G_f^{\dagger} / \Delta\Delta G_f$) quantifies changes in the free energy of activation ($\Delta\Delta G_f^{\dagger}$) relative to the ground state free energy of folding ($\Delta\Delta G_f$) between wild type and mutant proteins [17, 18] Computational modeling of Φ_M values is now possible for WW domains [14, 19], making direct comparisons with experiments achievable.

To obtain accurate Φ_M values that truly represent transition state energetics, one must design non-disruptive mutants that differ sufficiently in thermodynamic stability from the wild type reference protein [20–23], but are not so different that the folding landscape is substantially altered. A generally accepted strategy for Φ_M value analysis is to use conservative hydrophobic deletion mutations (e.g. Ile/Leu → Val → Ala ; Thr → Ser; Phe → Leu → Ala). This strategy avoids mutants that increase side chain size or introduce new functional groups (i.e. Ser → Thr, Phe → Trp), as well as mutation of solvent-exposed charged residues with long-range electrostatic interactions and/or protein-solvent interactions (e.g. Glu → Ala, Tyr → Phe). Several of the mutations that we employed in our previous side chain Φ_M analysis of hPin1 WW [6] do not meet these requirements. This has been discussed in detail in the literature [22].

One in four mutants studied here has a thermodynamic stability very close to wild type hPin1 WW ($\Delta\Delta G_f < 1$ kJ/mole, $\Delta T_m < 2.5$ °C, with a typical error in T_m of $0.5 - 1$ °C). These mutants were excluded from the Φ_M analysis discussed herein. Their thermodynamic and kinetic data (Table 1) should nonetheless provide a valuable resource for benchmarking upcoming molecular dynamics simulations because most of these mutants fold on the microsecond to millisecond time scale, accessible to all atom explicit [24], implicit [14] and coarse grained simulations [25]. We calculated Φ_M values at three representative temperatures (50 °C, 55 °C and 60 °C) (Table 1), where experimental data was available for almost all mutants without the need for error-prone extrapolation. For some of the more stable loop 1 deletion variants, we only report Φ_M values at 55 and/or 60 °C.

Outliers in the analysis—At 55 °C, the Φ_M values of the mutants that potentially qualify for Φ_M analysis ($\Delta\Delta G_f < 1$ kJ/mole and $\Delta T_m < 2.5$ °C) range from −0.20 (L7I) to 2.56 (S16A) (Fig. 2A, Fig. 2B, Table 1). With the exception of some loop 1 mutants that only

slightly destabilize the domain, there is no correlation between the magnitude of a Φ_M value and the extent of destabilization (G_f in Fig. 2A and Fig. 2B). Except for mutants E12Q, I28A, and Y23F, the estimated error in Φ_M was less than 10 %. A surprisingly high fraction of mutants yield Φ_M values that lie outside the classical range of Φ_M values (in particular $\Phi_M > 1$). Almost all mutants with non-classical Φ_M values map to the hydrophobic core 1 and loop 1 substructures in native hPin1 WW, pointing to the importance of these substructures for transition state energetics. Mutant L7I yields the only negative Φ_M value, which is, however, not supported by the L7A and L7V mutations (Fig. 2A). Also the large Φ_M value of V22A (β strand 2) can neither be cross-validated by Φ_M values of immediate sequence neighbors (R21A/H, Y23L/A) nor by its cross-strand neighbor (M15A, β strand 1). Finally, the Φ_M value of Y23F is almost twice as high as the Φ_M values of Y23L and Y23A that target the same residue (Fig. 2A). Y23F deletes a solvent-exposed hydroxyl-group that should not affect the side chain packing of hydrophobic core 1. Its unusual Φ_M value most likely reports on changes in solvation, rather than packing of the core. Mutants L7I, V22A and Y23F were thus excluded from further analysis.

Probing key residues for stability by multiple mutations—Several residues critical for thermodynamic stability, i.e. R14, Y23 and F25 that constitute hydrophobic core 2 (Fig. 1a), and T29 in loop 2 of hPin1 WW (Fig. 1b), were probed by multiple mutations (vertical Φ_M analysis). We find excellent agreement between the Φ_M value of the non-conservative mutants R14F/L and the classical R14A mutant, and the Φ_M values of the Leu and Ala mutants of F25 differ by 0.10 units (Fig. 2a, Table 1). This is clear evidence that hydrophobic cluster 2, although moderately conserved among WW domains, is rather robust towards perturbation by single side chain modifications.

Loop 2 of hPin1 WW is formed by residues H27-N30, and adopts a α_R-α_R-α_R-α_L, or $\pi_{\alpha L}$-conformation, with the first three residues being in a right-handed helical conformation, and N30 being in a left-handed helical conformation. The $\pi_{\alpha L}$-conformation is very common among four residue loops and is also found in the homologous hYap65 and FBP28 WW domains. We probed the contribution of T29 to transition state structure and energetics by the three classical mutations T29S/A/G. The non-conservative T29D mutation was also included in the analysis, as T29D is found in the homologous hYap65 WW domain, and T29D was utilized in our first Φ_M analysis study of hPin1 WW [6].

The Φ_M value of T29A (0.49 ± 0.01) is closest to the error-weighted average Φ_M value (0.53), with T29D yielding a slightly lower value ($\Phi_M = 0.44 \pm 0.01$) while T29S ($\Phi_M = 0.69 \pm 0.02$) and T29G ($\Phi_M = 0.79 \pm 0.01$) yielded higher values. Of all these, only the glycine mutant lies more than a standard deviation from the average. We also studied a double-mutant, I28N/T29G, which replaces the base of the helical $\pi_{\alpha L}$-turn with a sequence (Asn-Gly) that has a high propensity to form a tight 4-residue type-I' turn, a common loop type seen in hairpin structures. I28N/T29G is one of the most destabilized loop 2 mutants ($G_f = 8$ kJ/mol) and has a large Φ_M value (0.96 ± 0.01). The larger Φ_M value shows that loop 2 can become rate limiting when destabilized, moving the transition state towards the native state. As shown in the next section (Φ_T analysis), mutants T29G and I28N/T29G are perturbing mutants in that they shift the folding transition state with respect to wild type

hPin1 WW, so both mutants are not reliable reporters of the unperturbed wild type transition state structure.

Perturbation of hydrophobic cluster 1 disrupts the folding transition state— Molecular dynamics simulations of the fast-folding FiP variant of hPin1 WW suggest that hydrophobic cluster 1 is only weakly formed in the transition state. The simulated Φ_M values for hydrophobic core 1 residues (L7: -0.30 ± 0.50, P8: -0.3 ± 0.1, W11: ~ 0.4, Y24: 0.32 $\pm$ 0.1, P37: ~ 0) suggest that the native W11-Y24 side chain interaction is partially developed in the folding transition state, while other hydrophobic core contacts (e.g. P37 sandwiched between W11 and Y24 (SI Fig. 1)) must develop after crossing the folding barrier [17, 26, 27].

Because of its importance for stability (Fig. 1c), hydrophobic cluster 1 proves to be difficult to map experimentally by Φ_M analysis. Even though the negative Φ_M value of L7I (within error) agrees with the value from simulations, its Φ_M value is not supported by L7A and L7V mutations. Mutating residues W11, Y24 and P37 to either Ala or Leu resulted in unfolded proteins. Mutants P8A, W11F and Y24W, although (severely) destabilized, unfold cooperatively upon heating but yield non-classical Φ_M values significantly higher than the Φ_M values of other hydrophobic core 1 mutations (L7I/A/V, G10A, Y24F). As the W11F mutant of hPin1 WW folds into a native-like structure with a rigid core (SI Fig. 2), and because the conservative W11F mutation is unlikely to perturb unfolded state structure significantly, the high Φ_M value of W11F most likely results from a perturbation of transition state energetics, rather than ground state effects. The Y24W mutation replaces the phenol-moiety of Y24 with the indole ring of Trp. The larger side chain enables "gain-of-interactions" in the denatured and transition state ensembles, as well as steric clashes in the native state that are not present in the wild type protein. The Φ_M values of mutants G10A (0.57 ± 0.02) and Y24F (0.68 ± 0.02) agree reasonably well with simulation, but we observed that neither mutation is ideal for transition state mapping. Surface-exposed G10 acts as a hinge residue in hydrophobic core 1 formation, so it does not contribute to the side chain packing of the hydrophobic core *per se*, and Y24F removes a solvent-exposed OH-group without perturbing the side chain packing of the core (SI Fig. 1). Like Y23F in hydrophobic core 2, its Φ_M value may primarily report on changes in protein solvation energetics, rather than genuine hydrophobic core contacts. Unlike the disruptive mutations P8A, W11F and Y24W, mutants G10A and Y24F were included in further analysis.

In summary, the large number of disruptive hydrophobic core 1 mutants, the strong effect of the W11F mutation on the hPin1 WW folding kinetics, and the intermediate Φ_M values of the non-disruptive mutants L7A/V/I, G10A and Y24F, suggest that while hydrophobic cluster 1 is only partially structured in the transition state, it is very important for protein stability.

Non-classical Φ_M values in loop 1—The intrinsically dynamic loop 1 substructure of hPin1 WW (SI Fig. 3) was probed by both side chain and backbone hydrogen bond mutagenesis. Mutation S16s deletes the backbone hydrogen bond between residues S16 and R21, while mutation R17r weakens, but does not eliminate, the backbone hydrogen bond

between residue S16 and S19 (Fig. 1b). Mutants S16G, S19G, S18G/S19G and G20A perturb the native state by changing the backbone entropy.

Supporting our previous hypothesis that loop 1 formation is rate-limiting for hPin1 WW folding, all ten loop 1 mutants exhibit high Φ_M values close to or larger than 1 (Fig. 2B). The highest Φ_M values were calculated for mutants S16A (2.56 ± 0.02) and S16T (1.78 ± 0.02). The Φ_M value of S16A is about twice as high as that of all other loop 1 mutants, and is a clear outlier. From the structure of the folded hPin1 WW domain it is not immediately obvious why S16A would perturb transition state energetics and slow down folding so much, but similar observations have been made with the *fyn*SH3 domain [28], where a T47A substitution produces a Φ_M value twice as high as that of T47S and T47G.

Mutants S16G, R17r, S19G, S18G/S19G and G20A all share Φ_M values > 1 ($\Phi_M = 1.14$–1.43). Mutants S16G, R17r and G20A are significantly less stable than S19G and S18G/S19G, so at least their non-classical Φ_M values cannot be attributed to artifacts due to small differences in the stability between wild type and mutant proteins (G_f). Φ_M values close to 1 are obtained for side chain mutants R21A/H (loop 1/β strand 2 interface) and for mutant S16s that eliminates the backbone hydrogen bond between residues S16 and R21 that closes the 6-residue loop conformation. Except for S16A and S16T, all these mutants are used for further analysis.

3. Φ_T-value analysis

In folding studies that employ chemical denaturants (urea, guanidine hydrochloride) as the perturbation, transition state locations can be calculated from an analysis of the V-shaped folding relaxation rate *vs.* denaturant concentration plot, also known as "chevron plot." The Tanford β_T value from this analysis is an indicator of the relative compactness of the folding transition on the reaction coordinate in terms of solvent accessible surface area [29]. Using temperature as perturbant by analogy [6, 30, 31], a mutant's Φ_T value

$(\Phi_T = \dfrac{\partial \Delta G^{\dagger}/\partial T}{\partial \Delta G/\partial T} = \dfrac{\Delta S^{\dagger}}{\Delta S})$ can be used as a quantitative, entropic reaction coordinate that describes how much the transition state shifts along the reaction coordinate because of the mutation. It is worth emphasizing that the Φ_T value reports on the overall changes in entropy (i.e. it also includes changes in protein solvation), not just protein conformational entropy. Because the Φ_T value is calculated from two derivatives, it is also sensitive to the quality of the raw data with the best results obtained at temperatures close to the midpoint of unfolding (T_m).

We first calculated Φ_T values directly by taking the derivatives of the second order Taylor series in Table 1. Some of the quadratic coefficients have larger errors than others, and this results in unphysical values of Φ_T (SI Fig. 4A), of the temperature of maximal stability T_0 (where G is at a minimum), and of heat capacities. We therefore also analyzed the data by Taylor series expanding the free energy around the temperature of maximal stability using $G = G_0 + G^{(2)} (T-T_0)^2$. This "$\Phi_T$ T0-fit" yields essentially the same Φ_M values as the Taylor expansion about T_m in Table 1 (SI Fig. 4B), and Φ_T values with more realistic T_0 for all proteins, so we opt to discuss the "Φ_T T_m-fit" throughout this paper. For completeness, we summarize the connection between the Taylor expansion and the common Gibbs-

Helmholtz expansion (in terms of the more physical parameters H_0, S_0 and C_P) in the SI, and provide a table of heat capacities (SI Table 4).

Mutations N30A, T29G, I28N/T29G, S32s and W11F had Φ_T values > 0.7 (Fig. 3, dotted horizontal line), which we chose as a reasonable cut-off for distinguishing between conservative and perturbing mutants because the Φ_M values of mutants W11F, T29G and I28N/T29G either stand-out as clear outliers or are not cross-validated by other mutants (Fig. 2B). In these mutants, the transition state shifts closer to the native state such that their Φ_M values are no longer reliable indicators of the unperturbed "wild type" transition state ensemble, and thus must be excluded from consensus Φ_M analysis. Excluding the abovementioned 5 outliers, the remaining mutants fall within a 25 % interval around the average Φ_T value of 0.50 (Fig. 3, horizontal dashed line). Loop 2 mutants in general tend to have higher Φ_T values, indicative that loop 2 can compete with loop 1 for becoming rate-limiting at higher temperatures.

The ±0.2 spread in the transition state locations as quantified by Φ_T is similar to that reported for the FBP28 WW domain, analyzed using Tanford's β_T value [32]. Even though the individual Φ_T values were measured with high precision (error in $\Phi_T \sim 0.02$), the systematic error in Φ_T may be substantially larger. This is best seen when we compare the Φ_T values of multiple mutations for one residue. Mutants R21A and R21H have very similar Φ_M values (0.95 and 0.89) and essentially identical Φ_T values (0.44 and 0.45), while mutants R14A, R14L and R14F also have similar Φ_M values, but their Φ_T values that span 25 %.

The most dramatic shift in Φ_T is found for the I28N/T29G mutant, whose large Φ_M value (0.96 ± 0.02) also poorly agrees with other loop 2 mutants (Fig. 2B, Table 1). The double mutation I28N/R29G replaces the central two residues of loop 2 with a sequence that has a strong propensity to fold into a tight type-I' turn, suggesting that loop 2 is particularly prone to mutations that introduce residues that have a low propensity to adopt the helical α_R-α_R-α_R-α_L backbone conformation that is required to form loop 2. Indeed, the statistically preferred residues at position 29 are Ser and Thr, and at position 30, Arg, Lys, Gly or Asn. Glycines (position 29) and alanines (position 30) are rare, or not found at all among WW domains.

For mutant W11F, the shift in Φ_T is accompanied by a very large Φ_M value that clearly stands out as a outlier from the mutant pool (Fig. 2A), while the perturbing effect (shift in Φ_T) seen for loop 2 mutants T29G, I28N/T29G, N30A and S32s results in more subtle abnormalities in Φ_M that are more difficult to identify by merely looking at the context-dependent Φ_M values alone (SI Fig. 5). A third class of mutants (e.g. P8A, S16A, V22A and Y24W) shows clear outlier Φ_M values, but normal Φ_T values.

4. High-resolution mapping of the folding transition state of hPin1 WW

General features of the transition state—Our approach for mapping the folding transition state of hPin1 WW was to pick the most conservative mutant set with Φ_M values that were not outliers, based on cross-validation by multiple mutations, sequence neighbors, and backbone hydrogen bond neighbors, and whose Φ_T values indicate no excessive shift of the transition state. Thirty-nine mutants (34 side chain and 5 backbone hydrogen bond

variants) fulfill these criteria and form a consensus set for transition state analysis (Fig. 4A, Table 2). Except for S19G and I28V, all mutants had $G_f > 2$ kJ/mol, close to or above the empirical cutoff (> 2.50 kJ/mol) for reliable Φ_M analysis [33], and except for mutants I28A and E35Q/A, statistical errors in Φ_M were small.

Several residues (L7, E12, R14, R21, Y23, F25, I28, T29) in hPin1 WW were probed by more than one side chain mutation. For these residues, we can calculate more robust (and more representative) error-weighted average Φ_M values from the side chain Φ_M values of individual mutations (Table 2). Mapping the (error-weighted average) side chain Φ_M values onto the C_α-backbone of the folded protein reveals that loop 1 (S16-R21) is substantially more structured in the transition state than loop 2 (H27-N30) and hydrophobic cluster 1 (Fig. 4B).

The (error weighted) average side chain Φ_M plot is a smooth function of sequence (Fig. 5A, solid red line), indicating that the formation of transition state structure is governed mainly by local interactions. Even without the outlier mutants S16A/T, a peak at loop 1 is obvious (see SI Fig. 5 for an extended plot, including outliers). While hydrophobic cluster 1 contacts (probed by L7V/A, G10 and Y24F) are essential for hPin1 WW stability, their contribution to the folding rate is small, and folding of hPin1 WW is rate-controlled by the loop 1 substructure that contributes only slightly to thermodynamic stability. The high side chain Φ_M value of the C-terminal E35, although corroborated by two mutants (E35A/Q), may not truly report on transition state structure. E35 is a charged residue and solvent-exposed in the folded protein. Except for mutant S16A, we find good agreement between the Φ_M values of individual Ala mutants and the consensus average Φ_M value (SI Fig. 5).

Correlation between native-state disorder and non-classical Φ_M-values in loop 1—Here we propose the hypothesis that ϕ_M values >1 in loop 1 (see section 2) are due to native-state backbone dynamics. An NMR-solution structure of the apo-form of the isolated WW domain implies that loop 1 is intrinsically dynamic [34] (SI Fig. 3), and this dynamic nature appears to be preserved in the high-resolution X-ray structure (1.35 Å) of hPin1 WW in the context of the full-length hPin1 rotamase (Fig. 5B). Except for M15A in β strand 1, all mutations that yield non-classical Φ_M values > 1 mutate residues that map onto the intrinsically more disordered loop 1 region, and the concordance between the average consensus Φ_M values (Fig. 5A) and the thermal B factors (a convenient measure for native-state conformational disorder) (Fig. 5C) is striking. The reasonable correlation between the local disorder of a loop 1 residue and the magnitude of its Φ_M value (Fig. 5D) suggests that the Φ_M values in loop 1 are shifted upward further, from values near 1 that are indicative of the importance of loop 1 in the transition state, to even larger values indicative of native state disorder. A more disordered loop 1 may better accommodate mutations that change backbone and sidechain entropy or perturb backbone hydrogen bonds, and thus yields a lower G_f (and a higher Φ_M value), if at the same time the transition state is more sensitive to such mutations because other robust structure (e.g. hydrophobic core 1) have not yet formed.

Correlation between side chain and backbone hydrogen bond Φ_M values— Hydrophobic cluster 2 (R14-Y23-F25) that stabilizes the N-terminal β-hairpin is loosely

formed in the transition state, making an average of 73 % of its native contacts in the transition state (R14 = 77 %, Y23 = 72 %, F25 = 69 %, each calculated from the error-weighted average Φ_M, Table 2). The Φ_M value of mutant K13k that weakens the E12-F25 backbone hydrogen bond (0.80 ± 0.02) agrees well with the side chain Φ_M values of hydrophobic core 2 that protects the hydrogen bond from solvent in native hPin1 WW, suggesting that the E12-F25 backbone hydrogen bond and hydrophobic cluster 2 form cooperatively in the folding transition state.

To test whether this correlation between backbone hydrogen bond and side chain Φ_M values generally holds for hPin1 WW, it is helpful to compare the backbone and side chain Φ_M values at the level of individual residues. We thus assign the Φ_M value of a perturbed backbone hydrogen bond to the two residues that form such a bond, not the residue that is mutated to perturb the hydrogen bond (as done in a previous study [16]). For example, mutation S16s eliminates the S16-R21 backbone hydrogen bond by replacing the amide moiety of the M15-S16 backbone peptide bond that acts as a hydrogen bond donor to form the backbone hydrogen bond with the carbonyl moiety of residue R21 with an ester moiety that cannot engage in backbone hydrogen bond formation (Fig. 1B). Here, we assign the Φ_M of the S16s mutant to both residue S16 and R21. Likewise, mutation K13k perturbs, but does not eliminate, the backbone hydrogen bond between residues E12 and F25, by weakening the hydrogen bond acceptor (backbone carbonyl) of E12 (Fig. 1B). Here, however, it would be more correct to assign the Φ_M of K13k not to residue K13 but to residues E12 and F25 that form the backbone H, even though formally, the amide-moiety of residue K13 is mutated.

Overall, we find good agreement between the "residue-assigned" backbone Φ_M values (Fig. 5A, filled blue circles) and the Φ_M values from classical side chain mutation (Fig. 5A, filled red circles), in particular within the hairpin 2 region (Table 2). As the strength of a hydrogen bond is strongly dependent on the distance between the hydrogen bond donor (backbone amide) and hydrogen bond acceptor (backbone carbonyl), even fractional backbone hydrogen bond Φ_M values of ~0.5 imply that loop 2 is highly compact or that the measured fractional Φ_M values within hairpin 2 represent ensemble averages with about 50 % of the molecules having hairpin 2 fully formed in the transition state ensemble (Φ_M ~1), while in the other half of molecules hairpin 2 is disordered (Φ_M ~ 0). Such a scenario has been predicted in less extreme form from Markov-State-modeling of hPin1 WW folding [35–37].

The poor agreement between the side chain and backbone Φ_M values calculated for residue E12 probably stem from the removal of a solvent-exposed charged residue by mutations E12A/Q. Long-range electrostatic effects may play a role instead of just local contacts.

Variation of transition state structure with temperature—Probing the folding kinetics not just at a single temperature, but over a wider range of temperatures (here, 50, 55 and 60 °C), reveals the robustness of the transition state ensemble against thermodynamic stress. Folding studies at various temperatures also identify 'borderline' mutations that perturb the folding mechanism under increased thermal stress, but whose disruptive nature might escape detection under more favorable folding conditions.

On average, the Φ_M values increase by 0.07 units (Fig. 6A) and the Φ_T value increases by 0.15 units (Fig. 6B) upon raising the temperature from 50 to 60 °C (for data, see SI Table 1, Table 2). This suggests that the folding transition state becomes more structured and native-like at higher temperature, and the transition state ensemble shifts along the reaction coordinate closer to the native state, in agreement with Hammond's postulate [38].

A plot of $\Phi_M(60°C)/\Phi_M(50°C)$ *vs.* sequence in Fig. 6C reveals that structure within hairpin 1 (residues 12–25) at best changes only weakly with temperature. In contrast the loop 2 region (residues 27–30), the third β strand (residues 31–34) and hydrophobic core 1 (probed by L7A and L7V) increase by a larger margin and beyond experimental uncertainty. The absolute changes in Φ_M are, however, rather small such that hairpin 1 still dominates transition state structure at higher temperatures. The Ala mutant W34A may show unusual temperature tuning (although it has a large error bar in Fig. 6C), and we speculate on a possible origin in the SI.

Average fraction of native contacts and its temperature dependence—For the set of consensus mutants depicted in Fig. 4A, we calculate an average Φ_M value of 0.68 ± 0.04 at 55 °C, which is higher than the overall average Φ_T value (0.50 at 55 °C, excluding the 5 outliers discussed in sections 3 and 4). Mutants with a higher slope of ΔG *vs. T* (folding cooperativity) have a higher melting temperature (T_m) (Fig. 7A, where ΔG=0 at $T=T_m$ for all mutants). The average slope is +0.0017 kJ/mole/K, indicative of a negative folding entropy $\Delta S = -(\partial \Delta G/\partial T)$, and increases by about 0.1 kJ/mole/K over the 35–60 °C range of melting temperatures. The size-dependence of ΔS for folding has been discussed in the literature [39, 40]. From the temperature dependence of the folding barrier on protein stability (Fig. 7B), we calculate a slope ($\partial \Delta G^\dagger/\partial T$) ≈ 0.0024 kJ/mole/K (0.0028 for all mutants listed in Table 1). The ratio of the two slopes (activated/ground) is ~ 0.70 (0.63 for all mutants listed in Table 1). This value is also higher than the average Φ_T value of 0.50, and suggests that there is a significant unfolding cooperativity effect in the folding transition state, although not as high as the unfolding cooperativity seen in the native protein. The Φ_T value thus seems to slightly overestimate the distance of the transition state to the native state.

5. Φ_M analysis of loop 1 insertion and deletion mutants

Mutant design and structural analysis—We recently designed and biophysically characterized several hPin 1 WW variants in which the wild type loop 1 sequence is replaced by either a 5-residue type-I G-bulge turn (the preferred loop type in WW domains) or tighter, 4-residue type-I' turns that are not found among WW domains [7] (Fig. 8A).

The X-ray structures of the most stable type-I G-bulge variant (var1, or FiP, loop sequence: SADGR) and the most stable type-I' turn variant (var3, loop sequence: SNGR) have been solved at 1.90 and 1.50 Å resolution, respectively. Both variants essentially superimpose with the wild type structure (1.35 Å resolution), except for the redesigned loop 1 region (Fig. 8B). The thermal B factors of the FiP variant are consistently lower than that of wild type hPin 1 WW, while those of var3 are higher (SI Fig. 6). While the difference in the absolute values of the thermal B factors may result from different crystal packings, we note that turn

1 in the X-ray structure of FiP appears to be conformationally rigid, consistent with NMR-solution data of the same turn in its natural FBP28 WW context (SI Fig.3). The 4-residue type-I' turn of variant 3 shows a relative maximum in the B factor similar that of loop 1 in wild type hPin1 WW, suggesting that the type- I' turn, although stabilizing and hastening hPin1 WW folding, is conformationally flexible in the folded protein.

Group Φ_M analysis and Φ_M *vs.* AAG_f correlation—At 60 °C, and using wild type hPin1 WW as the reference protein, we calculate Φ_M values of 0.92 ± 0.01 for FiP and 0.91 ± 0.01 for the related variant 2. Both Φ_M values are cross-validated by the Φ_M value of variant 2 calculated with FiP as "pseudo wild type" reference (0.94 ± 0.05) (Fig. 8D), demonstrating that Φ_M analysis is surprisingly robust towards more severe sequence manipulations that simultaneously alter sequence and local chain topology. The Φ_M values of FiP and related variant 2 also agree well with the Φ_M values of mutants R21A, R21H and S16s (Φ_M = 0.83–0.97) measured in the wild type loop context (Fig. 8C, D). This correlation is remarkable in that the mutants differ by up to 15 kJ/mole in stability. It further implies that in the strictly sequential folding model (loop 1 first, then loop 2) proposed for FiP by Shaw et al., the energy barrier of the second transition (loop 2 nucleation) must be sufficiently small for FiP-variant 2 to yield a Φ_M value = 0.94 ± 0.05 (SI Fig. 7A). The GTT variant of FiP with an optimized loop 2 structure, however, significantly accelerates FiP folding (by a factor of three), suggesting that loop 2 formation in FiP is associated with a non-negligible barrier and rate-limiting for folding (SI Fig. 7B). Both observations are contradictory and difficult to reconcile in the framework of a sequential model, but perfectly compatible with a simple two-state mechanism, as in the latter case, stabilizing loop 1 and loop 2 mutations may additively lower the (single) transition barrier (SI Fig. 7C).

Type-I' turn variants also hasten wild type hPin1 WW folding, but by a smaller margin than in FiP. In contrast, the two Gly insertion variants 6 and 7 (both less stable than wild type) slow down folding, presumably because of an increased entropic penalty to form the longer 7- or 8-residue loop 1 substructure. All four variants yield Φ_M values greater than 1, similar in magnitude to the Φ_M values of wild type mutants S16G, S18G, S18G/S19G and G20A (Fig. 8D). As for wild type hPin1 WW (Fig. 5), increased local backbone dynamics around the type-I' turn may cause the already high Φ_M values to fall outside the classical range.

Hypothetical hybrid Φ_M-map of FiP and comparison with MD-simulations—Φ_M values are determined experimentally as a ratio of logarithms of rates to logarithms of equilibrium constants. This can be simulated directly by computation (using long trajectories or multiple shorter trajectories with Markov analysis to obtain rate and equilibrium constants), or it can be done by examining structure near the transition state (which has a $P_{fold} \approx 1/2$ folding probability) and comparing with native structure (based on native contacts). In principle, the kinetic/energetic method is the more direct comparison, but structural information may have smaller error bars than energy information, so there is a tradeoff between the two approaches. Extensive data sets such as those in the present paper should become amenable to both approaches in the next few years, to test the merits of the structural vs. energetic approach to simulated Φ_M values in detail. Here we present a brief comparison of our results, adapted to the FiP modification (see loop mutants in Table 1 for

example) of WW domain, and comparing with ref. [14], which presents both structure-based (native side chain contacts) and energy based (long trajectory kinetics) Φ_M values. In the case of [14], the difference between experiment and the two computational approaches still exceeds the difference between the computations, so it appears that force field errors currently still dominate over errors caused by the structural approximation.

We assume that replacing the wild type hPin1 WW loop with the FiP loop 1 sequence only affects the local loop 1 energetics. This assumption is justified by the smooth dependence of Φ_M on sequence, and by the nearly superimposable loop 2 and hydrophobic core 1 substructures of FiP and wild type hPin1 WW (Fig. 8B). A hypothetical "hybrid" Φ_M-map can be rendered for the ultrafast-folding FiP variant by combining the loop 1 Φ_M value of FiP variant 2 (0.94 ± 0.05, measured with FiP as the "pseudo wild type" reference) with the non-loop 1 Φ_M values obtained with wild type hPin1 WW (Fig. 9, red symbols and solid red line).

For loop 1 and its immediate sequence neighbors, our putative "hybrid" Φ_M map (60 °C) agrees well with the simulated Φ_M map calculated at slightly higher temperature (75 °C) [14]. This reinforces our hypothesis (previous paragraph) that replacing loop 1 in wild type hPin1 WW with more stable sequences hastens folding without changing the folding mechanism - either loop type is substantially (or fully) formed in the folding transition state. The Φ_M values within the loop 2 region, however, do not agree very well. Here, the experimental Φ_M values clearly suggest more structure within hairpin 2 than the MD-simulation [14]. As loop 2 slightly gains structure with temperature this discrepancy should be even more pronounced at 75 °C (the temperature used for MD-simulations).

Shaw *et al*. argue that the folding mechanism of FiP is a direct consequence of the difference in the thermal stability of the N- and C-terminal hairpins. Although the isolated hairpins fold about one order of magnitude faster than full-length FiP and at similar rates in simulations, hairpin 1 with the optimized loop 1 sequence is significantly more stable (25 % folded hairpin at equilibrium) than hairpin 2 (4 % folded hairpin at equilibrium), such that loop 1 nucleation is expected to kinetically outperform loop 2 nucleation. Although plausible, this model does not take into account the aforementioned significant (approximately 3-fold) increase in the folding rate that is seen experimentally with the GTT-FiP variant.

In hPin1 WW with the unstable and intrinsically flexible 6-residue loop 1 sequence, isolated hairpin 1 is expected to be much less stable, perhaps even less stable than isolated hairpin 2. This would open up three possible folding scenarios:

With both hairpins being similarly unstable, folding could occur through parallel pathways nucleated by either loop substructure (scenario 1), as predicted from Markov-state-modeling of hPin1 WW folding. In this case, the experimentally measured Φ_M values for the loop 1 and loop 2 regions would directly describe the relative flux along either pathway. In the simplest, and most extreme case, the hairpin whose loop segment nucleates folding is fully formed in the transition state ($\Phi_M \sim 1$) while the other hairpin is completely unstructured ($\Phi_M \sim 0$). For loop 2, we find average Φ_M values of ~ 0.60 at 60 °C. Therefore, if that

extreme model applied, one would expect Φ_M values of only ~ 0.40 for loop 1, which is clearly not what we observe experimentally (average $\Phi_M > 0.9$ at 60 °C).

Alternatively, both loop substructures may fluctuate between an open and a closed state, although not necessarily a native-like state, however a native-like N-terminal hairpin is mandatory for barrier-limited folding into the native state (scenario 2). In this model, loop 1 residues will by necessity yield the highest Φ_M values, while the loop 2 Φ_M values will be reporters about the equilibrium ratio of the open and closed hairpin 2 conformations before their interaction with the structured N-terminal hairpin occurs. As loop 2 formation could either occur before or after loop 1/hairpin 1 formation, hairpin 1 would "catalyze" the final transition of hairpin 2 from the closed to the native state. This folding model is unlikely for wild type hPin1 WW domain because an increase in temperature should shift the loop 2 equilibrium towards the open (less structured) conformation, so the loop 2 Φ_M should decrease with temperature, rather than (slightly) increase. It may, however, become a dominant mechanism in fast-folding WW domains such as FiP.

The most likely folding model for hPin1 WW thus remains a two-state folding mechanism, in which folding and docking of the hairpins occurs in a concerted fashion. The measured Φ_M values would then imply that the N-terminal hairpin is mainly formed in the transition state, while the second hairpin and the hydrophobic core are in the process of being formed in the transition state. Two-state folding of not only wild type hPin1 WW, but also the FiP variant, would also better explain why certain FiP variants such as FiP-GTT with stabilizing mutations within loop 2 and β strand 3 speed up its folding despite high Φ_M values near unity in the hairpin 1 turn region.

Conclusions

Φ_M-value analysis can provide valuable information about the structure of folding transition states by correlating changes in mutationally induced stability and folding kinetics. In its simplest manifestation, Φ_M-value analysis can be affected by probe perturbation of the folding mechanism, and by a trickle-down effect of mutations that lowers the structural resolution. Such trickle down effects can arise for instance from native state flexibility, or from solvent interactions that do not report on genuine structure *per se*.

Here we present a comprehensive Φ_M-value analysis with horizontal (sequence), vertical (multiple mutations at a single site) and chemical depth (side chain and "residue-assigned" backbone hydrogen bond mutations) to identify reliable mutations that can act as probes of the folding mechanism. The "conservatism" of mutations with respect to the folding mechanism is ascertained by multiple side chain substitutions at the same site (L7, E12, R14, S16, Y23, Y24, F25, I28 and T29), verification of individual Φ_M values by cross-β strand neighbors (M15 vs. V22, E12 vs. F25), residue assigned Φ_M values from backbone hydrogen bond mutagenesis (e.g. S16A/G/T vs. S16s, N26D vs. H27h) or immediate sequence neighbors (R21-V22-Y23 series), and temperature tuning (outliers in Φ_T).

For some residues (R14, T29), Φ_M values calculated from non-conservative mutations agree well with Φ_M values calculated from more conservative and structurally less perturbative

mutations, while other mutations yield Φ_M values that primarily report on the energetics of polar or charged residues with solvent (e.g. Y23F, E12A/Q, E35A/Q). Another subclass of mutations that target the flexible loop 1 substructure of hPin1 WW (S16G, R17r, S19G, S18G/S19G, G20A) yield Φ_M values that lie clearly outside the classical range ($\Phi_M > 1$). Based on the correlation with X-ray B factors, their high Φ_M values result at least in part from increased local backbone dynamics in the native state.

Although Ala mutations overall appear to be reliable reporters of transition state structure, as often assumed in the literature, we also identify clear outliers (P8A, S16A and V22A). Another Ala-mutant (W34A) shows an unusual dependence on temperature tuning. Its Φ_M value decreases with temperature, suggesting that the smaller Ala residue perturbs non-native interactions that are stable at low temperature, yet nevertheless speed up folding.

Aside from obvious mutant outliers that can be easily identified by cross-validating their Φ_M values with different mutants at the same sequence location, another subset of mutants perturb transition state structure and shift the transition state ensemble to a more native-like ensemble state, as evidenced by large Φ_T values for such mutations. Four of the five mutants that shift the transition state position in Fig. 5 map to the loop 2 region or immediately flanking residues. Although not dominating transition state structure, the wild type sequence of loop 2 can be perturbed sufficiently to affect folding rates. The ease with which the folding mechanism of the hPin1 WW domain can be changed by what appears to be subtle sequence modifications or perturbations of intermolecular forces (e.g. weakening a single, partially solvent-exposed backbone hydrogen bond as in amide-to-ester mutant S32s) argues against two-state folding with a well-defined, robust and narrow transition state and suggests a more complex, multidimensional energy surface with additional local extrema waiting to become rate limiting for folding, as shown experimentally and computationally for the FBP28 WW domain [4, 41]. The hPin1 WW domain is thus an apparent two-state folder, but not by a wide margin.

Using a more expanded set of consensus mutants, a detailed map of the folding transition state was generated that now covers 76 % of the hPin1 sequence (previous coverage: 50 %). Many of our earlier findings are supported in the present study, but some interpretations need to be modified or revisited. Loop 2 and β strand 3, which define the C-terminal hairpin in folded hPin 1 WW, appear to be more structured in the transition state than thought previously, and the discrepancy in the backbone and side chain Φ_M values within the loop 1 substructure can now be attributed to local backbone disorder in the folded protein, rather than a genuine variation in backbone and side chain structure. In fact, by assigning backbone hydrogen bond to the two residues that constitute the bond, we found good agreement between the Φ_M values measured by side chain and backbone hydrogen bond perturbation for most positions.

The mutants with a thermodynamically and kinetically optimized loop 1 substructure agree well with the native-like Φ_M values of the highly destabilized loop 1 variants R21A/H and S16s mutants that perturb the 6-residue wild type hPin1 WW loop. Clearly, in both wild type hPin1 and the redesigned variants, the tip of the loop/turn is fully developed in the transition state. These observations and the fact that stabilizing loop 2 in the already fast folding FiP

domain further speeds up folding by a factor of 3 are difficult to reconcile in a truly sequential (framework) model for folding, making a simple two-state folding mechanism more likely. Alternatively, as suggested by some simulations [35, 42] and experiments [43] of fast-folding WW domains, loop 2 could actually form before or after loop 1, or fluctuate between folded and unfolded conformations before loop 1 forms, while loop 1 remains rate-limiting due to its larger activation barrier. Additional experiments with mutations targeting loop 2 in FiP are needed to further discriminate between these alternatives.

Materials and Methods

Nomenclature

Residues of the hPin1 WW domain are abbreviated by a single capital letter, followed by the number of the residue in the sequence (e.g. W11). Amino acids are also abbreviated using the standard three letter code (e.g. Trp for tryptophan). Classical side chain mutants are indicated by single letter code (e.g. W11F), with the first and second letters representing the wild type and replacing residue, respectively, and the number indicates the sequence position. Non-classical backbone hydrogen bond mutations are also designated by single letter code. The first letter represents the mutated residue, and the same letter in small capitals is used for the replacing residue (e.g. S16s) to distinguish a non-classical amide-to-ester mutation from their classical counterparts.

Protein expression and sample preparation

The wild type hPin1 WW domain and mutants thereof with classical side chain mutations were prepared recombinantly, as described in detail in another publication [10]. hPin1 WW variants with amide-to-ester mutations were synthesized chemically, as described in detail in [16]. Protein identity and purity was ascertained by electrospray mass spectrometry, SDS-PAGE, and reversed-phase HPLC chromatography.

Experimental procedures

Equilibrium unfolding of hPin1 WW was monitored by far-UV spectroscopy at 229 nm as described in detail in [10]. Unfolding transitions were analyzed by using a two-state model, where the folding free energy G_f is expressed by a quadratic Taylor series approximation: $G_\mathrm{f}(T) = G_f^{(1)}(T_\mathrm{m}) \dot{} (T\text{-}T_\mathrm{m}) + G_f^{(2)}(T_\mathrm{m}) \dot{} (T\text{-}T_\mathrm{m})^{(2)}$. The two coefficients $G_f^{(i)}(T_\mathrm{m})$, $i=1\cdots 2$, represent the temperature-dependent free energy of folding, and T_m is the nominal midpoint of thermal denaturation ($G_\mathrm{f}(T_\mathrm{m}) = 0$). The inclusion of the quadratic term was necessary to fit the data of most mutants within experimental uncertainty. For selected mutants, the transition was also analyzed by expressing $G_\mathrm{f}(T)$ in terms of a constant heat capacity formula. As shown previously for the hYap65 WW domain, both procedures yield nearly identical results [31].

Laser temperature jumps around the protein's melting temperature were measured for each mutant as described in detail elsewhere [44, 45]. Briefly, a 10 ns Nd:YAG pulse Raman-shifted in H_2 heated the sample solution by ~ 5–10 °C, inducing kinetic relaxation of the WW domain to the new thermal equilibrium. 285 nm UV pulses, spaced 1 ns apart from a frequency-tripled, mode-locked titanium:sapphire laser, excited tryptophan fluorescence in

the hPin1 WW domain. Fluorescence emission was digitized in 0.5 ns time steps by a miniature photomultiplier tube with a 0.9 ns full-width-half-maximum response time. The sequence of fluorescence decays $f(t)$ was fitted within measurement uncertainty by the linear combination $a_1f_1(t)+a_2f_2(t)$ of decays just before and 0.5 ms after the T-jump. The normalized fraction $f(t)=a_1/(a_1+a_2)$ from $t\approx2$ μs to $t=0.5$ ms was fitted to a single exponential decay $\exp[-k_{obs}\,t]$ where $k_{obs}=k_f+k_u$. Thus the signal extraction and data analysis are consistently two-state.

The observed relaxation rate coefficient was combined with the equilibrium constant K_{eq} to compute the forward reaction rate coefficient $k_f=k_{obs}K_{eq}/(1+K_{eq})$. k_f was measured for several temperatures (typically around 10) below and above T_m, and $G_f^\dagger(T)$ was determined as a function of temperature using the relationship $k_f=A^\dagger.\exp(-G_f^\dagger(T)/RT)$ with the quadratic Taylor approximation $G^\dagger_f(T)=G_f^{\dagger(0)}(T_m)+G_f^{\dagger(1)}(T_m)(T-T_m)+G_f^{\dagger(2)}(T_m)(T-T_m)^2$, as well as expansions about the temperature of maximal stability (T_0), or the Gibbs-Helmholtz formula (see SI). The three coefficients $G_f^{\dagger(i)}$, $i=0\cdots2$, represent the temperature-dependent activation barrier. The frequency of activation $A^\dagger$ was fixed at 500 ns^{-1}, near the lower end of estimates of the folding speed limit [1], and the two coefficients $G_f^{\dagger(1)}(T_m)$ and $G_f^{\dagger(2)}(T_m)$ also incorporate some effects of temperature-dependent solvent friction. Because previous Φ_M analyses utilized a faster *ad hoc* frequency of 50 ns^{-1}, the Φ_M values of published mutants are shifted by a small constant from the recalculated values of these mutants in this study. Least squares fitting was carried out using IGOR Pro (Wavemetrics). Protein visualization was rendered using Pymol and Weblab viewer software packages (Accelerys, San Diego) [46].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Abbreviations used

B factor	thermal B factor, a measure for backbone dynamics from X-ray crystal structures
FBP28 WW	WW domain (residues 1–37) derived from mouse formin binding protein 28
FiP	hPin1 WW variant in which the wild type loop 1 sequence (SRSSGR) is replaced by a sequence that folds into a type-I G-bulge turn (sequence: SADGR)
FiP-GTT	stabilized FiP variant with the triple mutation N30G/A31T/S32T that hasten folding of FiP threefold at the thermal midpoint of unfolding

hPin1 WW	WW domain (residues 6–39) derived from human *cis/trans*-isomerase Pin1
hYap65 WW	WW domain (residues 1–45) derived from human Yes-associated protein 65
MD-simulation	molecular dynamics simulation
NMR	nuclear magnetic resonance
Φ_M	mutational phi value, an indicator for structure in the folding transition state
Φ_T	temperature-dependent phi value, a parameter for mapping the position of the folding transition state along an entropic reaction coordinate

References

1. Kubelka J, Hofrichter J, Eaton WA. The protein folding 'speed limit'. Curr. Opin. Struct. Biol. 2004; 14:76–88. [PubMed: 15102453]

2. Schaeffer RD, Fersht A, Daggett V. Combining experiment and simulation in protein folding: closing the gap for small model systems. Curr. Opin. Struct. Biol. 2008; 18:4–9. [PubMed: 18242977]

3. Cecconi F, Guardiani C, Livi R. Testing simplified proteins models of the hPin1 WW domain. Biophys. J. 2006; 91:694–704. [PubMed: 16648162]

4. Nguyen H, Jäger M, Moretto A, Gruebele M, Kelly JW. Tuning the free-energy landscape of a WW domain by temperature mutation and truncation. Proc. Natl. Acad. Sci. U. S. A. 2003; 100:3948–3953. [PubMed: 12651955]

5. Chen HI, Sudol M. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. Proc. Natl. Acad. Sci. U. S. A. 1995; 92:7819–7823. [PubMed: 7644498]

6. Jager M, Nguyen H, Crane JC, Kelly JW, Gruebele M. The folding mechanism of a beta-sheet: The WW domain. J. Mol. Biol. 2001; 311:373–393. [PubMed: 11478867]

7. Jäger M, Zhang Y, Bieschke J, Nguyen H, Dendle M, Bowman ME, et al. Structure-function-folding relationship in a WW domain. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:10648–10653. [PubMed: 16807295]

8. Gianni S, Camilloni C, Giri R, Toto A, Bonetti D, Morrone A, et al. Understanding the frustration arising from the competition between function, misfolding and aggregation in a globular protein. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:14141–14146. [PubMed: 25228761]

9. Piana S, Sarkar K, Lindorff-Larsen K, Guo M, Gruebele M, Shaw DE. Computational Design and Experimental Testing of the Fastest-Folding β-Sheet Protein. J. Mol. Biol. 2011; 405:43–48. [PubMed: 20974152]

10. Jäger M, Dendle M, Kelly JW. Sequence determinants of thermodynamic stability in a WW domain-An all-β-sheet protein. Protein Science : A Publication of the Protein Society. 2009; 18:1806–1813. [PubMed: 19565466]

11. Deechongkit S, Nguyen H, Powers ET, Dawson PE, Gruebele M. J.W. Kelly. Nature. 2004; 430:101. [PubMed: 15229605]

12. Northey JGB, Maxwell KL, Davidson AR. Protein Folding Kinetics Beyond the Φ Value: Using Multiple Amino Acid Substitutions to Investigate the Structure of the SH3 Domain Folding Transition State. J. Mol. Biol. 2002; 320:389–402. [PubMed: 12079394]

13. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How Fast-Folding Proteins Fold. Science. 2011; 334:517–520. [PubMed: 22034434]

14. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. Science. 2010; 330:341–346. [PubMed: 20947758]

15. Krivov SV. The Free Energy Landscape Analysis of Protein (FIP35) Folding Dynamics. J. Phys. Chem. B. 2011; 115:12315–12324. [PubMed: 21902225]

16. Deechongkit S, Dawson PE, Kelly JW. Toward Assessing the Position-Dependent Contributions of Backbone Hydrogen Bonding to β-Sheet Folding Thermodynamics Employing Amide-to-Ester Perturbations. J. Am. Chem. Soc. 2004; 126:16762–16771. [PubMed: 15612714]

17. Petrovich M, Jonsson AL, Ferguson N, Daggett V, Fersht AR. phi-Analysis at the experimental limits: Mechanism of beta-hairpin formation. J. Mol. Biol. 2006; 360:865–881. [PubMed: 16784750]

18. Guydosh, NR.; Fersht, AR. Protein Folding Handbook. 2005. A Guide to Measuring and Interpreting phi-values; p. 445-453.

19. Weikl TR. Transition States in Protein Folding Kinetics: Modeling Φ-Values of Small β-Sheet Proteins. Biophys. J. 2008; 94:929–937. [PubMed: 17905840]

20. De Los Rios MA, Muralidhara BK, Wildes D, Sosnick TR, Marqusee S, Wittung-Stafshede P, et al. On the precision of experimentally determined protein folding rates and φ -values. Protein Science : A Publication of the Protein Society. 2006; 15:553–563. [PubMed: 16501226]

21. Ruczinski I, Plaxco KW. Some recommendations for the practitioner to improve the precision of experimentally determined protein folding rates and Φ values. Proteins. 2009; 74:461–474. [PubMed: 18655053]

22. Ruczinski I, Sosnick TR, Plaxco KW. Methods for the accurate estimation of confidence intervals on protein folding Φ - values. Protein Sci. 2006; 15:2257–2264. [PubMed: 17008714]

23. Naganathan AN, Muñoz V. Insights into protein folding mechanisms from large scale analysis of mutational effects. Proceedings of the National Academy of Sciences. 2010; 107:8611–8616.

24. Ferrara P, Caflisch A. Folding simulations of a three-stranded antiparallel beta-sheet peptide. Proc. Natl. Acad. Sci. 2000; 97:10780–10785. [PubMed: 10984515]

25. Zhou R, Maisuradze GG, Suñol D, Todorovski T, Macias MJ, Xiao Y, et al. Folding kinetics of WW domains with the united residue force field for bridging microscopic motions and experimental measurements. Proceedings of the National Academy of Sciences. 2014; 111:18243–18248.

26. Fersht AR. Φ value versus ψ analysis. Proc. Natl. Acad. Sci. U. S. A. 2004; 101:17327–17328. [PubMed: 15583125]

27. Sharpe T, Jonsson AL, Rutherford TJ, Daggett V, Fersht AR. The role of the turn in beta-hairpin formation during WW domain folding. Protein Sci. 2007; 16:2233–2239. [PubMed: 17766370]

28. Lin SL, Zarrine-Afsar A, Davidson AR. The osmolyte trimethylamine-N-oxide stabilizes the Fyn SH3 domain without altering the structure of its folding transition state. Protein Sci. 2009; 18:526–536. [PubMed: 19241379]

29. Howland, JL. A guide to Enzyme Catalysis and Protein Folding: Alan Fersht. New York: W.H. Freeman and Company; 1999. Structure and Mechanism in Protein Science; p. 631ISBN 0-7167-3268-8, $53.002001

30. Ervin J, Gruebele M. Quantifying Protein Folding Transition States with Φ(T). J. Biol. Phys. 2002; 28:115–128. [PubMed: 23345761]

31. Crane JC, Koepf EK, Kelly JW, Gruebele M. Mapping the transition state of the WW domain β-sheet1. J. Mol. Biol. 2000; 298:283–292. [PubMed: 10764597]

32. Tanford C. Protein denaturation. Adv. Protein Chem. 1970; 24:95.

33. Fersht AR, Sato S. Φ-Value analysis and the nature of protein-folding transition states. Proc. Natl. Acad. Sci. U. S. A. 2004; 101:7976–7981. [PubMed: 15150406]

34. Kowalski JA, Liu K, Kelly JW. NMR solution structure of the isolated Apo Pin1 WW domain: Comparison to the x-ray crystal structures of Pin1. Biopolymers. 2002; 63:111–121. [PubMed: 11786999]

35. Ensign DL, Pande VS. The Fip35 WW Domain Folds with Structural and Mechanistic Heterogeneity in Molecular Dynamics Simulations. Biophys. J. 2009; 96:L53–L55. [PubMed: 19383445]

36. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR. Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. Proc. Nat. Acad. Sci. USA. 2009; 106:19011–19016. [PubMed: 19887634]

37. Lane TJ, Bowman GR, Beauchamp K, Voelz VA, Pande VS. Markov state model reveals folding and functional dynamics in ultra-long MD trajectories. J. Am. Chem. Soc. 2011; 133:18413–18419. [PubMed: 21988563]

38. Hammond GS. A correlation of reaction rates. J. Am. Chem. Soc. 1955; 77:334–338.

39. Brady GP, Sharp KA. Entropy in protein folding and in protein—protein interactions. Curr. Opin. Struct. Biol. 1997; 7:215–221. [PubMed: 9094326]

40. Li MS, Klimov DK, Thirumalai D. Finite Size Effects on Thermal Denaturation of Globular Proteins. Phys. Rev. Lett. 2004; 93:268107. [PubMed: 15698029]

41. Maisuradze GG, Zhou R, Liwo A, Xiao Y, Scheraga HA. Effects of Mutation Truncation and Temperature on the Folding Kinetics of a WW Domain. J. Mol. Biol. 2012; 420:350–365. [PubMed: 22560992]

42. a Beccara S, Škrbi T, Covino R, Faccioli P. Dominant folding pathways of a WW domain. Proceedings of the National Academy of Sciences. 2012; 109:2330–2335.

43. Wirth AJ, Liu Y, Prigozhin MB, Schulten K, Gruebele M. Comparing Fast Pressure Jump and Temperature Jump Protein Folding Experiments and Simulations. J. Am. Chem. Soc. 2015; 137:7152–7159. [PubMed: 25988868]

44. Ballew RM, Sabelko J, Reiner C. M. Gruebele. Rev. Sci. Instrum. 1996; 67:3694.

45. Ervin J, Sabelko J, Gruebele M. J. Photochem. Photobiol. 2000; B 54:1.

46. DeLano WL. The PyMOL Molecular Graphics System.

47. Jager M, Zhang Y, Bowman ME, Noel JP, Kelly JW. 2F21: human Pin1 Fip mutant. Worldwide Protein Data Bank. 2006

Highlights

- Folding kinetics of a comprehensive set of hPin1 WW mutants, spanning the whole sequence and multiple substitutions at many sites, has been studied.

- A very conservative phi value analysis, identifying and excluding disruptive mutations, has revealed the interplay between loops 1 and 2 in the transition state in unprecedented detail.

- Unusually large "non-classical" phi values are now explained by local native state disorder.

- This comprehensive experimental data set will be valuable for comparison with molecular dynamics simulation, and we begin by creating a hybrid Phi-value map for FiP WW domain for comparison with recent all-atom simulations.
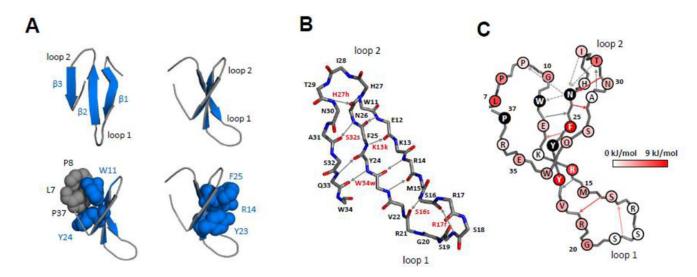
Figure 1. hPin1 WW structure and native state interactions

(**A**) Structural cartoon of the hPin1 WW fold, highlighting the two hydrophobic clusters that protrude from either side of the three-stranded β sheet. The individual β strand are color coded blue, while the loop segments and the N- and C-terminal extensions are shown in grey. Side chain contacts that constitute the hydrophobic clusters are shown as van der Waals surfaces. (**B**) C_α-backbone representation of the three-stranded β sheet region (residues W11-W34), highlighting the ten backbone hydrogen bonds that connect the three β strands and stabilize the 3-stranded β sheet topology. Hydrogen bonds that were perturbed by amine-to-ester mutations for Φ_M analysis are labeled in red. Residues are labeled in single letter code and are numbered. (**C**) Quantitative analysis of a complete Ala-scan, replacing each of the 33 non-Alanine residues individually with Ala. Destabilizations calculated at 55 °C range from near zero to ~ 9 kJ/mole and are mapped onto the C_α-backbone structure of the folded protein. Four Ala-mutants (labeled black) were either completely or significantly unfolded, even at low temperature (4 °C). For these four mutants,   G must exceed 9 kJ/mol, but no accurate thermodynamic data can be derived in aqueous buffer without invoking stabilizing co-solvents.
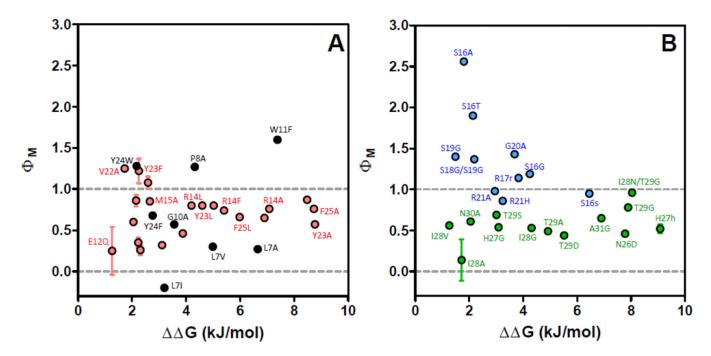
Figure 2. Φ_M-value analysis at 55 °C
(**A**) Plot of the Φ_M value *vs* the difference in free energy between wild type and mutant ($\Delta\Delta$G, in kJ/mol) for β strand (filled red circles) and hydrophobic cluster 1 mutants (filled black circles). (**B**) Plot of the Φ_M value *vs* the difference in free energy between wild type and mutant ($\Delta\Delta$G, in kJ/mol) of loop 1 (filled blue circles) and loop 2 mutants (filled green circles). Errors in Φ_M that exceed the symbol size are shown explicitly. For clarity, individual Φ_M-values are labeled with single letter code. Raw data used to render the plots are provided in Table 1.
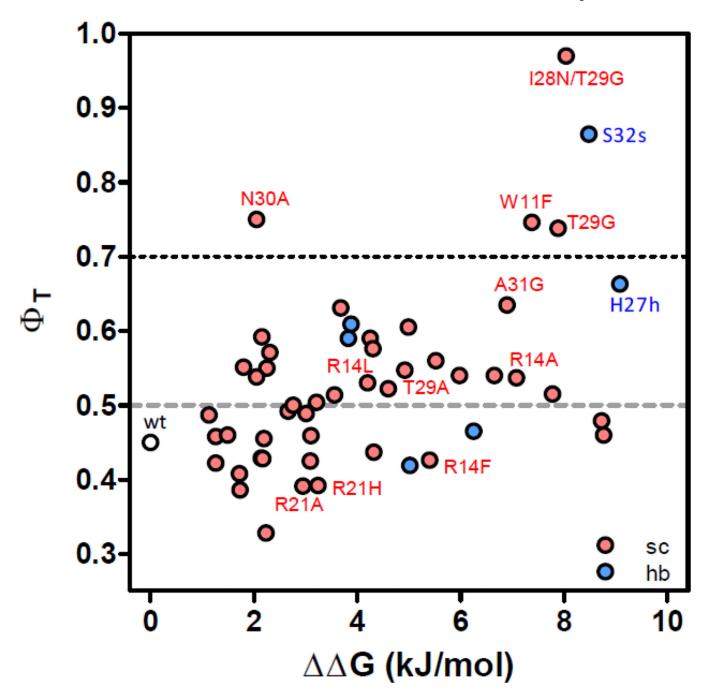
Figure 3. Φ_T analysis at 55 °C

Plot of the Φ_T value for wild type hPin1 WW and mutants thereof *vs* the change in free energy (G, in kJ/mol) between wild type and mutant. Φ_T values are calculated using the T_0-fitting procedure (for details, see SI supporting text 1). Φ_T values of side chain and backbone hydrogen bond mutants are color coded red and blue, respectively. Except the obvious five outliers (mutants W11F, T29G, I28N/T29G, N30A, S32s), the Φ_T values are within a ± 25 % error margin of the average Φ_T (0.50, dashed grey horizontal line). The outlier Φ_T values (> 0.70, dotted grey line) are indicative of perturbing mutations that shift the transition state ensemble along the reaction coordinate closer to the native state.

Mutational Φ_M values calculated from these mutants are no longer reliable indicators of the unperturbed "wild type" transition state ensemble, and must be excluded from the consensus Φ_M analysis of hPin1 WW transition state structure.
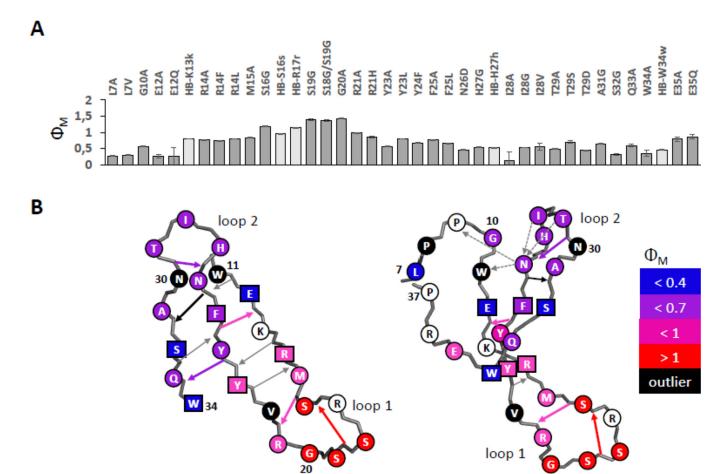
Figure 4. Analysis of the folding transition state of the hPin1 WW domain

(A) Φ_M values of the 34 single and double mutants (dark grey) and the 5 amide-to-ester backbone hydrogen bonds mutants (light grey) that qualify for Φ_M analysis, and that were used for consensus Φ_M mapping of the folding transition state. (B) Φ_M map of the folding transition state, with Φ_M values for 25 of the 34 residues (single letter representation) mapped onto the C-α backbone structure of the N-terminally truncated folded protein (residues 6–39). Left panel: residues W11-W34 that define the 3-stranded β sheet. Right panel: Residues L7-P37 that includes hydrophobic cluster 1 and the N- and C-terminal extensions. For clarity, Φ_M values were grouped and color-coded ($0 < \Phi_M < 0.30$, blue; $0.3 < \Phi_M < 0.6$, purple, $0.6 < \Phi_M < 0.90$, pink; $\Phi_M > 0.90$, red). Residues for which classical hydrophobic deletion mutagenesis yields very high, or negative, Φ_M values that are not supported by other mutations or structural context are color coded black. Residues for which no mutant suitable for Φ_M analysis is available are color coded white. Backbone hydrogen bonds that were studied by amide-to-ester mutagenesis are indicated by arrows (same color code as for side chains). Data used to render the figure are provided in Tables 1 and 2.
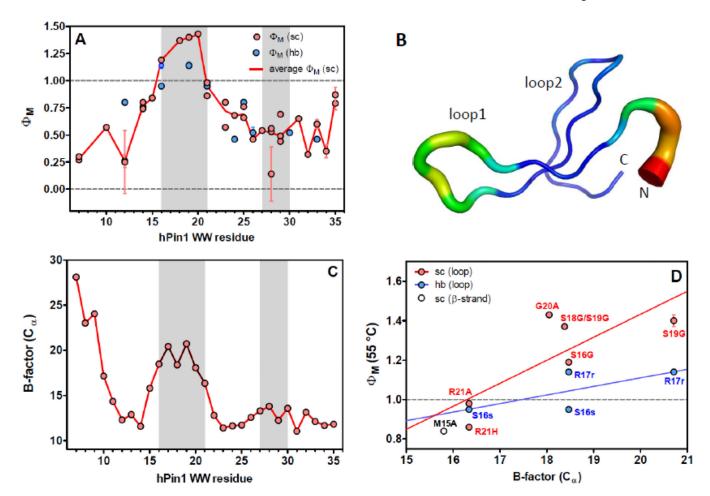
Figure 5. Φ_M *vs* sequence map and Φ_M *vs* backbone disorder correlation
(A) Plot of Φ_M values *vs.* the hPin1 WW sequence used for transition state analysis. Individual side chain Φ_M values are color coded red, while those calculated from backbone hydrogen bond mutants are color-coded blue. The solid red line represents the error-weighted average trend of the side chain Φ_M (see Table 2 for data). **(B)** Tube plot showing the distribution of thermal B factors from the X-ray crystal structure [47] along the backbone of hPin1 WW domain. **(C)** Plot of thermal B factors vs. the hPin1 WW sequence, showing a pronounced maximum in loop 1, and a smaller maximum in loop 2. **(D)** Correlation between Φ_M values and thermal B factors for residues M15-R21 with increased local backbone disorder at 55 °C. Side chain (sc) loop 1 mutants are color coded red and backbone hydrogen bond mutants (hb) are color coded blue. The solid lines represent best fits of the experimental data.
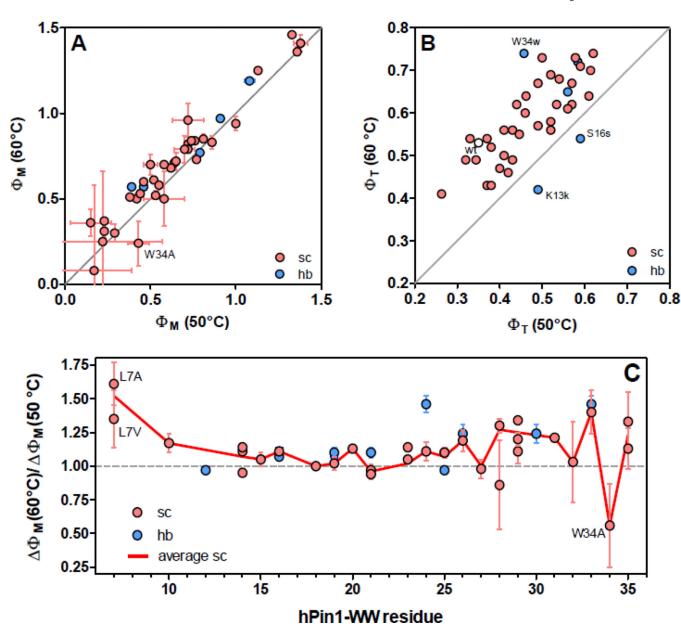
Figure 6. Variation of transition state structure with temperature

(A) Plot of Φ_M (60 °C) *vs* Φ_M (50 °C). On average, Φ_M values increase by 0.07 units when raising the temperature from 50 °C to 60 °C, suggesting that the transition state overall gains native structure upon heating. (B) Plot of Φ_T (60 °C) *vs* Φ_T (50 °C). On average, Φ_T values increase by 0.15 units when raising the temperature from 50 °C to 60 °C, suggesting that the transition state becomes more native-like at elevated temperature, consistent with Hammond's postulate. (C) Plot of the Φ_M (60 °C)/ Φ_M (50 °C)-ratio *vs* the residue number of the hPin1 WW sequence. Data from individual side chain mutants are color coded red. Data from individual backbone hydrogen bond mutants are color coded blue. The solid red line represents the error-weighted average side chain trend. For clarity, the side chain data of E12 (large errors, see Table 2) are not shown.
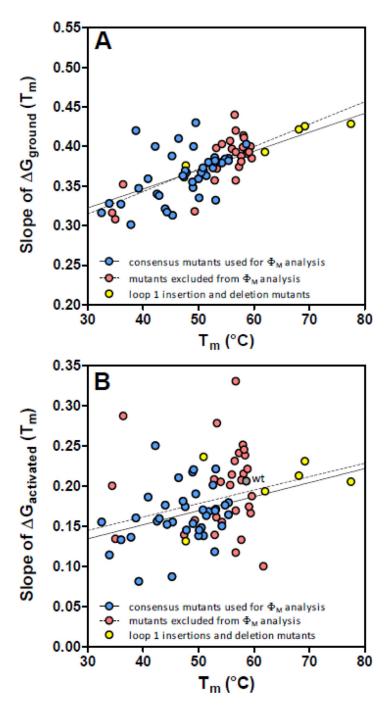
Figure 7. Average number of native contacts in the folding transition state
(A) Slope of the ground state free energy ($\Delta G_{ground}(T)/\Delta T$) of the 39 consensus mutants used for Φ_M analysis (filled red circles, solid black line) or the entire set of single and double mutants (excluding the 6 loop 1 insertion and deletion variants) (filled grey circles, dashed black line) at the midpoint of unfolding (T = T_m, with $\Delta G_{ground}(T_m = 0)$. **(B)** Corresponding plot as in (A) showing the slope of the free energy of activation ($\Delta G_{activated}(T)/\Delta T$ at the midpoint of unfolding (T = T_m). The ratio of the two slopes (activated/ground) of ~ 0.70 for the 39 consensus mutants (0.63 for the entire mutant set) suggests that about 70 % of the

native contacts are developed in the folding transition state, a value that agrees well with the average calculated from the Φ_M data (Table 2), but that is higher than the average Φ_T value (0.50). The loop 1 insertion and deletion variants that change local changes in backbone topology (filled yellow circles) were excluded from the fit, but their values agree well with the extrapolated fits of the mutants with the 6-residue wild type hPin1 WW loop 1.
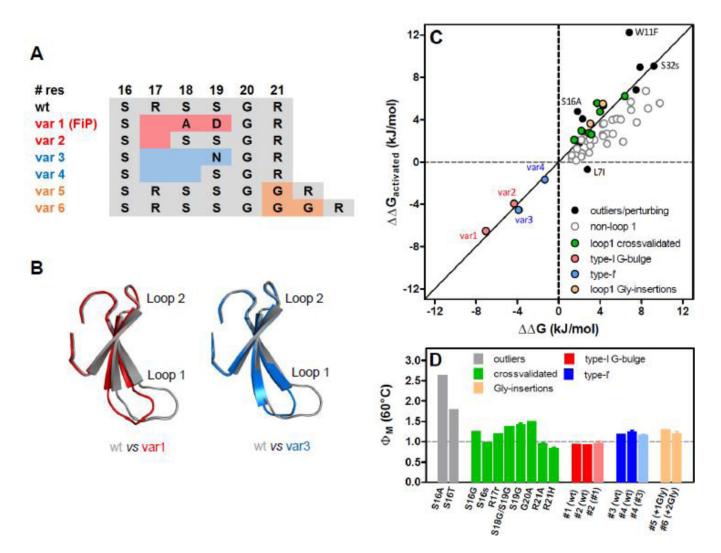
Figure 8. Φ_M analysis of hPin1 WW variants with loop 1 deletions or insertions mutations
(A) Loop 1 sequences of the hPin1 WW loop 1 deletions or insertions variants. Wild type residues are numbered and color coded grey. Mutated or deleted residues in the loop deletion variants are color coded red (type-I G-bulge turn) and blue (type-I' turn), while the inserted Gly residues in the loop 1 insertion mutants are highlighted in orange. **(B)** Superposition of the high resolution X-ray structures of type-I G-bulge variant FiP (1.90 Å resolution, color coded red, left) and the type-I' variant 3 (1.50 Å resolution, color coded blue, right) with wild type hPin1 WW structure (1.35 Å resolution, color coded grey). **(C)** Brønsted plot for folding of the loop 1 variants of hPin1 WW at 60 °C, rendered from the data provided in SI Table 2. Filled red circles: 5-residue type-I G-bulge turn mutants (var1, var2). Filled blue circles: 4-residue type-I' turn variants (var3, var4). Filled green circles: Cross-validated loop 1 side chain and backbone hydrogen bond mutants (6-residue wild type loop 1 context). Filled orange circles: Gly insertion variants (var5, var6). Filled black circles: Outlier/perturbing mutants. Open light grey circles: Non-loop 1 consensus mutants. The solid black line is the line predicted for $\Phi_M = 1$. **(D)** Bar plot of Φ_M-values for selected mutants shown in (C). Φ_M values calculated for the redesigned loop 1 variants using wild type hPin1 WW as

reference are color coded red (5-residue type-I G-bulge variants) and blue (4-residue type-I' variants). Φ_M values calculated for variants 2 and 4 in the type-I G-bulge (var1, FiP) and type-I' context (var3) are shown in light red and light blue, respectively.
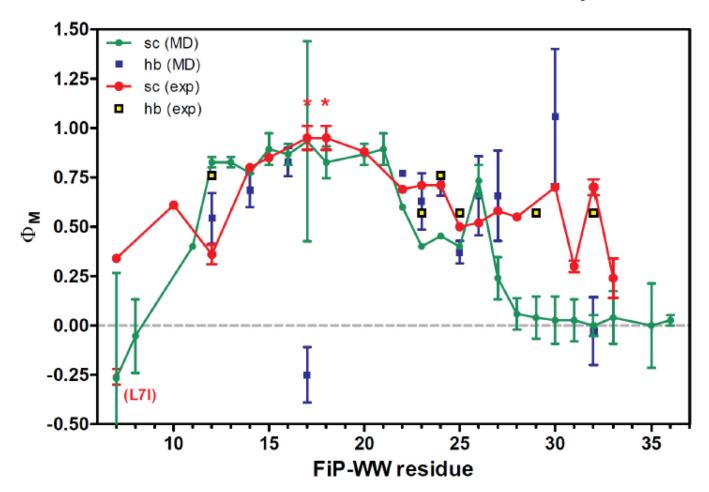
Figure 9. Hypothetical "hybrid" Φ_M map for the fast-folding FiP variant of hPin1 WW
Hypothetical side chain Φ_M map (red circles and solid red line) for the fast folding FiP variant of hPin1 WW, rendered with side chain Φ_M values of non-loop 1 mutants measured with wild type hPin1 WW as reference (see Fig. 3, SI Table 2 for details) and the side chain Φ_M value for loop 1 FiP WW variant 2 (loop 1 sequence: SSSGR) measured with FiP as "pseudo wild type" reference (loop 1 sequence: SADGR). As two residues were replaced simultaneously in FiP variant 2 (A18S, D19S, see Fig. 8A), the Φ_M value calculated for variant 2 ($\Phi_M = 0.94 \pm 0.05$) was assigned to either mutated residue (labeled by asterisks) in FiP. For residues that are probed by multiple side chain mutations, the error-weighted average Φ_M value is shown (see SI Table 2 for details). Experimentally measured backbone hydrogen bond Φ_M values (filled yellow squares) are those measured for wild type hPin1 WW and are assigned to the two residues that engage in the perturbed hydrogen bond (see SI Table 2 for details). The simulated side chain and backbone hydrogen bond Φ_M values and associated errors are shown in green and blue, respectively and were rendered from Fig. 2E in [14]. Residue numbers correspond to the 33-residue FiP sequence and thus account for the shorter loop 1 substructure (deletion of Arg17 of wild type hPin1 WW).

Table 1

Thermodynamic and kinetic data for wild type hPin1 WW and mutants thereof

Variant	T_m (°C)	$G_f^{(1)}$	$G_f^{(2)}$	$G^{\dagger(0)}$	$G^{\dagger(1)}$	$G^{\dagger(2)}$	Φ_M (50 °C)[1]	Φ_M (55 °C)[1]	Φ_M (60 °C)[1]	Ref.
1. Wildtype and single-site mutants										
wt hPin1	58.6	0.403	0.00272	14.92	0.206	0.00472	-	-	-	[10]
K6A	59.4	0.400	0.00153	11.16	0.166	0.00173	-	-	-	[10]
K6M	58.1	0.414	0.00180	11.76	0.215	0.00162	-	-	-	N.3
L7A	37.8	0.301	0.00022	13.16	0.136	0.00192	0.23 (0.02)	0.27 (0.02)	0.31 (0.03)	[6, 10]
L7I	49.3	0.318	0.00050	12.66	0.157	0.00141	-0.21 (0.04)	-0.20 (0.04)	-0.26 (0.04)	[10]
L7V	44.0	0.321	0.00041	13.56	0.176	0.00218	0.23 (0.02)	0.30 (0.02)	0.37 (0.02)	[10]
P8A	47.4	0.361	0.00293	18.56	0.139	0.00237	1.29 (0.01)	1.27 (0.01)	1.23 (0.01)	[10]
P9A	56.0	0.397	0.00229	19.10	0.214	0.00272	-	-	-	[10]
G10A	49.0	0.348	0.00151	15.23	0.153	0.00341	0.52 (0.02)	0.57 (0.02)	0.61 (0.02)	[10]
W11F	3.05	0.308	-0.00050	21.62	0.134	0.00399	1.42 (0.01)	1.58 (0.01)	1.79 (0.01)	[10]
E12A	52.6	0.373	0.00104	14.33	0.201	0.00396	0.15 (0.12)	0.26 (0.06)	0.36 (0.05)	[10]
E12Q	55.4	0.385	0.00308	14.62	0.179	0.00421	0.22 (0.35)	0.25 (0.30)	0.25 (0.29)	[6, 10]
K13A	59.6	0.385	0.00285	16.11	0.187	0.00139	-	-	-	[10]
K13V	62.8	0.401	0.00322	15.85	0.215	0.00213	-	-	-	N
K13Y	51.7	0.349	0.00237	16.63	0.125	0.00120	1.09 (0.07)	1.09 (0.07)	1.01 (0.08)	N
R14A	39.2	0.347	0.00074	17.21	0.081	0.00464	0.72 (0.01)	0.76 (0.01)	0.82 (0.01)	[10]
R14F	45.2	0.388	0.00195	16.87	0.087	0.00517	0.76 (0.01)	0.74 (0.01)	0.73 (0.01)	[6]
R14L	47.8	0.367	0.00234	16.31	0.145	0.00482	0.77 (0.01)	0.80 (0.01)	0.84 (0.01)	N
M15A	51.8	0.380	0.00289	15.88	0.168	0.00434	0.81 (0.02)	0.84 (0.02)	0.85 (0.02)	[6, 10]
S16A	54.0	0.380	0.00313	18.63	0.205	0.00372	2.44 (0.03)	2.56 (0.02)	2.62 (0.02)	[10]
S16G	47.6	0.369	0.00194	17.75	0.174	0.00452	1.13 (0.01)	1.19 (0.01)	1.25 (0.01)	[10]
S16T	53.2	0.398	0.00325	18.01	0.161	0.00401	1.99 (0.02)	1.90 (0.02)	1.78 (0.01)	[6]
R17A	58.8	0.391	0.00232	19.23	0.221	0.00276	-	-	-	[10]
R17G	57.3	0.374	0.00277	18.76	0.241	0.00301	-	-	-	[10]
S18A	58.4	0.398	0.00185	22.34	0.238	0.00614	-	-	-	[10]
S18G	56.5	0.440	0.00227	16.49	0.231	0.00670	-	-	-	[10]
S19G	54.8	0.384	0.00248	16.29	0.176	0.00432	1.38 (0.04)	1.40 (0.01)	1.41 (0.04)	[6, 10]

Variant	T_m (°C)	$G_f^{(1)}$	$G_f^{(2)}$	$G^{†(0)}$	$G^{†(1)}$	$G^{†(2)}$	Φ_M (50 °C)[l]	Φ_M (55 °C)[l]	Φ_M (60 °C)[l]	Ref.
G20A	48.9	0.355	0.00270	18.11	0.217	0.00216	1.33 (0.01)	1.43 (0.01)	1.50 (0.01)	[10]
R21A	50.9	0.369	0.00144	16.54	0.138	0.00181	1.00 (0.02)	0.98 (0.02)	0.94 (0.02)	[10]
R21H	50.0	0.359	0.00130	16.31	0.138	0.00127	0.86 (0.02)	0.86 (0.02)	0.83 (0.02)	N
R21L[4]	55.9	0.521	-0.00010	15.63	0.217	0.00111	-	-	-	N
V22A	54.2	0.403	0.00116	16.29	0.155	0.00146	1.36 (0.05)	1.25 (0.04)	1.12 (0.04)	[6, 10]
Y23A	33.9	0.328	0.00098	15.99	0.114	0.00193	0.55 (0.01)	0.57 (0.01)	0.58 (0.01)	[10]
Y23F	52.8	0.376	0.00254	16.54	0.208	0.00141	1.11 (0.02)	1.23 (0.02)	1.27 (0.02)	[10]
Y23L	45.3	0.313	0.00153	16.24	0.155	0.00159	0.74 (0.01)	0.80 (0.01)	0.84 (0.01)	[6, 10]
Y24F	51.4	0.363	0.00279	15.49	0.163	0.00392	0.64 (0.02)	0.68 (0.02)	0.71 (0.02)	[10]
Y24W	52.9	0.357	0.00230	16.72	0.139	0.00436	1.27 (0.02)	1.28 (0.02)	1.30 (0.02)	[10]
F25A	32.5	0.316	0.00042	16.92	0.155	0.00098	0.72 (0.01)	0.76 (0.02)	0.79 (0.02)	[10]
F25L	42.5	0.340	0.00202	15.85	0.156	0.00239	0.62 (0.01)	0.66 (0.01)	0.68 (0.01)	[6, 10]
N26D	36.0	0.327	0.00044	14.56	0.133	0.00211	0.42 (0.01)	0.46 (0.02)	0.50 (0.03)	[6, 10]
H27A	57.7	0.388	0.00262	14.76	0.207	0.00245	-	-	-	[10]
H27G	50.5	0.367	0.00130	15.20	0.148	0.00197	0.53 (0.02)	0.54 (0.02)	0.52 (0.02)	[10]
I28A	54.2	0.379	0.00165	14.35	0.150	0.00404	0.17 (0.22)	0.14 (0.25)	0.08 (0.44)	[6, 10]
I28G	47.2	0.363	0.00105	14.93	0.181	0.00326	0.46 (0.01)	0.53 (0.01)	0.60 (0.01)	[10]
I28V	55.4	0.382	0.00328	15.01	0.164	0.00413	0.58 (0.12)	0.56 (0.10)	0.50 (0.12)	[10]
T29A	44.3	0.317	0.00100	14.80	0.152	0.00205	0.44 (0.01)	0.49 (0.01)	0.53 (0.01)	[10]
T29D	42.9	0.338	0.00009	14.38	0.159	0.00262	0.38 (0.01)	0.44 (0.01)	0.51 (0.01)	[6]
T29G	34.4	0.316	0.00001	15.32	0.200	0.00243	0.68 (0.01)	0.79 (0.01)	0.91 (0.02)	[10]
T29S	50.8	0.373	0.00159	15.57	0.170	0.00278	0.65 (0.03)	0.70 (0.03)	0.72 (0.04)	[10]
N30A	53.3	0.372	0.00208	15.02	0.278	0.00302	0.31 0.07)	0.61 (0.03)	0.89 (0.03)	[10]
A31G	40.9	0.359	0.00197	15.45	0.186	0.00311	0.58 (0.01)	0.65 (0.01)	0.70 (0.01)	[6, 10]
A31S	57.7	0.381	0.00283	15.76	0.133	0.00373	-	-	-	[10]
S32G	50.1	0.335	0.00200	14.46	0.145	0.00198	0.29 (0.03)	0.32 (0.03)	0.30 (0.04)	[10]
S32T	61.7	0.398	0.00356	14.70	0.100	0.00240	-	-	-	[6]
Q33A	53.1	0.332	0.00103	15.13	0.171	0.00326	0.50 (0.04)	0.60 (0.04)	0.70 (0.04)	N
W34A	52.9	0.386	0.00067	14.75	0.118	0.00295	0.43 (0.06)	0.35 (0.06)	0.24 (0.10)	[6, 10]
W34F	58.0	0.399	0.00326	15.81	0.251	0.00212	-	-	-	[10]
E35Q	53.1	0.380	0.00280	15.67	0.221	0.00265	0.72 (0.09)	0.87 (0.06)	0.96 (0.06)	[10]

Variant	T_m (°C)	$G_f^{(1)}$	$G_f^{(2)}$	$G^{‡(0)}$	$G^{‡(1)}$	$G^{‡(2)}$	Φ_M (50 °C)[1]	Φ_M (55 °C)[1]	Φ_M (60 °C)[1]	Ref.
E35A	50.3	0.369	0.00283	16.13	0.154	0.00203	0.82 (0.07)	0.83 (0.07)	0.79 (0.06)	[10]
R36A	56.7	0.357	0.00225	16.44	0.117	0.00231	-	-	-	[10]
S38A	59.1	0.393	0.00204	17.13	0.174	0.00327	-	-	-	[10]
S38G	58.2	0.411	0.00382	18.43	0.245	0.00295	-	-	-	[10]
S38T	58.2	0.390	0.00327	18.22	0.232	0.00337	-	-	-	N
2. Double-site mutants										
S18G/S19G	53.0	0.382	0.00163	16.88	0.169	0.00246	1.36 (0.02)	1.37 (0.02)	1.36 (0.02)	N
S19G/G20S	56.7	0.393	0.00288	16.88	0.169	0.00246	-	-	-	N
I28N/T29G	36.4	0.352	0.00024	15.25	0.287	0.00387	0.79 (0.01)	0.96 (0.01)	1.14 (0.01)	N
3. Loop1 insertion and deletion mutants[2]										
var1 (FiP)	77.5	0.428	0.00327	10.65	0.2052	0.00532	-	-	0.92 (0.01)	[7]
var2	69.2	0.425	0.00191	13.01	0.2305	0.00457	-	0.84 (0.01)	0.91 (0.01)	[7]
var3	68.1	0.422	0.00220	12.07	0.2126	0.00498	-	-	1.18 (0.01)	[7]
var4	62.0	0.393	0.00228	13.92	0.1931	0.00216	-	1.28 (0.07)	1.24 (0.04)	[7]
var5 (+1G)	47.7	0.396	0.00139	18.73	0.1310	0.00256	1.34 (0.01)	1.32 (0.01)	1.32 (0.01)	N
var6 (+2G)	50.9	0.366	0.00347	16.47	0.2360	0.00281	0.94 (0.01)	1.09 (0.01)	1.09 (0.01)	N
4. Backbone hydrogen bond amide-to-ester mutants										
K13k	46.4	0.410	0.0010	16.52	0.21	0.00100	0.79 (0.01)	0.80 (0.01)	0.77 (0.01)	[16]
S16s	42.2	0.400	-0.0005	17.37	0.25	0.00120	0.91 (0.01)	0.95 (0.01)	0.97 (0.01)	[16]
R17r	49.1	0.400	0.0016	17.20	0.22	0.00300	1.08 (0.03)	1.14 (0.03)	1.19 (0.03)	[16]
V22v	56.7	0.420	0.0034	16.64	0.33	0.00340	-	-	-	[16]
H27h	38.7	0.420	0.0031	14.83	0.16	0.00560	0.46 (0.01)	0.52 (0.01)	0.57 (0.01)	[16]
S32s	41.5	0.510	0.0010	14.70	0.50	0.00090	0.72 (0.01)	0.87 (0.01)	0.98 (0.01)	[16]
W34w	49.5	0.430	0.0032	14.74	0.19	0.00840	0.39 (0.03)	0.46 (0.02)	0.57 (0.01)	[16]

[1]Mutants that differ < 1 kJ/mole in stability from wild type hPin1 WW resulted in large errors in Φ_M, so no Φ_M-values are listed. Φ_M-value were also not calculated at 50 and/or 55 °C for the more stable loop 1 deletion mutants with thermodynamically optimized loop 1 substructures, to avoid errors in Φ_M due to extrapolation of the data. Rounded errors in Φ_M of all other mutants are given in brackets.

[2]Var1: Type-I G-bulge turn, sequence: SADGR. Var2: Type-I G-bulge turn, sequence: SSSGR. Var3: Type-I′ turn, sequence: SNGR. Var4: Type-I′ turn, sequence: SSGR. Var5: Single Gly insertion, sequence: SRSSGGR. Var6: Double Gly insertion, sequence: SRSSGGGR.

[3]N= new mutant.

[4]Mutant R21L forms a dimer at protein concentrations employed for T-jump relaxation (10–30 μM) and was thus excluded from Φ_M analysis.

Table 2

Summary of Φ_M values of consensus mutants used for transition state mapping at 55 °C

Residue	Mutation	Type [1]	G (kJ/mol)	Φ_M (55 °C)	Average Φ_M (sc)	Average Φ_M (hb)
L7	L7A	sc	6.65	0.27 (0.02)	0.28	-
	L7V	sc	5.00	0.30 (0.02)		
G10	G10A	sc	3.56	0.57 (0.02)	0.57	-
E12	E12A	sc	2.31	0.26 (0.06)	0.26	0.80
	E12Q	sc	1.26	0.25 (0.29)		
	K13k	hb	5.01	0.80 (0.01)		
R14	R14A	sc	7.08	0.76 (0.01)	0.77	-
	R14F	sc	5.41	0.74 (0.01)		
	R14L	sc	4.18	0.80 (0.01)		
M15	M15A	sc	2.66	0.84 (0.02)	0.84	-
S16	S16G	sc	4.25	1.19 (0.01)	1.19	1.01
	S16s	hb	6.45	0.95 (0.01)		
	R17r	hb	3.38	1.14 (0.02)		
S18 [3]	S18G/S19G	sc	2.19	1.37 (0.02)	1.37	-
S19	S19G	sc	1.49	1.40 (0.03)	1.40	1.19
G20	G20A	sc	3.68	1.43 (0.01)	1.42	-
R21	R21A	sc	2.95	0.98 (0.02)	0.92	0.95
	R21H	sc	3.24	0.86 (0.02)		
Y23	Y23A	sc	8.77	0.57 (0.01)	0.72	-
	Y23L	sc	4.60	0.80 (0.01)		
Y24	Y24F	sc	2.76	0.68 (0.02)	0.68	0.46
F25	F25A	sc	8.73	0.76 (0.02)	0.69	0.80
	F25L	sc	5.98	0.66 (0.01)		
N26	N26D	sc	7.79	0.46 (0.02)	0.46	0.52
	H27h	hb	9.08	0.52 (0.01)		
H27	H27G	sc	3.09	0.54 (0.02)	0.54	-
I28	I28A	sc	1.72	0.14 (0.25)	0.52	-
	I28V	sc	1.26	0.56 (0.10)		

Residue	Mutation	Type [1]	G (kJ/mol)	Φ_M (55 °C)	Average Φ_M (sc)	Average Φ_M (hb)
	I28G	sc	4.31	0.53 (0.01)		
T29	T29A	sc	4.92	0.49 (0.01)	0.49	-
	T29S	sc	3.01	0.70 (0.04)		
	T29D	sc	5.52	0.44 (0.01)		
N30	H27h	hb	9.08	0.52 (0.01)	-	0.52
A31	A31G	sc	6.87	0.65 (0.01)	0.65	-
S32	S32G	sc	3.10	0.32 (0.03)	0.32	-
Q33	Q33A	sc	2.05	0.60 (0.04)	0.60	0.46
	W34w	hb	3.87	0.46 (0.01)		
W34	W34A	sc	2.23	0.35 (0.10)	0.35	-
E35	E35A	sc	3.27	0.83 (0.06)	0.85	-
	E35Q	sc	2.14	0.87 (0.07)		

[1]Type of mutation: side chain (sc), backbone hydrogen bond (hb).

[2]Error weighted average Φ_M-value for residues probed my multiple mutations.

[3]ϕ_M-value of the S18G/S19G was assigned to S18.