BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2011,** *50,* 125–135 125
DOI: 10.1021/bi1016793

Structural Insights into the Two Sequential Folding Transition States of the PB1 Domain of NBR1 from Φ Value Analysis and Biased Molecular Dynamics Simulations†

Ping Chen,[‡,∥] Clare-Louise Evans,[§] Jonathan D. Hirst,[§] and Mark S. Searle*[,‡,§]

[‡]*Centre for Biomolecular Sciences, School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, U.K., and* [§]*School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.* [∥]*Current address: Institute of Biophysics, Chinese Academy of Science, 15 Datun Rd., Chaoyang District, Beijing 100101, China.*

Received October 19, 2010; Revised Manuscript Received November 30, 2010

ABSTRACT: The PB1 domain of NBR1 folds via a single pathway mechanism involving two sequential energy barriers separated by a high-energy intermediate. The structural ensemble representing each of the two transition states (TS1 and TS2) has been calculated using experimental Φ values and biased molecular dynamics simulations. Both TS1 and TS2 represent compact states (β_{TS1} = 0.71, and β_{TS2} = 0.93) but are defined by quite different distributions of Φ values, degrees of structural heterogeneity, and nativelike secondary structure. TS1 forms a heterogeneous ensemble of dynamic structures, representing a global collapse of the polypeptide chain around a set of weak nativelike contacts. In contrast, TS2 has a high proportion of nativelike secondary structure, which is reflected in an extensive distribution of high Φ values. Two snapshots along the folding pathway of the PB1 domain reveal insights into the malleability, the solvent accessibility, and the timing of nativelike core packing that stabilizes the folded state.

The folding pathways of many small proteins have been described by one of two distinct classical mechanisms defined by the nature of the interactions identified in the transition state for folding (*1*−*7*). However, a unifying view of the protein folding landscape has more recently emerged in which the classical nucleation−condensation and diffusion−collision models are viewed as different manifestations of folding on the same rough energy landscape (*8*, *9*). This hypothesis would appear to rationalize the observation that structurally highly homologous members of the same family can fold through either mechanism (*8*).

The stochastic nature of protein folding suggests that it is a highly heterogeneous process involving many different initial trajectories that allow for the possibility of parallel folding pathways with structurally unique transition states possessing different degrees of nativelike bias. Nonlinear effects in the unfolding kinetics, manifested as curvature in chevron plots, attest to transition state movements under different experimental conditions (*10*−*12*). This movement has variously been interpreted as a change in the flux of molecules along competing pathways each with a structurally distinct transition state, as a switch between discrete sequential transition state structures on a single folding pathway, or as Hammond-like shifts of the transition state across a broad energy barrier representing a diffuse ensemble of structures (*10*−*19*). Implicit in the parallel and sequential pathway models is the notion that molecules are funnelled through a narrowly defined ensemble of states representing discrete transition states and obligatory (often undetected metastable) folding intermediates that play a key role in channeling the polypeptide chain toward the native structure.

The 90-residue N-terminal Phox and Bem1p (PB1)[1] domain of NBR1 (Next-to-Breast Cancer Gene 1), which has a ubiquitin-like structural topology (*20*, *21*), folds with nonlinear effects evident in the denaturant dependence of the unfolding rate. The kinetics of wild-type (wt) PB1 and several mutants showed that this curvature is consistent with a single pathway mechanism involving movement of the transition state between two sequential energy barriers (TS1 and TS2) separated by a transiently populated high-energy intermediate (Figure 1) (*12*). We have now characterized the two transition states at high resolution using Φ value analysis, in combination with biased molecular dynamics (BMD) simulations (*8*, *10*, *11*, *22*), to generate an ensemble of structures. Both TS1 and TS2 represent compact states (β_{TS1} = 0.71, and β_{TS2} = 0.93) but are defined by quite different distributions of Φ values, degrees of structural heterogeneity, and nativelike secondary structure. The two transition states are correlated by sharing a subset of the same nativelike interactions but reflect different expanded forms of the native state. The observation of sequential transition states provides two snapshots along the hierarchical folding pathway of the PB1 domain and reveals insights into the malleability, the solvent accessibility, and the timing of nativelike core packing that stabilizes the folded state.

MATERIALS AND METHODS

Mutagenesis and Protein Expression. The wild-type NBR1-PB1 gene corresponding to residues 1−100 was cloned between the BamHI and XhoI sites of the pGEX 4T-1 plasmid (Amersham Biosciences). All mutations were introduced by site-directed mutagenesis (QuikChange Mutagenesis Kit, Stratagene) and the constructs confirmed by DNA sequencing. We expressed

†P.C. is grateful for an Overseas Research Scholarship (ORS) and for funding from the School of Chemistry at the University of Nottingham. C.-L.E. acknowledges support from the BBSRC.

*To whom correspondence should be addressed. E-mail: mark.searle@nottingham.ac.uk. Telephone: (44) 115 951 3567. Fax: (44)115846 6059.

[1]Abbreviations: PB1, Phox and Bem1p; NBR1, Next-to-Breast Cancer Gene 1; TS, transition state; rmsd, root-mean-square deviation; BMD, biased molecular dynamics.

Published on Web 12/01/2010

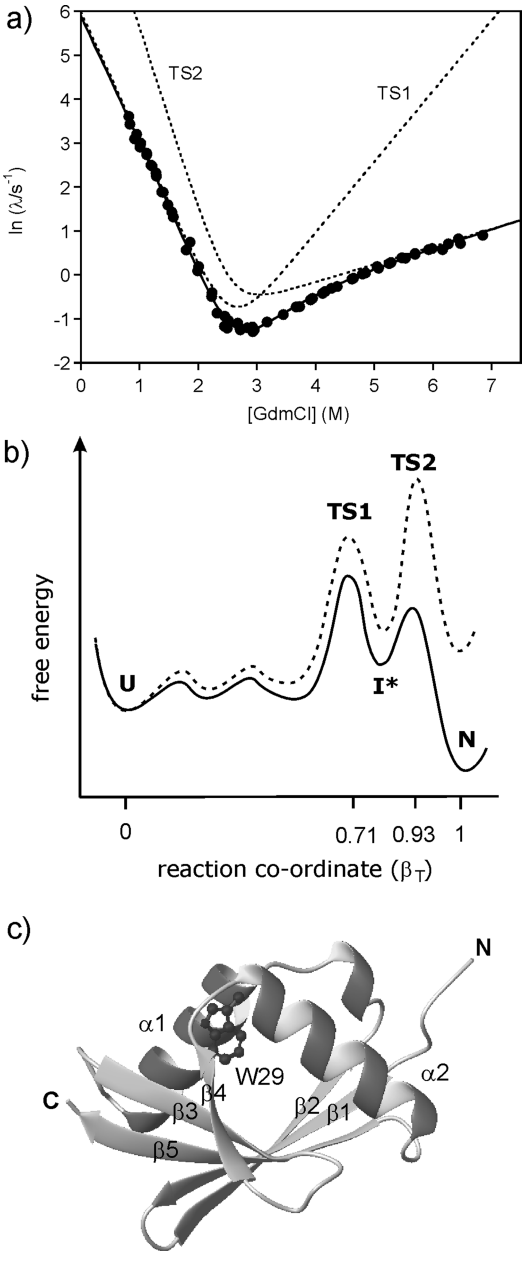

FIGURE 1: Kinetic analysis of the folding pathway of the PB1 domain of NBR1 at 298 K. (a) Curved chevron plot showing the fit to a sequential transition state model resulting in curvature in the refolding arm. The dotted lines represent the linear chevron analysis under conditions in which either TS1 or TS2 is rate-limiting. (b) Representation of the shift from the TS1-limited (—) to TS2-limited (···) kinetics under strongly renaturing and strongly denaturing conditions, respectively. (c) Structure of the PB1 domain [Protein Data Bank entry 2BKF (*20*)] showing the position of W29 with labeling of secondary structure elements.

proteins by transforming *Escherichia coli* strain BL21(DE3), growing the cells at 30 °C to an OD_{595} of 0.7, and inducing the cells with IPTG at a final concentration of 300 μM for 20 h. Proteins were overexpressed, isolated, and purified as previously described (*12*).

Kinetic Experiments. Kinetic measurements were performed using an Applied Photophysics Pi-Star-180 spectrophotometer and stopped-flow apparatus, with a dead time of ~2 ms. The temperature at the observation cell was regulated using a Neslab RTE-300 circulating programmable water bath. The fluorescence was excited at 280 nm using an Hg−Xe lamp and the emitted fluorescence monitored above 320 nm. Unfolding and refolding kinetics were initiated by 11-fold dilution into various GdmCl concentrations of the native or unfolded protein at initial concentrations between 5 and 15 μM. The change in fluorescence intensity over time was measured until no significant change was observed (typically <10 s^{-1}). All experiments were performed in 25 mM sodium acetate buffer (pH 5.0), and five traces were averaged for each experiment (*23*, *24*).

Analysis of Kinetic Data. The kinetic data were analyzed using a multiexponential fitting procedures (two refolding phases and one unfolding phase were observed). Chevron plots were constructed by plotting ln k_{obs} versus denaturant concentration [D]. The curvature in the unfolding data was interpreted in terms of two discrete sequential transition states (TS1 and TS2) separated by a high-energy intermediate (I*), with TS1 and TS2 becoming rate-limiting under different experimental conditions (see Figure 1) (*12*, *14*, *15*). At a low denaturant concentration, TS1 is rate-limiting and the rate of folding k_f is approximated by the relationship $k_f^{TS1} = k_{UI*}$ and unfolding can be considered in terms of a pre-equilibrium between N and I* such that $k_u^{TS1} = (k_{NI*}/k_{I*N})k_{I*U}$.

Similarly, under the TS2 rate-limiting regime, the rates of folding and unfolding are expressed as follows: $k_f^{TS2} = (k_{UI*}/k_{I*U})k_{I*N}$, and $k_u^{TS2} = k_{NI*}$. The curvature in the unfolding arm of the chevron is described by a switch from a TS1 to a TS2 rate-limited folding mechanism. Because I* is not significantly populated, its stability can not be determined directly. It remains higher in energy than either U or N with the k_{I*U}/k_{I*N} ratio representing a measure of the difference in free energy between the two transition states, where

$$\Delta\Delta G_{TS1-TS2} = -RT \ln(k_{I*N}/k_{I*U}) \quad (1)$$

The analytical solution to the three-state model allows the dependence of ln k_{obs} on [D] to be analyzed according to

$$k_{obs} = [-B \pm (B^2 - 4C)^{1/2}]/2 \quad (2)$$

$$B = -(k_{UI*} + k_{I*U} + k_{I*N} + k_{NI*}) \quad (3)$$

$$C = k_{UI*}(k_{I*N} + k_{NI*}) + k_{I*U}k_{NI*} \quad (4)$$

The microscopic rate constants can be determined, along with their denaturant dependence (m values), by fitting ln k_{obs} versus [D] (*25*). To facilitate fitting by minimizing the number of iterated variables and to ensure that I* remains higher in energy than U or N under all conditions, we fixed k_{I*U} and m_{I*U}, allowing all other parameters to be fitted iteratively. Experimenting with a wide variation in k_{I*U} and m_{I*U} values produced little effect on other variables provided that I* remains metastable. Thus, k_{I*U} and m_{I*U} values were arbitrarily fixed at 10^4 s^{-1} and 0 kJ mol^{-1} M^{-1}, respectively (*14*).

The positions of the two transition states, TS1 and TS2, along the reaction coordinate were described by the Tanford β values β_{TS1} and β_{TS2}, respectively, which are given by the expressions

$$\beta_{TS1} = m_{UI*}/(m_{UI*} + m_{I*U} + m_{I*N} + m_{NI*}) \quad (5)$$

$$\beta_{TS2} = 1 - m_{NI*}/(m_{UI*} + m_{I*U} + m_{I*N} + m_{NI*}) \quad (6)$$

Φ values were calculated from the relative change in activation free energy in the transition state ($\Delta\Delta G_{TS}$) and the native state

($\Delta\Delta G_{UN}$) upon mutation according to the following expressions:

$$\Phi = \Delta\Delta G_{TS}/\Delta\Delta G_{UN} \quad (7)$$

$$\Delta\Delta G_{TS} = -RT \ln(k_f^{wt}/k_f^{mut}) \quad (8)$$

$$\Delta\Delta G_{UN} = -RT[\ln(k_f^{wt}/k_u^{wt}) - \ln(k_f^{mut}/k_u^{mut})] \quad (9)$$

In considering Φ value analysis for TS1, $k_f^{TS1} = k_{UI*}$ and $k_u^{TS1} = (k_{NI*}/k_{I*N})k_{I*U}$, with values determined for both wt-PB1 and the various mutants. Similarly, Φ value analysis for TS2 uses the relationships $k_f^{TS2} = (k_{UI*}/k_{I*U})k_{I*N}$ and $k_u^{TS2} = k_{NI*}$ from data derived for wt-PB1 and mutants (see Table S1 of the Supporting Information).

The data for wt-PB1 and the various mutants were analyzed globally using IGOR (Wavemetrics) by simultaneously fitting the observed refolding and unfolding rate constants, normalized amplitudes, and initial and final fluorescence signal to determine the microscopic rate constants while using shared *m* values (denaturant dependencies of the rate constants). A number of independent experiments for wt-PB1 gave consistent kinetic parameters indicating that standard curve fitting errors are similar to standard deviations determined from repeat measurements. The fitting errors were propagated to the determination of $\Delta\Delta G_{UN}$, $\Delta\Delta G_{TS1}$, $\Delta\Delta G_{TS2}$, and Φ values. In a number of cases, stability changes for a mutation were too small ($\Delta\Delta G_{UN} \leq 2$ kJ/mol) to determine accurate Φ values. To identify potentially anomalous data, we also considered the variation in Φ values for clusters of residues close together in sequence and structure. In this context, there appeared to be consistency between Φ values determined from very different stability changes (see Table 1 and Figure 5).

MD Simulations of the Native State. The X-ray crystal structure [Protein Data Bank (PDB) entry 2BKF] of the PB1 domain of NBR1 was selected for MD simulations (*20*). Hydrogen atoms were added, and the structure was validated using Molprobity before it was used in any MD simulations (*26*). MD simulations were performed using the CHARMM 19 force field with version EEF1.1 (*27*) of the united-atom topology and parameter files with an excluded volume implicit solvation model used within CHARMM version c34b1 (*28*). The termini and titratable side chain protonation states were chosen so that they were consistent with those at neutral pH. To generate an initial degree of sampling from the X-ray crystal structure, five replicas were generated and the MD simulations initiated using random seeds for each replica.

The replica systems were first minimized using steepest descent for 200 steps to remove any severe contacts and then using the adopted basis Newton−Raphson (ABNR) algorithm for a total of 10000 steps, with a force constant of 200 kcal mol^{-1} Å^{-2} for harmonic restraints applied to the backbone atoms that was decreased by 5 kcal mol^{-1} Å^{-2} every 250 steps. Once minimized, the system was heated to 300 K using the Berendsen algorithm (*29*), where the temperature was increased 30 K every 20 ps with harmonic restraints were again applied to the backbone atoms (force constant of 50 kcal mol^{-1} Å^{-2}). During heating, the harmonic restraints on the backbone were gradually reduced and then removed once the target temperature of 300 K was reached. Each system was allowed to equilibrate at a constant temperature of 300 K for a further 200 ps. Unconstrained dynamics were continued for an additional 1 ns to acquire the production dynamics using a time step of 2 fs.

Φ *Value Restrained Simulations*. The final five snapshots taken from the EEF1.1 native simulations were used as starting structures for the restrained simulations. Two sets of BMD simulations were performed using the HQBMD (half-quadratic BMD) module of CHARMM to drive the system toward the two transition states, TS1 and TS2 (*30*). To generate a greater degree of sampling, we conducted the BMD simulations at five different temperatures (300, 360, 500, 640, and 780 K), as advocated by Paci et al. (*30*, *31*), to aid the generation of transition state ensembles. The putative transition states were restrained using the experimental Φ values for the NBR1 PB1 domain as shown in Table 1. The simulations were restrained using an energy penalty, $W(r,t)$, which was applied to the force field to restrain the simulations to the set of Φ values based upon the mean-square deviation, $\rho(t)$, between the experimental and calculated Φ values throughout the trajectory (eqs 1−3), where the calculated Φ value is determined from the fraction of native contacts formed by an individual residue (eq 3) and ρ_α is the previously measured deviation (*30*, *31*). The original X-ray crystal structure was used as the reference structure in the BMD calculations from which native contacts were computed.

$$W(r,t) = \frac{\alpha}{2}(\rho - \rho_\alpha)^2 \quad (10)$$

$$\rho(t) = \frac{1}{N_\Phi}\sum_{i\in}[\Phi_i^{calc}(t) - \Phi_i^{exp}(t)]^2 \quad (11)$$

$$\Phi_i^{calc} = \frac{N_i(t)}{N_i^{nat}} \quad (12)$$

A further parameter, β, is used in the calculation of the native contacts throughout the duration of the trajectory. On the basis of previous protocols (*30*, *31*), the parameters for restraining the simulations, α and β, were first optimized through repetitive short time scale calculations to restrain the systems to the experimental values satisfactorily. The BMD simulations were driven away from the reference structure over 5 ns using a β of 5.0 in the smoothing function used to calculate the native contacts (eq 13) based on a cutoff distance, r, of 6.5 Å with the force, α, set to 10000 and doubled every 400 ps.

$$\psi_r = \frac{1}{1 + e^{\beta_r}} \quad (13)$$

Once a correlation greater than 0.90 was reached between the experimental and calculated Φ values, the force α was reduced to that required to ensure that the system remained within the locality of the transition state, where the correlation coefficient remained above 0.9 and $\rho \leq 0.01$. The force was then kept constant while the temperature of the trajectories was increased, followed by equilibration for a further 1 ns at the required temperature. Once the sample had been equilibrated, with a stable temperature and fluctuations about the mean experimental Φ value, 5 ns of production dynamics at the respective temperature was collected. At all temperatures, the force was sufficient that the systems remained within the locality of the transition state with a correlation coefficient above 0.9 and a ρ of ≤0.01.

MD Trajectory Analysis. Production dynamics were analyzed using the available analytical modules within CHARMM to calculate the time-dependent properties: rmsd, radius of gyration, native contacts, and SASA. Before analysis, the domain was reorientated and rms fitted to the backbone of the starting

Table 1: Φ Values[a] for the Two Transition States of the PB1 Domain of NBR1

mutant[b]	mutation site	$\Delta\Delta G_{U-N}$ (kJ/mol)	Φ_{TS1}	Φ_{TS2}
L8A	β1, core	14.5 ± 0.2	0.24 ± 0.02	0.67 ± 0.03
N9A	β1	4.5 ± 0.2	0.01 ± 0.03	0.54 ± 0.03
V10A	β1, core	7.6 ± 0.3	0.40 ± 0.03	0.97 ± 0.03
I16A	β2	2.2 ± 0.1	0.47 ± 0.02	0.73 ± 0.02
F19L	β2, core	7.7 ± 0.3	0.11 ± 0.03	0.59 ± 0.04
L20A	β2	4.8 ± 0.2	0.22 ± 0.02	0.72 ± 0.03
A30G	α1	4.4 ± 0.2	0.65 ± 0.05	0.98 ± 0.04
I32A	α1, core	16.8 ± 0.4	0.56 ± 0.04	0.87 ± 0.03
V36A	α1, core	13.4 ± 1.3	[d]	0.74 ± 0.11
A37G[c]	α1	5.2 ± 0.4	0.28 ± 0.03	0.77 ± 0.03
I45A	β3, core	6.0 ± 0.6	0.47 ± 0.03	0.90 ± 0.03
Q46A	β3	2.8 ± 0.2	0.48 ± 0.04	0.47 ± 0.04
A58G[c]	β4	6.4 ± 0.6	0.33 ± 0.03	0.25 ± 0.03
Y65L	α2, core	9.4 ± 0.3	0.11 ± 0.02	[d]
A67G[c]	α2	8.4 ± 0.5	0.39 ± 0.03	0.65 ± 0.04
L69A	α2, core	14.1 ± 0.6	0.35 ± 0.03	0.35 ± 0.03
A70G	α2	6.5 ± 0.4	0.42 ± 0.04	0.76 ± 0.03
A73G[c]	α2	5.5 ± 0.4	0.40 ± 0.05	0.65 ± 0.05
A74G[c]	α2	2.9 ± 0.2	0.47 ± 0.03	0.77 ± 0.03
A80G[c]	β5	6.7 ± 0.5	−0.01 ± 0.02	0.52 ± 0.04
V83A	β5, core	2.4 ± 0.3	0.99 ± 0.05	1.00 ± 0.04

[a]Estimated from the data for wt-p62 for the corresponding TS. [b]A total of 28 single-point mutations were used to generate 38 Φ values (19 pairs of values for TS1 and TS2, excluding V36A and Y65L). [c]Φ values calculated from the double mutation of Xaa to Ala and Gly at this position. [d]Data fit to a simple two-state model which is either TS1-limited (Y65L) or TS2-limited (V36A).

X-ray structure to remove center-of-mass motion from the trajectories. The rmsd values of the C_α atoms and all heavy atoms from the X-ray structure were calculated for each frame of the simulation, and the percentage of native contacts was calculated over all trajectories for all residues. A native contact was counted between all heavy atom side chain−side chain contacts within a 6.5 Å cutoff in the X-ray crystal structure. The SASA was determined using Lee and Richard's algorithm (*32*), with a spherical water probe with a radius of 1.4 Å. The secondary structure content over the ensemble of the five replicas was calculated using the DSSP algorithm (*33*) within Simulaid (*34*).

The trajectory data, which satisfied the experimental restraints, were clustered using the ART2′ algorithm within CHARMM (*35*). Because of the potential vulnerability of the ART2′ algorithm to generate results dependent on the data order, both randomized and ordered data sets were used. A cutoff radius of 500 $Å^2$ was selected to reflect approximately twice the standard deviation observed for the simulated native state SASA as well as the generation of consistent clusters without creating single-member clusters.

RESULTS

Point Mutation Analysis of TS1 and TS2. We constructed 28 single-point mutants for detailed kinetic studies and were able to generate 38 Φ values (19 pairs of values at each position for TS1 and TS2) (see Table 1) (*36*, *37*). Mutations (typically Ala) were introduced within each of the five strands of the β-sheet and within each helix to probe both secondary structure formation and core hydrophobic interactions between elements of structure. In a number of cases where surface-exposed charged side chains were targeted, or where Ala mutations produced only small changes in stability, both Ala and Gly mutations were introduced at the same site and Φ values were calculated via comparison of the Ala and Gly variant at that position. All mutants were analyzed under the same conditions using GdmCl denaturant at 298 K and pH 5.0 in 25 mM acetate buffer (*12*).

The kinetic data for wt-PB1 and the vast majority of the mutants gave rise to a single-exponential refolding rate constant that accounted for >90% of the fluorescence signal. The remaining <10% could be ascribed to a slow phase with little variation in rate or amplitude with denaturant concentration, typical of an isomerization-limited process. Unfolding was described well by a single-exponential process. However, chevron plots reveal downward curvature in the unfolding arm that is consistent with a denaturant-dependent switch between two different rate-limiting folding barriers. The data were fitted to a sequential barrier model in which the two transition states are separated by a metastable intermediate (*12*). The curvature in the unfolding arm of the chevron is associated with a change in the limiting slope (m_{N-TS}) as either the early or late transition state becomes rate-limiting under different experimental conditions. The two limiting chevrons, each of which is associated with a different β_{TS} value (0.71 for TS1 and 0.93 for TS2 for wt-PB1), are shown in Figure 1a, together with a schematic representation of the switch in the rate-limiting step at high and low denaturant concentrations (Figure 1c).

The chevron plots were initially fitted individually to the analytical solutions of the three-state model to yield values for k_{UI*}, k_{NI*}, and the k_{I*N}/k_{I*U} ratio. Fitting parameters were restrained to ensure that the intermediate (I*) remained high in energy and is never populated (see Materials and Methods). As previously described in a number of studies (*10*, *12*, *14*), k_{I*U} and m_{I*U} were given arbitrary fixed values of 10^4 s^{-1} and 0 kJ/mol, respectively, although wide variation in these parameters had little effect on the derived rate constants and relative *m* values but reduced errors by minimizing the number of iterated variables. Kinetic data were then fitted globally taking into consideration fluorescence amplitudes, the denaturant dependence of the rate constants (*m* values), and fluorescence end points. In the case of a number of the mutants, the curvature within the unfolding arm of the chevron is less well-defined within the accessible range of denaturant concentration because of differential stabilization of one transition state or the other. Individual fitting of chevron plots resulted in large variations in *m* values (*10*, *11*, *38*), which were inconsistent with equilibrium fluorescence unfolding studies. However, global fitting using shared *m* values (m_{UI*} = 7.3 ± 0.1 kJ mol^{-1} M^{-1}, m_{NI*} = 0.66 ± 0.05 kJ mol^{-1} M^{-1}, and $m_{I*N} - m_{I*U}$ = 2.30 ± 0.1 kJ mol^{-1} M^{-1}, which give the β_{TS1} and β_{TS2} values indicated above) resulted in a significant reduction in errors and a higher-quality fit, as illustrated for the selection of chevrons shown in Figure 2 (see also Figure S1 of the Supporting Information). The kinetic data for all of the mutants are presented in Table S1 of the Supporting Information.

Although the energy of the intermediate could not be determined directly by this approach, the sizes of the kinetic barriers, and hence the relative stability of the two transition states, could be determined from k_{I*N}/k_{I*U} (see Figure 1b). The folding and unfolding rate constants for TS1 and TS2, defined by various combinations of microscopic rate constants (see Materials and Methods), allowed us to generate 32 Φ values from this data set, taking a lower limit for changes in stability ($\Delta\Delta G_{UN}$) of >2 kJ/mol as a prerequisite for the determination of accurate Φ values (*37*, *39*, *40*). Experimental variations in $\Delta\Delta G_{UN}$ for point mutations were in the range of 0−17 kJ/mol (see Table 1).

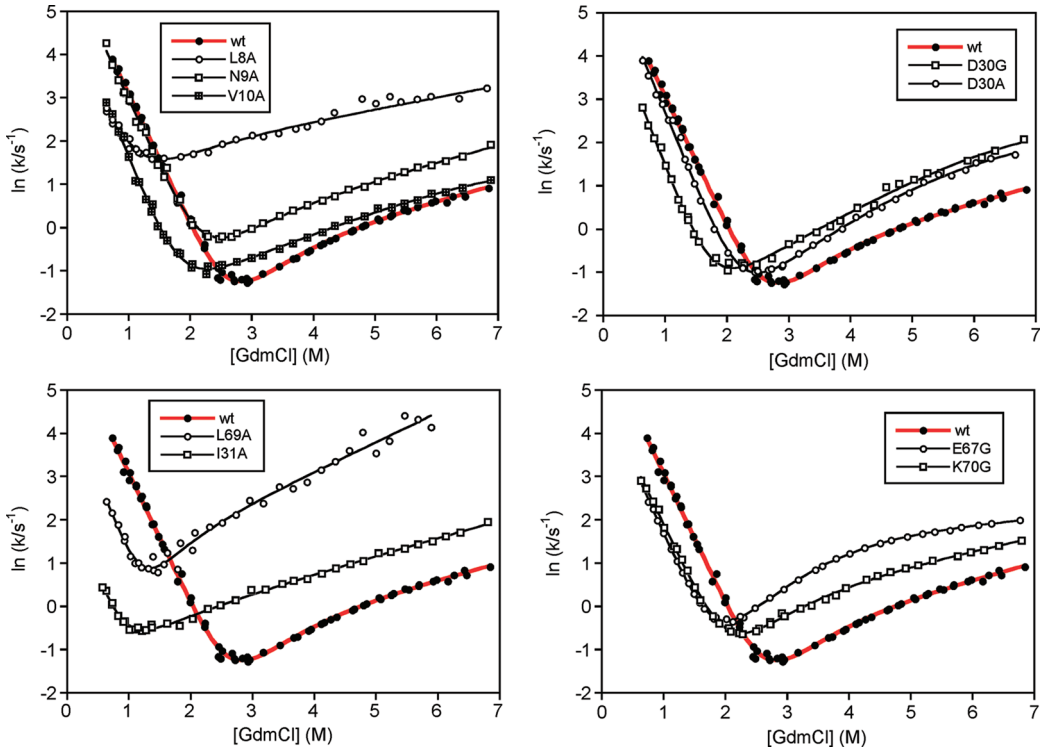

FIGURE 2: Chevron plots of the kinetic data for wt-PB1 and a selection of mutants used for Φ value analysis (see also Figure S1 of the Supporting Information). In all cases, the data for wt-PB1 are shown as black dots with a red line. All of the data were fitted globally assuming the same values for the denaturant dependence of the refolding and unfolded rates; changes in stability and calculated Φ values are listed in Table 1. All kinetic data were recorded at 298 K in 25 mM acetate buffer (pH 5.0).

Leffler Analysis of Folding Transition States. The effects of point mutations on folding kinetics were analyzed using Leffler plots, where the change in the activation barrier for either TS1 ($\Delta\Delta G_{TS1}$) or TS2 ($\Delta\Delta G_{TS2}$) was correlated with the overall change in protein stability ($\Delta\Delta G_{UN}$) (*41*). Leffler plots for TS1 and TS2 are shown in panels a and b of Figure 3, bounded by the $\alpha = 0$ (unfolded TS) and $\alpha = 1$ (nativelike TS structure) limiting slopes. The TS1 data fit with a linear correlation ($R = 0.85$) and a slope α_{TS1} of 0.38 ± 0.05. Although a few points lie close to the $\alpha = 1$ line, a number of these are associated with relatively small changes in protein stability. However, the majority correspond to low to medium Φ values, characteristic of a diffuse transition state. In contrast, the data for TS2 ($R = 0.90$) have a higher mean value of α_{TS2} (0.67 ± 0.07). Very few individual points lie close to the $\alpha = 0$ line, whereas a significant number show values close to $\alpha = 1$, indicating that TS2 is structurally more nativelike.

Although both TS1 and TS2 represent relatively compact states ($\beta_{TS1} = 0.71$, and $\beta_{TS2} = 0.93$), they are structurally distinct species with quite different Leffler plots. The shallower slope for TS1 indicates that the widely distributed residues sampled in the analysis are weakly formed in a loose nativelike network of contacts, typical of nucleation−condensation mechanisms of folding. The observation that TS1 is relatively compact ($\beta_{TS1} = 0.71$) suggests that weak long-range tertiary interactions are in place and hydrophobic side chains are already substantially buried. This has been suggested as an evolutionary pressure on protein folding pathways to limit the possibilities of aggregation (*22*, *42*). The compact polypeptide chain appears to be poised for further consolidation and stabilization of native secondary and tertiary structure. Plotting the change in activation energy for TS1 versus TS2 for the family of mutants reveals a linear correlation (Figure 3c), which suggests that the two transition states are related as different expanded versions of the native state, each with a subset of the same nativelike interactions (*10*).

Structural Ensembles for TS1 and TS2 Derived from Molecular Dynamics Simulations. We explored the structural characteristics of TS1 and TS2 using BMD simulations restrained using experimental Φ values (*30*, *31*). Ensembles of structures representing the two folding transition states were generated over a range of temperatures between 300 and 780 K to ensure the widest possible conformational sampling. An initial selection of conformers that had satisfied the experimental Φ value restraints were clustered on the basis of the correspondence of the solvent accessible surface areas (SASA) with the experimental β_T values (*43*). The average SASA of the native state calculated from equilibrium MD simulations is 5351 ± 196 Å^2. On the basis of an experimentally determined β_{TS1} of 0.71 and a β_{TS2} of 0.93, the expected ranges for the ensemble averages of TS1 and TS2 were 6333−7632 and 5347−6143 Å^2, respectively. Using a cluster radius of 500 Å^2 to reflect the expected variance, 11 and 6 clusters were identified for TS1 and TS2, respectively. All conformers from the generated clusters were used to calculate SASA ensemble averages of 6905 ± 381 and 5426 ± 102 Å^2, respectively (Table 2). Representative structures from each of the TS1 and TS2 clusters are shown in Figure 4, demonstrating the contrast between the secondary structure content of the two transition states. The experimental Φ value restraints are well satisfied for both ensembles, as shown in Figure 5, with correlation coefficients of > 0.95 between Φ^{calc} and Φ^{expt} for each transition state.

The distribution of relatively low Φ values for TS1, which suggests that it is a rather diffuse collapsed state, is supported by the simulations that show a heterogeneous broad ensemble of structures. The radius of gyration of the TS1 ensemble increases by a modest ~16% compared with the native state ensemble.

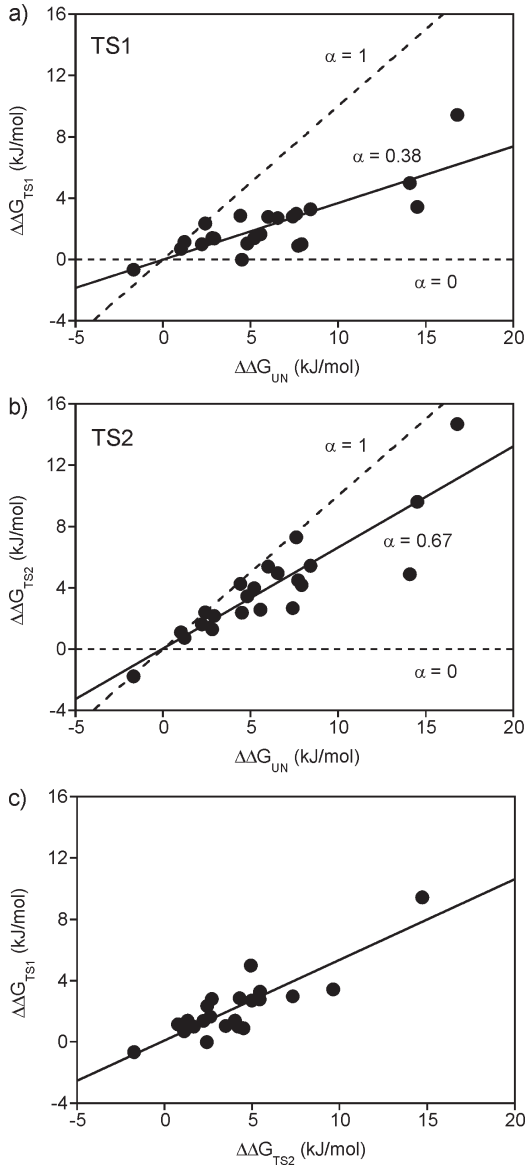

FIGURE 3: Leffler analysis of the kinetic data for the PB1 mutants. Plot of the change in equilibrium stability ($\Delta\Delta G_{UN}$) vs the change in activation free energy for TS1 ($\Delta\Delta G_{TS1}$) (a) and TS2 ($\Delta\Delta G_{TS2}$) (b) for each mutant vs wt-PB1 derived from differences in the folding and unfolding rates. Lines of best fit are shown as solid lines with slope α. In the case of TS1, $\alpha = 0.38 \pm 0.05$ ($R = 0.85$), and for TS2, $\alpha = 0.67 \pm 0.07$ ($R = 0.90$). The limiting slopes corresponding to $\alpha = 1$ and $\alpha = 0$ are shown as dotted lines in both panels a and b. In panel c, the changes in the activation parameters for TS1 and TS2 are shown to be linearly correlated ($R = 0.88$).

However, there is a significant increase in the average C_α rmsd between pairs of structures, from 1.5 ± 0.2 Å in the native state to 8.6 ± 2.1 Å for the TS1 ensemble (Table 2). A closer examination of secondary structure content also reveals significant disorder in TS1 compared with the native state (see Table 2 and Figure 6). The average structures for TS1 and TS2 shown in panels c and d of Figure 6 demonstrate the structural regions in which the variability is at its greatest across both transition state ensembles. The contrast between the two TS ensembles is also apparent in the side chain contact maps of TS1 and TS2 shown in Figure 7. Although all elements of secondary structure are represented within the heterogeneous ensemble of TS1 (Figure 4), few interact or are consolidated in a nativelike manner. As a result, TS1 also exhibits evidence of some transient non-native secondary structure.

Table 2: Structural Parameters from the Ensemble of Structures Representing TS1 and TS2 from Biased Molecular Dynamics Simulations

	SASA (Å^2)	C_α rmsd (Å)	R_g (Å)	% α-helix	% β-sheet
native	5326 ± 105	1.5 ± 0.2	12.4 ± 0.1	30 ± 3	37 ± 5
TS1	6905 ± 381	8.6 ± 2.1	14.4 ± 0.8	6 ± 8	16 ± 7
TS2	5426 ± 102	2.4 ± 0.4	12.7 ± 0.1	29 ± 5	39 ± 5

In the native state, β1, flanked by β5 and β2 (Figure 1d), participates in mixed antiparallel/parallel hydrogen bonding patterns; however, in TS1, we observe a non-native parallel conformation in a subset of structures with persistent hydrogen bonding between β1 and β5.

Within the broader heterogeneous profile of TS1, there are some examples of persistent structure associated with both high experimental and calculated Φ values (> 0.6) [namely, residues T11, K13, A30, A34, V38, K48, Q80, Q82, V83, and H84 (see Figure 5a,c)]. The average structure for the TS1 ensemble (Figure 6c) shows that this cluster involves residues from across the β-sheet (β1, β3, and β5) and from α1, forming a weak folding nucleus. Our earlier analysis suggested that helix α2 might act as a folding nucleation site because of its high intrinsic helix propensity (~30%) (*12*). However, five experimental TS1 Φ values for residues in α2 fall below 0.5, and high structural variability is seen for α2 within the ensemble of TS1 structures.

In contrast, the general distribution of much higher Φ values for TS2, determined experimentally and from the simulated ensemble (Figure 5b,d), reveals that TS2 is much more compact and nativelike than the broad ensemble seen for TS1 (Figure 6d) with extensive native side chain contacts (Figure 7). The ensemble of structures selected for TS2 in terms of buried surface area, radius of gyration, and average C_α rmsd identifies a highly native-like topology (Table 2). The simulations also show that an essentially nativelike proportion of secondary structure is present and persists throughout (Figure 6b). The TS2 Φ values identify some residues within loops that are still substantially disordered (notably the β2−α1 loop) but likely provide additional "gatekeeping" conformational flexibility in TS2 to facilitate consolidation of core packing and secondary structure alignment.

Mutants Dominated by One Folding Transition State or the Other. Many folding pathways for small proteins are dominated by a single transition state under a wide range of experimental conditions, rather than exhibiting the switch between transition states on a sequential pathway as different barriers become rate-limiting under different conditions. The fortuitously small energy difference between TS1 and TS2 for the PB1 domain is comparable to the effect on stability of a single point mutation and could alter the relative free energies of TS1 and TS2, allowing one barrier or the other on the sequential transition state pathway to dominate the folding kinetics. During the analysis, we identified the core mutations V36A and Y65L that destabilize the wt-PB1 domain by 9−13 kJ/mol and result in chevron plots lacking curvature in the unfolding arm (Figure 8a). In both cases, the chevrons fit to the sequential transition state model but with significant changes in the relative stabilities of TS1 and TS2 compared with those of wt-PB1. While TS1 is higher in energy than TS2 for wt-PB1 under conditions that strongly favor the native folded state ($\Delta\Delta G_{TS1/TS2} \sim 9$ kJ/mol), the relative energies of TS1 and TS2 are reversed for V36A ($\Delta\Delta G_{TS1/TS2} \sim -10$ kJ/mol), ensuring that TS2 is always the rate-limiting barrier. In the case of Y65L, the difference between TS1

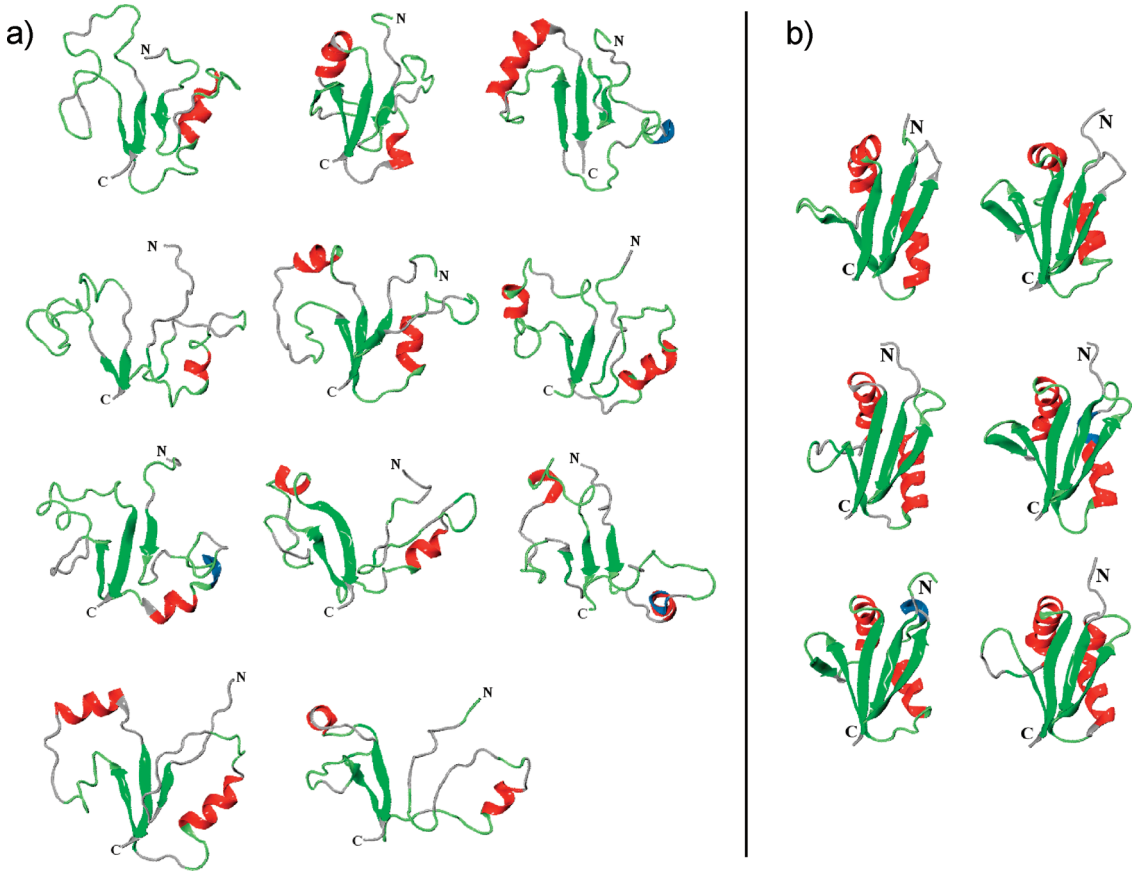

FIGURE 4: Structures from the TS1 and TS2 clusters generated using BMD simulations. Ensembles of structures were generated over the temperature range of 300−700 K, and clustering was performed on the basis of the fit to experimental Φ value restraints and solvent accessibility (β_T) values. In panel a, 11 clusters were identified for TS1 within the expected variance range with representative structures from each cluster shown. (b) Representatives of the six clusters identified for TS2.

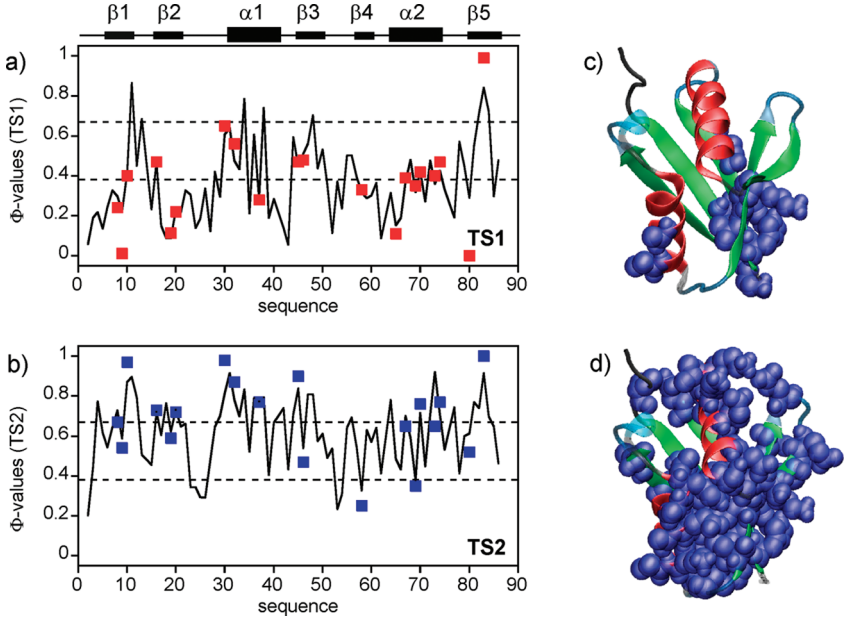

FIGURE 5: Experimental vs calculated Φ values from BMD simulations. Comparison of experimental and calculated data as a function of the position along the PB1 sequence for TS1 (a) and TS2 (b). In both cases, the experimental data are shown as squares and the calculated data at every position as the solid line; the correlation coefficient between Φ^{calc} and Φ^{expt} for each transition state is > 0.95. The dotted lines shown in both plots represent the Leffler α values derived for the same data set from the slopes of the plots in panels a and b of Figure 3 ($\alpha = 0.38$, and $\alpha = 0.67$). The ribbon structure of the PB1 domain is shown to highlight residues that have high Φ values ($\Phi_{TS1} > 0.67$) in the BMD calculations of TS1 (c) and TS2 (d).

and TS2 has the same sign as that of wt-PB1 but is much larger ($\Delta\Delta G_{TS1/TS2} \sim 19$ kJ/mol), showing selective stabilization of TS2, ensuring that in this case TS1 is always rate-limiting (Figure 8b). These mutants give negative Φ values for one transition state or

the other and are not compatible with the Leffler analysis, which measures incremental changes in stability and activation parameters, assuming minimal perturbations to the nature of the two transition states. As a consequence, the data for V36A and Y65L were excluded from the Leffler analysis in Figure 3. Assuming a simple two-state folding model, the calculation of the β_T values for V36A and Y65L reveals $\beta_{TS}^{V36A} = 0.93$ and $\beta_{TS}^{Y65L} = 0.73$, in good agreement with values determined for TS1 and TS2 from the sequential TS model for wt-PB1.

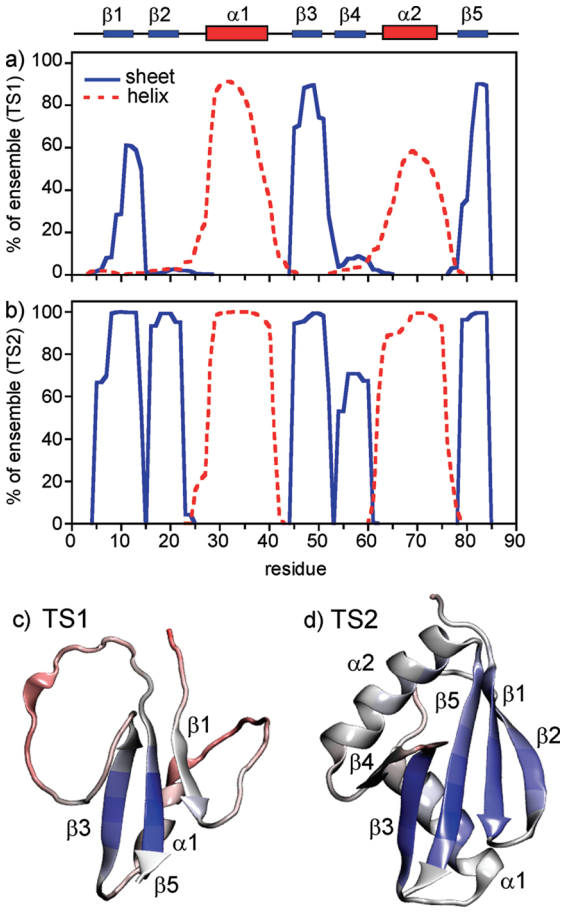

FIGURE 6: Secondary structure analysis of the (a) TS1 and (b) TS2 ensembles generated using the BMD simulations. The secondary structure calculation was performed using the DSSP algorithm (*33*) within Simulaid (*34*), and the structural content has been calculated by type as a percentage of the ensemble for each residue within the NBR1 PB1 domain. Average structures computed for the (c) TS1 and (d) TS2 ensembles shaded with respect to the rmsd across the ensembles. Blue shading indicates areas for which a low rmsd was calculated, changing through gray to red as the rmsd increases. The average structure for TS1 suggests that β-strands β1, β3, and β5 and helix 1 constitute a weak folding nucleus with low Φ values within an otherwise heterogeneous ensemble of structures.

DISCUSSION

Similarities of TS1 and TS2 to Transition States of Classical Folding Mechanisms. We have characterized the structural ensemble of the two sequential transition states on the folding pathway of the PB1 domain of NBR1. The pattern of fractional Φ values for TS1 is consistent with formation of a globally collapsed state with a loosely nativelike topology ($\beta_{TS1} = 0.73$). The mean α_{TS1} value of 0.38 derived from the Leffler analysis agrees well with slopes of 0.36 and 0.31 reported for CI2 and the homeodomain protein c-Myb (*8*), respectively, both of which have been described in terms of a nucleation−condensation mechanism. A high degree of conformational heterogeneity is apparent in the BMD simulations for the PB1 domain, which suggest that TS1 is a diffuse ensemble of structures representing a broad transition state barrier. However, a cluster of residues involving β-strands β1, β3, and β5 and part of helix 1, which have higher experimental and calculated Φ values, shows more persistent nativelike structural propensities, consistent with formation of a weak folding nucleus (Figures 6c and 7). This rather dispersed nucleus is associated with a high "effective contact order" in terms of the distance between interacting residues along the polypeptide chain (*44*).

In contrast, polarized transition states are more typically associated with local stable structural elements (β-turns or β-hairpins), as described for SH3, WW domains, and ubiquitin (*45*−*48*). The PB1 domain adopts a ubiquitin-like fold, and a comparison of the chevron plots for the two proteins shows that the TS1-limited data for the PB1 domain closely resemble that of ubiquitin with similar kinetic *m* values and solvent accessibilities of the two respective transition states (Figure 8c). Both Φ value analysis and ψ value analysis of the native state of ubiquitin (*47*−*49*), as well as peptide fragment folding studies by NMR (*50*), have shown evidence of a folding nucleation site centered around the β-turn of the N-terminal hairpin. In contrast,

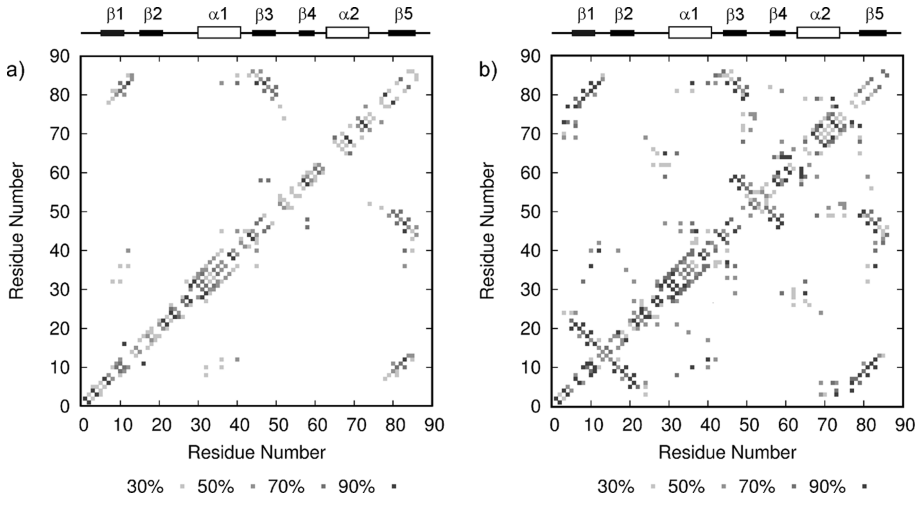

FIGURE 7: Contact maps for the (a) TS1 and (b) TS2 ensembles shaded with respect to the percentage of the ensembles for which the contacts were present (see the key). Secondary structure elements are highlighted in panel b.

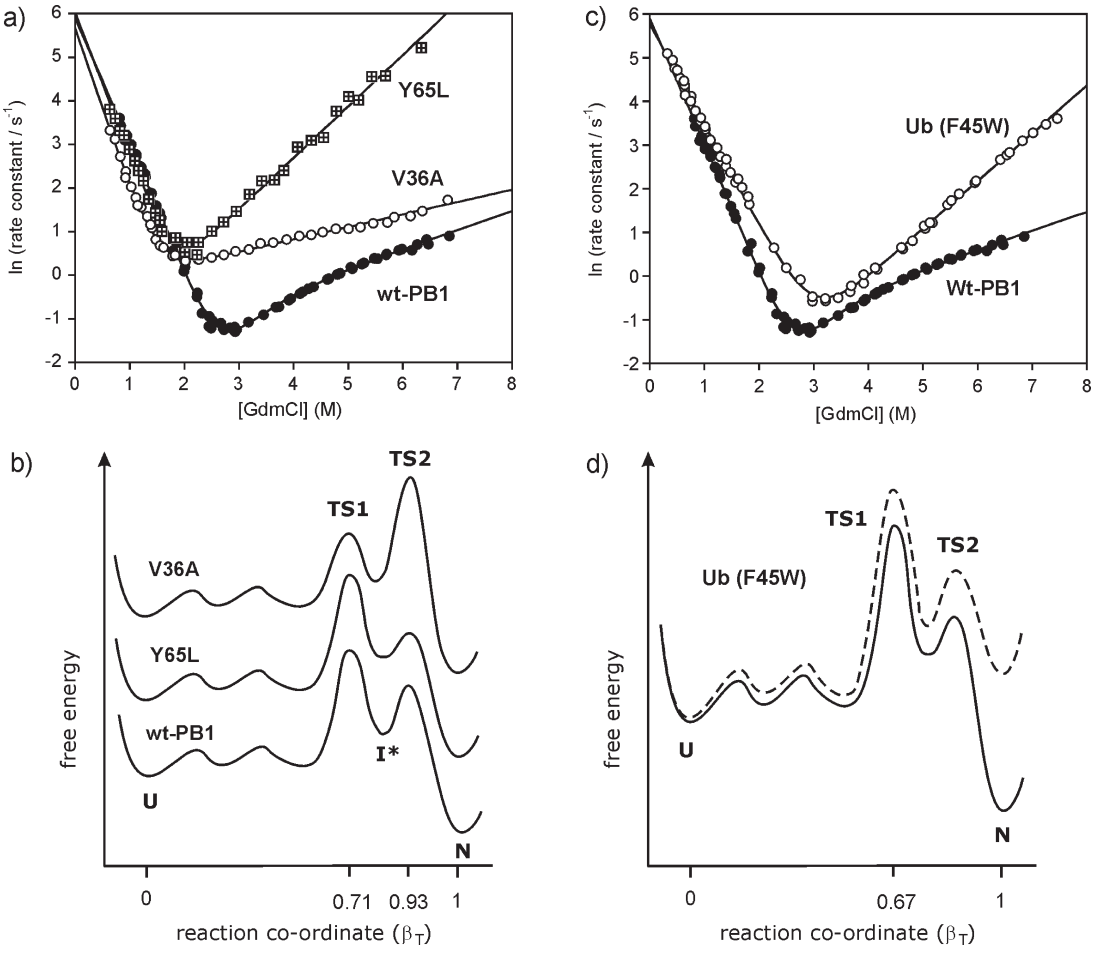

FIGURE 8: (a) Kinetic analysis of PB1 mutants V36A and Y65L (comparison with wt-PB1) showing linear chevron plots consistent with a single dominant rate-limiting barrier under all experimental conditions. (b) Schematic energy landscape showing how the mutations appear to affect the relative barrier heights such that TS2 is rate-limiting for V36A and TS1 rate-limiting for Y65L. The curves are separated for the sake of clarity, but the energy of the unfolded state (U) is the same in each case. (c) The chevron plot for wt-PB1 is compared with that of F45W yeast ubiquitin [Ub (F45W)], where the data approximate an apparent two-state model at 298 K. Analysis of *m* values shows that the TS for ubiquitin is similar to that for TS1 of wt-PB1 and that of the Y65L PB1 mutant. By analogy, similar energy landscapes can be envisaged for both wt-PB1 and Ub in which only small differences in the relative energies of the activation barriers account for the observation that TS1 is always rate-limiting for Ub (F45W). (d) Energy landscape for Ub (F45W) under native (—) and unfolding (---) conditions.

the N-terminal β-hairpin of the PB1 domain (residues 6−21) shows a cluster of six experimental Φ values in the range of 0−0.47 [mean value of 0.24 (see Table 1)], which suggest a contribution to a diffuse transition state rather than one that is highly polarized toward the N-terminal hairpin.

In contrast to the broad barrier associated with TS1, the TS2 ensemble has many individual Φ values lying close to unity, demonstrating that nativelike secondary structure and tertiary contacts are well formed in TS2. In the BMD simulations, this is reflected in a nativelike radius of gyration and low average pairwise C_α rmsd between structures in the TS2 ensemble (see Table 2). We see few low Φ values for TS2 and a relatively homogeneous distribution of high Φ values, which appear to indicate that a high proportion of native contacts are formed simultaneously. The correlation between the stabilities of TS1 and TS2 (Figure 3c) strongly supports the notion that, despite the heterogeneity in the TS1 ensemble, both TS1 and TS2 represent different expanded or distorted forms of the native state with the relative stabilities acquired through a subset of the same nativelike interactions (*7*, *8*, *10*, *11*, *22*). Studies of the folding mechanism of two homologous PDZ domains (*10*, *11*) similarly conclude that the early folding events reflect a much weaker nativelike bias resulting in a diffuse ensemble of isoenergetic structures (*5*−*7*).

The development and consolidation of the hydrophobic core not only determine protein stability but also drive the folding process. The relevant timing of desolvation events and the formation of the tightly packed core remain key questions in understanding the folding mechanism. In this regard, several recent studies have used multiple substitutions at specific sites to probe solvent accessibility within transition states and intermediates, as well as their malleability and tolerance to overpacking and the nonspecific hydrophobic burial of side chains (*39*, *51*, *52*). Studies of the α-spectrin SH3 domain and immunity protein Im7 have used Val → Thr substitutions to probe the solvation of core residues at key junctures along the folding pathway, revealing that different extents of solvation are evident for different core residues in the TS ensemble. Im7 folds via a populated intermediate, and many high Φ_I and Φ_{TS2} values for these atypical residue substitutions indicate a significant proportion of the core being buried in the early stages of folding. For other residues, low Φ values indicate that some portions of the structure remain solvated until the very late stages of consolidation of the native core. The heterogeneity observed in the ensemble simulations for TS1 of the PB1 domain, together with the distribution in observed Φ values, similarly suggests that a subset of residues are protected at an early stage in folding but others remain

dynamic and highly solvent accessible within the context of the collapsed polypeptide chain (Figure 5a). The conclusions from the BMD simulations warrant further experimental verification; however, differential desolvation appears to be emerging as a key factor in driving folding around a nucleus of nativelike interactions.

ACKNOWLEDGMENT

We are grateful for access to the University of Nottingham High Performance Computer. We also thank Dr. J. Long for assistance with the optimization of protein expression and purification protocols and Jonathan Phillips for providing some additional kinetic data.

SUPPORTING INFORMATION AVAILABLE

Kinetic data for mutants of the PB1 domain of NBR1 in 25 mM sodium acetate buffer pH 5.0 at 298K (Table S1) and Chevron plots of the kinetic data for *wt*-PB1 (wt) and all of the mutants used for Φ-value analysis (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Kim, P. S., and Baldwin, R. L. (1990) Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem. 59*, 631–660.
2. Karplus, M., and Weaver, D. L. (1994) Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Sci. 3*, 650–668.
3. Jackson, S. E. (1998) How do single domain proteins fold? *Folding Des. 3*, R81–R91.
4. Wong, K. B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M., Fersht, A. R., and Daggett, V. (2000) Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol. 296*, 1257–1282.
5. Daggett, V., and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding? *Trends Biochem. Sci. 28*, 18–25.
6. Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (2005) Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem. 48*, 545–600.
7. Fersht, A. R. (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U.S.A. 92*, 10869–10873.
8. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. N., DeMarco, M. L., Daggett, V., and Fersht, A. R. (2003) Unifying features in protein folding mechanisms. *Proc. Natl. Acad. Sci. U.S.A. 100*, 13286–13291.
9. White, G. W. N., Gianni, S., Grossman, J. G., Jemth, P., Fersht, A. R., and Daggett, V. (2005) Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J. Mol. Biol. 350*, 757–775.
10. Gianni, S., Geierhaas, C. D., Calosci, N., Jemth, P., Vuister, G. W., Travaglini- Allocatelli, C., Vendruscolo, M., and Brunori, M. (2007) A PDZ domain recapitulates a unifying mechanism for protein folding. *Proc. Natl. Acad. Sci. U.S.A. 104*, 128–133.
11. Calosci, N., Chi, C. N., Richter, B., Camilloni, C., Engstrom, A., Eklund, L., Travaglini-Allocatelli, C., Gianni, S., Vendruscolo, M., and Jemth, P. (2008) Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proc. Natl. Acad. Sci. U.S.A. 105*, 19240–19245.
12. Chen, P., Long, J., and Searle, M. S. (2008) Sequential barriers and an obligatory metastable intermediate define the apparent two-state folding pathway of the ubiquitin-like PB1 domain of NBR1. *J. Mol. Biol. 376*, 1463–1477.
13. Wright, C. F., Lindorff-Larsen, K., Randles, L. G., and Clarke, J. (2003) Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. Biol. 8*, 658–662.
14. Bachmann, A., and Kiefhaber, T. (2001) Apparent two-state Tendamistat folding is a sequential process along a defined route. *J. Mol. Biol. 306*, 375–386.
15. Sanchez, I. E., and Kiefhaber, T. (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol. 325*, 367–376.
16. Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E., and Fersht, A. R. (1995) Movement of the position of the transition state in protein folding. *Biochemistry 34*, 13656–13662.
17. Oliveberg, M., Tan, Y.-J., Silow, M., and Fersht, A. R. (1998) The changing nature of the protein folding transition state: Implications for the shape of the free energy profile for folding. *J. Mol. Biol. 277*, 933–943.
18. Otzen, D. E., Kristensen, O., Proctor, M., and Oliveberg, M. (1999) Structural changes in the transition state of protein folding: Alternative interpretations of curved chevron plots. *Biochemistry 38*, 6499–6511.
19. Sanchez, I. E., and Kiefhaber, T. (2003) Hammond behaviour versus ground state effects in protein folding: Evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol. 327*, 867–884.
20. Muller, S., Kursula, I., Zou, P., and Wilmanns, M. (2006) Crystal structure of the PB1 domain of NBR1. *FEBS Lett. 580*, 341–344.
21. Ito, T., Matsui, Y., Ago, T., Ota, K., and Sumimoto, H. (2001) Novel modular domain PB1 recognises PC motif to mediate functional protein-protein interactions. *EMBO J. 20*, 3938–3946.
22. Geierhaas, C. D., Best, R. B., Paci, E., Vendruscolo, M., and Clarke, J. (2006) Structural comparison of the two alternative transition states for folding of TI I27. *Biophys. J. 91*, 263–275.
23. Crespo, M. D., Simpson, E. R., and Searle, M. S. (2006) Population of on-pathway intermediates in the folding of ubiquitin. *J. Mol. Biol. 360*, 1053–1066.
24. Bofill, R., Simpson, E. R., Platt, G. W., Crespo, M. D., and Searle, M. S. (2005) Extending the folding nucleus of ubiquitin with an independently folding β-hairpin finger: Hurdles to rapid folding arising from the stabilisation of local interactions. *J. Mol. Biol. 349*, 205–221.
25. Bachmann, A., and Kiefhaber, T. (2005) in Kinetic mechanisms in protein folding (Kiefhaber, T., and Buchner, J., Eds.) Wiley-VCH, Weinheim, Germany.
26. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., III, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) Molprobity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res. 35*, W375–W383.
27. Lazaridis, T., and Karplus, M. (1997) "New View" of protein folding reconciled with the old through multiple unfolding simulations. *Science 278*, 1928–1931.
28. Brooks, B. R., Brooks, C. L., III, MacKerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., and Karplus, M. (2009) CHARMM: The biomolecular simulation program. *J. Comput. Chem. 30*, 1545–1614.
29. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684–3690.
30. Paci, E., and Karplus, M. (1999) Forced unfolding of fibronectin type 3 modules: An analysis by biased molecular dynamics simulations. *J. Mol. Biol. 288*, 441–459.
31. Paci, E., Vendruscolo, M., Dobson, C. M., and Karplus, M. (2002) Determination of a transition state at atomic resolution from protein engineering data. *J. Mol. Biol. 324*, 151–163.
32. Lee, B., and Richards, F. M. (1971) The interpretation of protein structure: Estimation of static accessibility. *J. Mol. Biol. 55*, 379–400.
33. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577–2637.
34. Mezei, M. (1997) Optimal position of solute for simulations. *J. Comput. Chem. 18*, 812–815.
35. Karpen, M. E., Tobias, D. J., and Brooks, C. L., III (1993) Clustering techniques for the analysis of long molecular dynamics trajectories: Analysis of 2.2-ns trajectories of YPGDV. *Biochemistry 32*, 412–420.
36. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol. 224*, 771–782.
37. Fersht, A. R., and Sato, S. (2004) Φ-Value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. U.S.A. 101*, 7976–7981.
38. Lowe, A. R., and Itzhaki, L. S. (2007) Rational redesign of the folding pathway of a modular protein. *Proc. Natl. Acad. Sci. U.S.A. 104*, 2679–2684.
39. Bartlett, A. I., and Radford, S. E. (2010) Desolvation and development of specific hydrophobic core packing during Im7 folding. *J. Mol. Biol. 396*, 1329–1345.
40. Sanchez, I. E., and Kiefhaber, T. (2003) Origin of unusual f-values in protein folding: Evidence against specific nucleation sites. *J. Mol. Biol. 334*, 1077–1085.

41. Otzen, D. E., Itzhaki, L. S., ElMasry, N. F., Jackson, S. E., and Fersht, A. R. (1994) Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc. Natl. Acad. Sci. U.S.A. 91*, 10422–10425.
42. Dobson, C. M. (1999) Protein misfolding, evolution and disease. *Trends Biochem. Sci. 24*, 329–332.
43. Creamer, T. P., Srinivasan, R., and Rose, G. D. (1997) Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry 36*, 2832–2835.
44. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and refolding rates of single domain proteins. *J. Mol. Biol. 277*, 985–994.
45. Grantcharova, V. P., Riddle, D. S., Santiago, J. V., and Baker, D. (1998) Important role of hydrogen bonds in the structurally polarised transition state for folding of the src SH3 domain. *Nat. Struct. Biol. 5*, 714–720.
46. Ferguson, N., Pires, J. R., Toepert, F., Johnson, C. M., Pan, Y. P., Volkmer-Engert, R., Schneider-Mergener, J., Daggett, V., Oschkinat, H., and Fersht, A. R. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 13008–13013.
47. Went, H. M., and Jackson, S. E. (2005) Ubiquitin folds through a highly polarised transition state. *Protein Eng., Des. Sel. 18*, 239–246.
48. Simpson, E. R., Meldrum, J. K., and Searle, M. S. (2006) Engineering diverse changes in β-turn propensities in the N-terminal β-hairpin of ubiquitin reveals significant effects on stability and kinetics but a robust folding transition state. *Biochemistry 45*, 4220–4230.
49. Krantz, B. A., Dothager, R. S., and Sosnick, T. R. (2004) Discerning the structure and energy of multiple transition states in protein folding using Ψ-analysis. *J. Mol. Biol. 337*, 463–475.
50. Searle, M. S., and Ciani, B. (2004) Design of β-sheet systems for under standing the thermodynamics and kinetics of protein folding. *Curr. Opin. Struct. Biol. 14*, 458–464.
51. Weikl, T. R., and Dill, K. A. (2003) Folding rates and low entropy-loss routes of two-state proteins. *J. Mol. Biol. 329*, 585–598.
52. Fernandez-Escamilla, A. M., Cheung, M. S., Vega, M. C., Wilmanns, M., Onuchi, J. N., and Serrano, L. (2004) Solvation in protein folding analysis: Combination of theoretical and experimental approaches. *Proc. Natl. Acad. Sci. U.S.A. 101*, 2834–2839.