Malleability of folding intermediates in the homeodomain superfamily

Wiktor Banachewicz[a], Tomasz L. Religa[a], R. D. Schaeffer[b], Valerie Daggett[b], and Alan R. Fersht[a,1]

[a]Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, United Kingdom; and [b]Department of Bioengineering, P.O. Box 355013, University of Washington, Seattle, WA 98195-5013

Contributed by Alan R. Fersht, January 31, 2011 (sent for review January 21, 2011)

Members of the homeodomain superfamily are three-helix bundle proteins whose second and third helices form a helix-turn-helix motif (HTH). Their folding mechanism slides from the ultrafast, three-state framework mechanism for the engrailed homeodomain (EnHD), in which the HTH motif is independently stable, to an apparent two-state nucleation-condensation model for family members with an unstable HTH motif. The folding intermediate of EnHD has nearly native HTH structure, but it is not docked with helix1. The determinant of whether two- or three-state folding was hypothesized to be the stability of the HTH substructure. Here, we describe a detailed Φ-value analysis of the folding of the Pit1 homeodomain, which has similar ultrafast kinetics to that of EnHD. Formation of helix1 was strongly coupled with formation of HTH, which was initially surprising because they are uncoupled in the EnHD folding intermediate. However, we found a key difference between Pit1 and EnHD: The isolated peptide corresponding to the HTH motif in Pit1 was not folded in the absence of H1. Independent molecular dynamics simulations of Pit1 unfolding found an intermediate with H1 misfolded onto the HTH motif. The Pit1 folding pathway is the connection between that of EnHD and the slower folding homeodomains and provides a link in the transition of mechanisms from two- to three-state folding in this superfamily. The malleability of folding intermediates can lead to unstable substructures being stabilized by a variety of nonnative interactions, adding to the continuum of folding mechanisms.

protein folding | temperature jump | diffusion collision | transition state

The nucleation-condensation mechanism was proposed from a Φ-value analysis of the folding of the protein chymotrypsin inhibitor 2 (CI2) (1, 2). The individual elements of secondary structure in CI2 are not independently stable, leading to highly cooperative formation of structure in the transition state (TS) for folding in which the helix is the best-formed region, stabilized by long-range interactions. No element of secondary structure is fully formed in the TS and its folding is kinetically two-state (3). It was hypothesized that multistate folding requires independently stable elements of structure or an unstable element of structure being stabilized by interacting with other elements. At the other extreme, the three-helix engrailed homeodomain, EnHD, is a small protein that folds ultrafast via a stable intermediate in line with the framework mechanism, whereby secondary structure forms first followed by docking of the helices (4, 5). Its folding intermediate consists of helix1 (H1) being helical but not docked on to the helix2-turn-helix3 (HTH) DNA binding motif (5, 6). The HTH motif is independently stable in the on-pathway intermediate (7). Studies of other members of the homoedomain superfamily show a slide from this multistate framework folding mechanism to two-state folding and nucleation condensation for members with less stable HTH motifs, consistent with our earlier hypothesis (8–10).

EnHD folds via an on-pathway intermediate in 1.7 μs, which rearranges in 17 μs to the native structure. The structure of the intermediate, which has been solved by NMR and molecular dynamics (MD) simulation, consists of the HTH motif in a native-like conformation with H1 being helical but undocked (5, 6). Such HTH motifs do not usually fold independently but that of EnHD is a marginally stable domain (7). Temperature-jump experiments using IR to monitor secondary structure formation and Trp fluorescence for tertiary interactions on wild-type protein and the sequence of residues 16–59 that encompass the HTH domain show conclusively that the 1.7-μs phase is for the formation of the HTH domain with concerted consolidation of its secondary and tertiary structures (7). Direct measurements of the motion of H1 by photoinduced electron transfer fluorescence-quenching correlation spectroscopy shows that the slower step of 17 μs represents the segmental motion of H1 as it docks onto the HTH motif (11). The folding is sufficiently fast that full atomistic MD simulations have been performed of the protein on similar time scales to those of experiment to fill in the folding pathway (4, 12).

The rate-determining TS for EnHD and other homeodomains have been characterized by Φ-analysis (4, 8); however, the Φ-values for formation of the EnHD folding intermediate have proved difficult to determine because the two phases are only separated by a factor of 10. We have found that Pit1, the 63-residue homeodomain from pituitary-specific transcription factor (13), folds in better separated phases of 2.3 and 46 μs with more convenient fluorescence changes from its endogenous Trp probe, Trp48, analogous to that in EnHD (14). The 2.3-μs phase is a partly diffusion-limited collapse of the denatured state to give a folding intermediate, I. The TS for the formation of the intermediate, TS1, has very little structure, and structure is formed cooperatively after the major TS, TS2, for formation of the native state.

Here, we have performed a detailed Φ-analysis of both TSs and also the folding intermediate of Pit1. These results have been combined with MD simulations and NMR experiments and they reveal an unexpected twist: The Pit1 intermediate is different from that of EnHD, illustrating a malleability of folding intermediates in this simple three-helix bundle family dictated by changes in stability.

Results

Stability of the HTH Motif. We synthesized the peptide spanning the HTH motif (residues 16–60) of Pit1 and recorded 1D NMR at different values of pH and ionic strength to determine if it formed a folded structure. In all cases, the spectra were those expected for an unstructured peptide, with the most upfield-shifted 1H resonance at 0.8 ppm (Fig. S1).

Rate Constants for Folding and Unfolding: Comparison of the Folding of Pit1 with EnHD and Its Fragments. EnHD and Pit1 both fold with biphasic kinetics, whose rate constants vary with urea concentration (Fig. 1). The fast phases for the formation of EnHD is slightly higher than that for Pit1, as are the single fast phases

Author contributions: W.B., T.L.R., V.D., and A.R.F. designed research; W.B., T.L.R., and R.D.S. performed research; W.B., T.L.R., R.D.S., V.D., and A.R.F. analyzed data; and W.B., V.D., and A.R.F. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: arf25@cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1101752108/-/DCSupplemental.

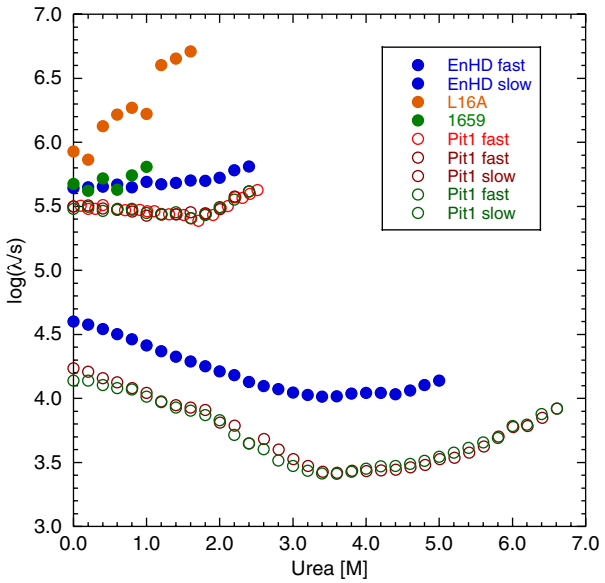

Fig. 1. Folding and unfolding kinetics of EnHD, the L16A EnHD mutant (L16A), the EnHD HTH motif (residues 16–59, 1659) and Pit1 (triplicate determinations of fast phase and duplicate of slow phase).

for the folding of the 16–59 HTH fragment of EnHD and that of the L16A EnHD mutant, which folds only as far as the HTH motif under these conditions (4–7). The single phases for these constructs increased with increasing concentrations of urea as they are both unstable and the unfolding limb of the chevron is manifested. The fast phases are flatter for EnHD and Pit1. The slower phase for Pit1 paralleled that of EnHD, but it is 2–3 times slower.

The rate constants for the fast and slow phases are sufficiently close together that they are coupled. Determination of the individual rate constants requires fitting the data in Fig. 2 simultaneously to the two equations that are the solutions to a general three-state kinetic scheme (15, 16):

$$\mathrm{U} \underset{k_{\mathrm{IU}}}{\overset{k_{\mathrm{UI}}}{\rightleftarrows}} \mathrm{I} \underset{k_{\mathrm{NI}}}{\overset{k_{\mathrm{IN}}}{\rightleftarrows}} \mathrm{N}$$

to give the solutions $\lambda_1 = 0.5[p + (p^2 - 4q)^{1/2}]$ and $\lambda_2 = 0.5[p\text{-}(p^2 - 4q)^{1/2}]$, where p and q are defined as $p = k_{\mathrm{UI}} + k_{\mathrm{IU}} + k_{\mathrm{IN}} + k_{\mathrm{NI}}$ and $q = k_{\mathrm{UI}}k_{\mathrm{IN}} + k_{\mathrm{IU}}k_{\mathrm{NI}} + k_{\mathrm{UI}}k_{\mathrm{NI}}$.

Conducting a complete Φ-value analysis of such a three-state reaction is both difficult and rare because it requires the deconvolution of a set of four rate constants for each mutant and their variation with denaturant concentration. Further, the required changes in the free energy of denaturation, $\Delta G_{\mathrm{U\text{-}N}}$, have to be calculated from these rate constants ($=\Delta G_{(\mathrm{U\text{-}N)kin}}$) because they cannot be measured easily from equilibrium experiments, unlike those for two-state reactions. Rare examples of such an analysis are the work on the Im7 protein (17) and a PDZ domain (18). We used our home-built T-jump apparatus to collect extensive data on wild-type Pit1 and 44 mutants (Fig. S2).

The denaturant dependence of the rate constant was modeled with the standard linear equation, $k = k_0 \exp(m[\mathrm{urea}]/RT)$, where k_0 is the value of rate constant in the absence of denaturant, m (either positive or negative) is the slope of the dependence of $\log k$ on urea concentration, R is gas constant, and T is temperature. Such individual plots—the "free fits"—generate data with considerable scatter in m values, which leads to high errors in k_{NI} extrapolated to 0 M denaturant (Table S1). To take advantage of the huge amount of kinetic data, we then fitted all of the data simultaneously to the two equations assuming that m_{UI}, m_{IU}, m_{IN}, and m_{NI}, the m values associated for k_{UI}, k_{IU}, k_{IN}, and k_{NI},

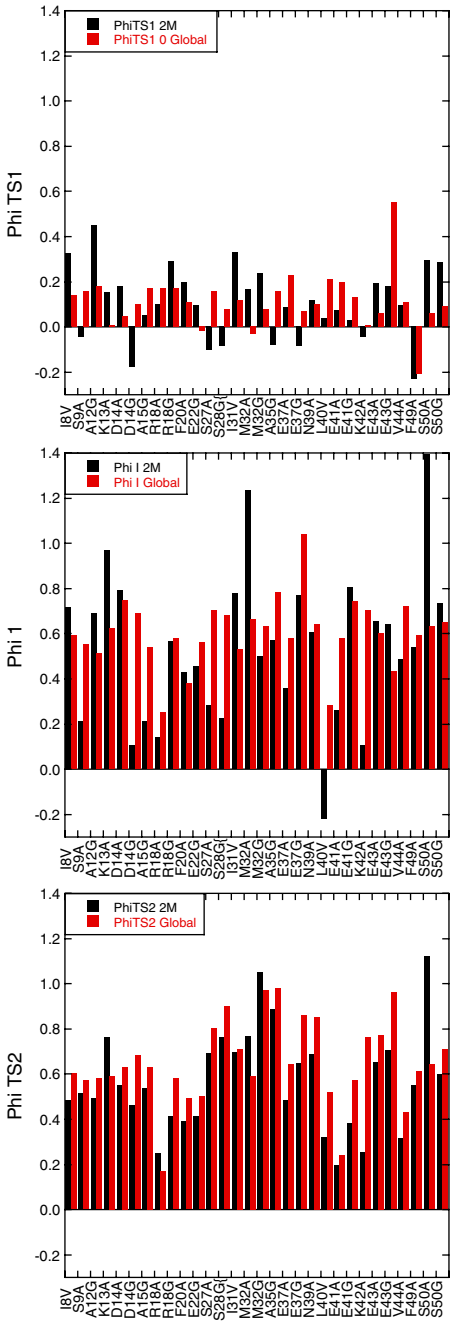

Fig. 2. Comparison of Φ-values calculated from global fitting of data at 0 M denaturant and free fitting of data at 2 M urea.

respectively, are the same for each mutant (Table S2). This assumption is reasonable: A plot of the individual values of m versus the $\Delta\Delta G_{\mathrm{D\text{-}N}}$ between wild type and mutant calculated from the rate constants extrapolated to 2 M urea (Fig. S3) indicated little systematic change of m with stability of the mutant. Extrapolation was to 2 M urea for the calculation of Φ to shorten the long extrapolations of $\log k_{\mathrm{NI}}$ and thereby lower errors. The value of $\Delta G_{\mathrm{D\text{-}N}}$ for wild type calculated from the global fit is 4.2 kcal mol^{-1}, and from the free fit, with a larger error, 4 kcal mol^{-1}. ^{2}H/H-exchange data previously gave a value of 4–5 kcal mol^{-1} (14).

Φ-Value Analysis of Transition and Intermediate States. The known crystal structure of Pit1 is from a complex with DNA (13). We checked the solution structure of the free protein by residual

dipolar couplings and chemical shift analysis. The correlation coefficient between the measured residual dipolar couplings and those calculated from the crystal structure of 2.3-Å resolution is 0.95 and the Q factor is 0.23, showing satisfactory agreement (19). The only difference is that H3 is shorter in solution, terminating at residue 52. In the following sequence, H1 spans residues 10–22, H2 spans 28–38, and H3 spans 42–52. There is a turn composed of residues 39–41 (NLE). The helical regions in the following sequence are in bold:

^{1}KRKRRTTIS^{10}**IAAKDALERH**^{20}**FGE**HSKPS**SQ**

× ^{30}**EIMRMAEEL**N^{40}LE**KEVVRVWF**^{50}**SNR**RQREKRV^{60}K.

Various mutations were introduced to probe different interactions (Table S3). In addition, we probed secondary structure for solvent-exposed side chains in helices by Ala → Gly scanning (20, 21). All mutants studied were folded, as shown from circular dichroism.

Calculation of Φ-Values. Φ_F for folding is defined as zero for U and one for the native N state, so that $\Phi_{TS1} = \Delta\Delta G_{U\text{-}TS1}/\Delta\Delta G_{U\text{-}N}$, $\Phi_I = \Delta\Delta G_{U\text{-}I}/\Delta\Delta G_{U\text{-}N}$, and $\Phi_{TS2} = \Delta\Delta G_{U\text{-}TS2}/\Delta\Delta G_{U\text{-}N}$. Systematic errors in the measurement and extrapolation of rate constants should be similar for wild type and mutant and so tend to cancel out in calculating Φ. We used the data from the global analysis to calculate Φ-values at 31 positions plus a series of eight secondary mutations of Ala → Gly in helices (Tables S4 and S5). To check that we were not biasing the results by global fitting, we compared their derived values with those from the free fits calculated at 2 M urea (Fig. 2). We discarded all values where $\Delta\Delta G$ in the denominator was <0.5 kcal/mol (22). The overall agreement was acceptable, given that there are large standard errors in the free-fitted data. The most reliable Φ-values from the free fits should be those for TS2 because they are, in effect, measured by dividing logarithms of ratios of rate constants for unfolding by those for the overall equilibria, and the extrapolation to 2 M urea is relatively short.

General Picture Emerging from the Pattern of Φ-Values. The folding pathway of Pit1 emerging from Φ-values suggests that TS1 has little internal structure (Fig. 2). The low values of β_T, = 0.14 (ref. 14 and here), reflects little compaction of solvent accessible surface area. With one or two exceptions, all measured values of Φ_{TS1} (Tables S4 and S5) were in the range 0–0.2 (mean = 0.11 ± 0.11 SD), showing very little native secondary or tertiary structure had been formed in TS1. The most significant consolidation of the structure took place between TS1 and the intermediate, where $\langle\Phi_I\rangle = 0.56 \pm 0.15$ SD. The presence of high compaction in I was corroborated by the significant m value for the fast unfolding rate constant (0.73 kcal mol^{-1} M^{-1}). The Φ_I-values suggest that the intermediate had all three helices significantly formed. They interacted with each other, and the hydrophobic core had started to consolidate. On progressing from the intermediate state to TS2, there was a slight consolidation with $\langle\Phi_{TS2}\rangle = 0.60 \pm 0.15$ (Tables S4). Furthermore, Ala scanning (Table S5) detected consolidation of secondary structure to near-native levels in H2 in TS2. H1 displayed little change in Φ. The final consolidation of the structure took place after crossing TS2.

Overview of MD Simulations of Pit1 Folding. The folding pathway of Pit1 was modeled via high-temperature MD simulations of the reverse reaction, unfolding. In two independent long (100 ns) 498 K simulations, the unfolding pathway consisted of a sharp transition (TS2) from the native state (N) to the intermediate (I). After lingering in the intermediate for ∼20 ns, the protein then unfolded through another broader transition (TS1) into the denatured state (D) (Fig. 3*A*). To facilitate comparison with experiment, we present the process in the opposite direction, i.e., in terms of folding. The denatured state was relatively compact but the helix content was low and the Cα rmsd from the starting structure was approximately 8 Å (Fig. 3 *A* and *B*). In moving from D to TS1, some fluctuating helical structure formed, but there was little consolidated, internal structure (Fig. 3*A*). In moving from TS1 to I, all three helices formed to some extent with the average helix content increasing to approximately 40% (Fig. 3 *A* and *B*). The structure consolidated but the hydrophobic core was only partially formed (more details regarding I are presented below). Nonetheless, the intermediate had persistent contacts within the HTH motif and fluctuating helical structure in H1. On moving from I to TS2, the helix content increased and the packing involving H3 improved (Fig. 3*A*), but H1 remained frayed and consolidation of the helices was incomplete. Finally, from TS2 to N, the H1 helix folded completely and there was rearrangement of the docking of the helices to break nonnative interactions and replace them with the proper packing interactions.

Properties of the MD-Generated Pit1 Intermediate State Ensemble. An intermediate was detected in the simulations, both by reaction coordinate analysis of property space (23) (Fig. 3*B*, *Right*) and clustering the all-versus-all Cα rmsd matrix (24). For example, Pit1 contained two persistent nonnative clusters: one representing the intermediate and the other the denatured state. The intermediate was treated as all time points between 0.9—22.55 ns, resulting in 21,650 structures in this ensemble. This intermediate had an average Cα rmsd to the native state of 5.6 ± 0.8 Å and substantial helical content (41 ± 10%, note that N has 62% helix). That said, the structure was fluid and the orientations of the helices were dynamic (see stereoview in Fig. 3*C*). The Φ-values and S values calculated over the MD-generated intermediate ensemble in good agreement (Fig. 3*D*). As with Φ-values, S values reflect the extent of structure of the residue of interest and generally take values from zero to one (24).

Despite the fluctuating nature of the helix packing, there were some persistent native contacts within the HTH motif (H2 and H3) and fluctuation of helical structure in H1 (Fig. 3*C*, *Left*). There were 90 distinct native contacts observed in I. The majority of native contacts were comprised of intrahelical backbone contacts (approximately 80 native contacts). The HTH motif averaged 49 ± 15 native contacts (of 84 native contacts in the HTH motif in the crystal structure) and 14 ± 5 nonnative contacts. Consequently, the loss of interhelical native contacts was countered somewhat by the gain of six nonnative interactions. There were three persistent native contacts between the C-terminal residues of H1 and the C-terminal residues of H3 (Fig. 3*E*, *Left*). Native contacts within the turn region were not maintained. Contacts between the HTH motif and H1 fluctuated throughout the intermediate. All native contacts between H2 and H1 had low populations (<20%). Altogether, these transactions led to an intermediate ensemble where the HTH motif was compact and maintained by a combination of native and nonnative contacts. H1 lost its contact to H2 and most of its contacts to H3, except near where they are joined by the loop. A persistent electrostatic network between the N terminus and the turn kept H1 near but not always in direct contact with the HTH motif. Although H1 was poorly folded compared with H2 and H3, it exerted an effect on the stability of the HTH motif, and when H1 pulled away from the other helices, the HTH motif fell apart (Fig. 3*A*).

In addition to the native contacts within the HTH motif, nonnative salt bridges stabilized I. Specifically, nonnative contacts between the N terminus and the turn region and the N-terminal region of H3 were highly populated in the intermediate (Fig. 3*E*, *Center*). A single native salt bridge, K3-E41, was populated 99.9% of the time in the intermediate state. It was joined by other non-

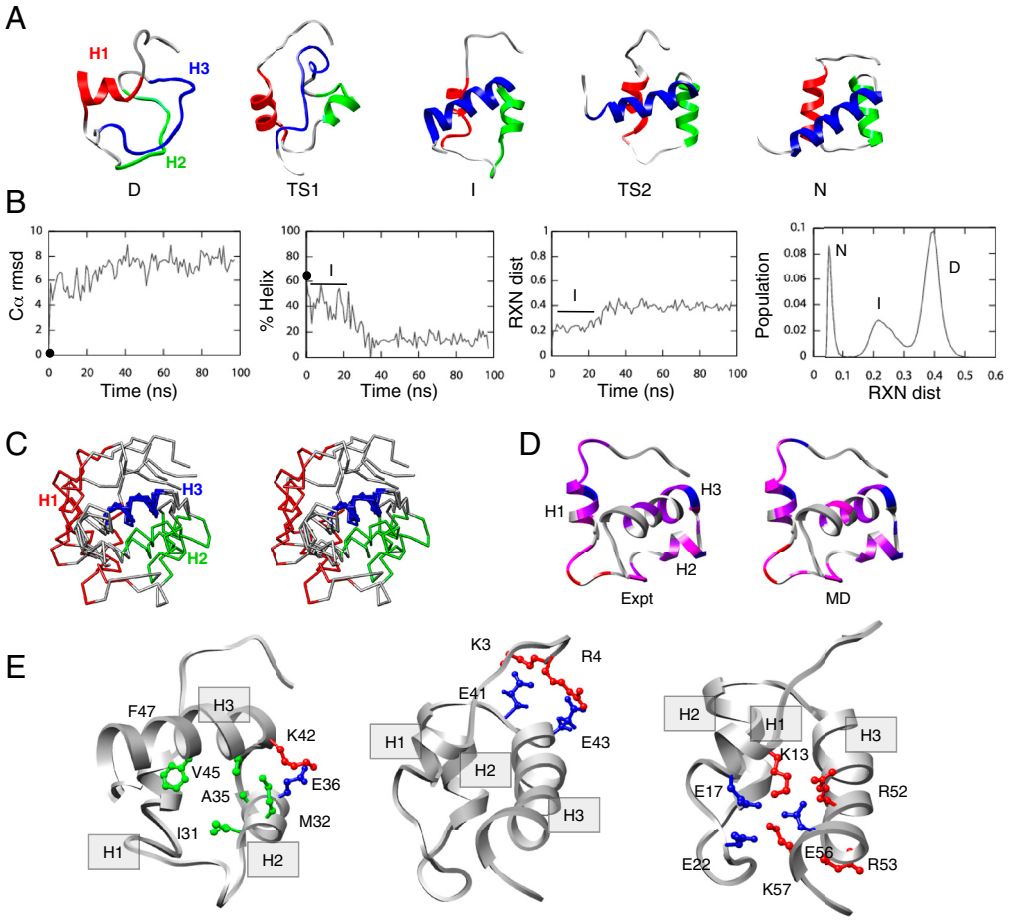

Fig. 3. MD simulations of the thermal unfolding of Pit1. (*A*) Representative structures from run 1 of MD depicted in the direction of folding: the denatured state (D, 97 ns), the initial transition state (TS1, 43.2 ns), the intermediate state (I, 20 ns), and the final transition state (TS2, 372 ps). The crystal structure is provided as the native (N) reference. The protein is colored as follows: H1, red, residues 10–22; H2, green, residues 28–37; and H3, blue, residues 42–50. (*B*) Various physical properties are plotted as a function of time (from left to right): the Cα rmsd (in angstrom) for run 1; the helix content (based on Φ,ψ angles) for the same simulation; reaction coordinate (RXN) distance from native reference over time, which is based on a property-space distance relative to 298-K control; and distributions of RXN distances to create a multidimensional-embedded one-dimensional reaction coordinate, showing breakdown into I and D ensembles. Although the presence of an intermediate is not evident in the rmsd plot, it is can be seen in the helical content and the property space reaction coordinate. The value of each property at the beginning of the simulation is indicated with a filled circle at 0 ns. (*C*) Stereoview of overlay of four structures from the intermediate state (5, 10, 15, and 20 ns) colored as in *A* and aligned on H3. (*D*) Comparison between Φ (expt) and *S* values (MD) for the intermediate state. The 20-ns MD-generated structure is colored from low degree of structure in red to highly structured in blue (red, Φ; *S*, 0–0.25; magenta, 0.25–0.5; purple, 0.5–0.75; blue, >0.75). (*E*) Residues participating in highly populated contacts in I (20-ns snapshot) from left to right: residues stabilizing the HTH motif, residues participating in H1-H3 salt bridges, and interactions between the N terminus and turn region of the HTH motif.

native salt bridges linking these same regions: R4-E41 (97% occupancy) and R4-E43 (95%). This electrostatic network kept I compact and stabilized the turn of the HTH motif.

It is worth noting that only the K3-E41 salt bridge was populated to any significant extent in the native 298-K simulation. That is, there were not any nonnative salt bridges in the N terminus to turn and C-terminal H3 region in the 298-K control. In addition to a network involving the N terminus and the turn and N-terminal H3 residues, there was a small electrostatic network near the C terminus of H1 and H3 in the intermediate state (Fig. 3*E*, *Right*). This network involved the native E17-R50 salt bridge of and several nonnative salt bridges (K13-R52, K13-E56, E17-K57, E22-R52, E22-R52, and E22-K57). This network was less stable than the former one involving the N terminus. Together these two networks stabilized both "ends" of the protein along the axis of H3. The "slack" from the N terminus' interactions allowed H1 to pull away from the HTH motif, while H2 and H3 remained closely aligned, although not in the precise native HTH geometry (Fig. 3*E*). With respect to unfolding, the secondary structure of Pit1 was fairly well maintained until the protein exited the intermediate state. As the protein unfolded further, H2 and H3 unraveled while H1 contained fluctuating helical structure for the duration of the 100-ns simulations.

Discussion

The folding kinetics of wild-type Pit1 resembles that of EnHD, which folds via a discrete intermediate in which the HTH motif is an autonomous, stable domain, whose formation and stability are uncoupled from that of H1 (14). But, a detailed Φ-value analysis shows clearly that formation of its intermediate has the formation of all three helices closely coupled. The peptide corresponding to the HTH motif of Pit1 is unstructured (Fig. S1), unlike that of EnHD, which forms a stable structure (7). MD simulations detected a Pit1 folding intermediate in which H1 is collapsed on to the rest of the molecule, unlike in the previous simulations on EnHD.

We previously noted that the mechanism of folding can slide from nucleation condensation to framework as the stability of individual domains increases (2, 8–10). For EnHD, the stability of the HTH motif is exceptionally high with the measured helical propensity of H1, H2, and H3 at approximately 30%, 15%, and 25% and predicted at approximately 70%, 30%, and 2% (4) by

the program Agadir (25). In contrast, the predicted values for Pit1 are approximately 8%, 8%, and 1% (Fig. S4). Close inspection of the amino acid frequency matrix from the Dodd and Egan (26) algorithm suggests that EnHD has optimal amino acids at key positions that would stabilize the HTH motif: hydrophobic leucines at positions 34, 38, 40; small glycine at position 39 to allow for unusual backbone dihedral angles; and hydrophobic amino acids at positions I45, W48, F49. The only highly conserved position that does not have the "optimal" amino acid in EnHD is S35, where alanine is expected. For Pit1, the positions M34 and N39 are suboptimal with good conservation at key positions for H3. Consequently, the predicted HTH motif propensity (26) is 5.88 for EnHD ("strong") and 3.88 for Pit1 ("medium") as opposed to 0 for c-Myb and hTRF1 (where the turn regions are extended) and which do not fold via a detectable intermediate (8). Interestingly, mutating c-Myb to enhance the predicted stability of the HTH motif leads to the detection of an intermediate (9). The HTH of EnHD is so stable that it forms an independently folded domain (7), but the HTH for Pit1 is not sufficiently stable and requires stabilization through interactions, many nonnative, with H1 (Fig. 3*A*). Along with misfolded intermediates found for EnHD (12), Im7 (27, 28), the PDZ (18) and FF domains (29, 30), stabilization by nonnative interactions is a general mechanism for the formation of intermediates.

The TSs for the folding of the homeodomain superfamily vary in detailed structure with changes in sequence and local stability, which has also been documented in other systems (20, 31–34). Transition states can be quite malleable (31) but one unifying feature is the extended nucleus model (35), whereby the TS for folding (or final TS for multistep reactions) resembles an expanded and distorted native state. The degree of distortion varies, but the overall topology is maintained because of the high cooperativity of folding.

Materials and Methods

Protein Expression. All the chemicals were analytical grade or superior. Unless otherwise stated, reagents were obtained from Sigma, BDH, or Fisher Scientific, whereas ultra-pure urea was purchased from Fluka. The plasmid pSEA100, encoding Pit1 gene under T7 promoter was used to express the protein. Cysteine at the position 50 was mutated to serine in order to prevent intermolecular disulfide bond formation, and the C50S mutant is referred to as WT. The protein was expressed by standard procedures and purified by using Q-Sepharose XL column (Pharmacia) followed by a Source 30S column (Amersham Biosciences) and the reverse-phase C18 monomeric column using Waters HPLC system. Purity and identity were checked by mass spectroscopy. The genes for mutants were prepared by using QuikChange (Stratagene), and the proteins were produced as described above. All proteins were >95% pure. All experiments were performed in 50 mM sodium acetate, 100 mM NaCl at pH 5.5, 25.0 °C unless otherwise stated.

Peptide Synthesis. The peptide spanning the region 16–60 of Pit1 was synthesized according to a solid-phase Fmoc/tBu strategy on a Fmoc-PAL-PEG-PS resin (Applied Biosystems). After deprotection and cleavage from the resin, it was purified by HPLC on a preparative C18 column with a gradient from 40–60% solvent B (solvent A, 0.1% TFA in water; solvent B, 0.1% TFA, 5% water in acetonitrile), and then lyophilized.

Kinetics. All samples for kinetic experiments were degassed and filtered. The range of concentration used was 100–200 μM. The methodology of kinetic experiments is described in ref. 7. All measurements were performed in 50 mM sodium acetate, 100 mM NaCl at pH 5.5, 25.0 °C buffer.

NMR Experiments. One-dimensional NMR data were recorded on Bruker DRX500 spectrometer at 25 °C using WATERGATE pulse sequence (36). All samples were prepared by dissolving lyophilized peptide in appropriate buffers containing 10% D_2O and 2,2-dimethyl-2-silapentane-5-sulfonate as an internal standard. Buffers used were 50 mM sodium phosphate pH 3.0, 50 mM sodium acetate (deuterated) pH 5.5, 50 mM sodium phosphate pH 7.0, and each repeated at 100 and 500 mM NaCl.

Data Analysis. Data analysis was performed using Kaleidagraph (Synergy Software), Prism (GraphPad Software), and R (R Development Core Team).

MD Simulations. Simulations of Pit1 were performed at 498 K to characterize the thermal unfolding pathway as well as at 298 K as a control. Residues 103–160 from chain B of Protien Data Bank ID 1AU7 (13) were used as the starting structure for all simulations. Missing side-chain density for Lys 125 was modeled in by selecting the lowest energy rotamer in the context of the protein from our dynamic rotamer library (37). The two histidines (H17 and H21) were neutral with protonation on the Nδ1 atom. Two independent 100-ns 498-K simulations were performed to sample each state along the unfolding pathway. Three additional 498-K simulations (each 5-ns long) were performed to further sample TS2 (the early TS in unfolding).

Simulations were performed using standard protocols (38, 39). The minimized starting structure was solvated in a rectangular water box extending at least 10 Å from any protein atom, with the water density set to the experimental value (0.829 g/mL at 498 K and 0.999 g/mL at 298 K) (40). A flexible three-center water model (41) was used along with periodic boundary conditions to reduce edge effects. Simulations were performed using the Levitt et al. force field (42) and the microcanonical (number of molecules, volume, and total energy held fixed) ensemble using the *in lucem* molecular mechanics modeling package (43). In all 498-K simulations, an 8-Å force-shifted nonbonded cutoff was used and the nonbonded list was updated every two steps; a 10-Å nonbonded cutoff was used at 298 K. Nonbonded interactions between charge groups separated by three bonds were included and scaled to 0.4 (44). A 2-fs time step was used, and structures were saved 0.2–1 ps for analysis.

ACKNOWLEDGMENTS. V.D. is grateful for support from the National Institutes of Health (GM50789).

1. Otzen DE, Itzhaki LS, elMasry NF, Jackson SE, Fersht AR (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 USA* 91:10422–10425.
2. Itzhaki LS, Otzen DE, Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 254:260–288.
3. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor 2.1. Evidence for a two-state transition. *Biochemistry* 30:10428–10435.
4. Mayor U, et al. (2003) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* 421:863–867.
5. Mayor U, Johnson CM, Daggett V, Fersht AR (2000) Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc Natl Acad Sci USA* 97:13518–13522.
6. Religa TL, Markson JS, Mayor U, Freund SM, Fersht AR (2005) Solution structure of a protein denatured state and folding intermediate. *Nature* 437:1053–1056.
7. Religa TL, et al. (2007) The helix-turn-helix motif as an ultrafast independently folding domain: The pathway of folding of engrailed homeodomain. *Proc Natl Acad Sci USA* 104:9272–9277.
8. Gianni S, et al. (2003) Unifying features in protein-folding mechanisms. *Proc Natl Acad Sci USA* 100:13286–13291.
9. White GW, et al. (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. Fersht AR, Daggett V (2002) Protein folding and unfolding at atomic resolution. *Cell* 108:573–582.
11. Neuweiler H, Banachewicz W, Fersht AR (2010) Kinetics of chain motions within a protein folding intermediate. *Proc Natl Acad Sci USA* 107:22106–22110.
12. McCully ME, Beck DA, Fersht AR, Daggett V (2010) Refolding the engrailed homeodomain: Structural basis for the accumulation of a folding intermediate. *Biophys J* 99:1628–1636.
13. Jacobson EM, Li P, Leon-del-Rio A, Rosenfeld MG, Aggarwal AK (1997) Structure of Pit-1 POU domain bound to DNA as a dimer: Unexpected arrangement and flexibility. *Genes Dev* 11:198–212.
14. Banachewicz W, Johnson CM, Fersht AR (2010) Folding of the Pit1 homeodomain near the speed limit. *Proc Natl Acad Sci USA* 108:569–573.
15. Ikai A, Tanford C (1973) Kinetics of unfolding and refolding of proteins. I. Mathematical analysis. *J Mol Biol* 73:145–163.
16. Gianni S, Ivarsson Y, Jemth P, Brunori M, Travaglini-Allocatelli C (2007) Identification and characterization of protein folding intermediates. *Biophys Chem* 128:105–113.
17. Friel CT, Smith DA, Vendruscolo M, Gsponer J, Radford SE (2009) The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat Struct Mol Biol* 16:318–324.
18. Gianni S, Ivarsson Y, Travaglini-Allocatelli C, Brunori M, Vendruscolo M (2010) Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nat Struct Mol Biol* 17:1431–1437.
19. Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci* 12:1–16.

20. Matthews JM, Fersht AR (1995) Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: Observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for structural elements for unfolding/folding of barnase. *Biochemistry* 34:6805–6814.
21. Scott KA, Alonso DO, Sato S, Fersht AR, Daggett V (2007) Conformational entropy of alanine versus glycine in protein denatured states. *Proc Natl Acad Sci USA* 104:2661–2666.
22. Fersht AR, Sato S (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc Natl Acad Sci USA* 101:7976–7981.
23. Toofanny RD, Jonsson AL, Daggett V (2010) A comprehensive multidimensional-embedded, one-dimensional reaction coordinate for protein unfolding/folding. *Biophys J* 98:2671–2681.
24. Daggett V, Li A, Itzhaki LS, Otzen DE, Fersht AR (1996) Structure of the transition state for folding of a protein derived from experiment and simulation. *J Mol Biol* 257:430–440.
25. Munoz V, Serrano L (1997) Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* 41:495–509.
26. Dodd IB, Egan JB (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acids Res* 18:5019–5026.
27. Whittaker SB, Spence GR, Gunter Grossmann J, Radford SE, Moore GR (2007) NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J Mol Biol* 366:1001–1015.
28. Gsponer J, et al. (2006) Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc Natl Acad Sci USA* 103:99–104.
29. Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE (2010) A transient and low-populated protein-folding intermediate at atomic resolution. *Science* 329:1312–1316.
30. Jemth P, et al. (2004) Demonstration of a low energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. *Proc Natl Acad Sci USA* 101:6450–6455.
31. Lindberg MO, Oliveberg M (2007) Malleability of protein folding pathways: A simple reason for complex behaviour. *Curr Opin Struct Biol* 17:21–29.
32. Matouschek A, Otzen DE, Itzhaki LS, Jackson SE, Fersht AR (1995) Movement of the position of the transition state in protein folding. *Biochemistry* 34:13656–13662.
33. Shen T, Hofmann CP, Oliveberg M, Wolynes PG (2005) Scanning malleable transition state ensembles: Comparing theory and experiment for folding protein U1A. *Biochemistry* 44:6433–6439.
34. Daggett V, Li A, Fersht AR (1998) A combined molecular dynamics and Φ-value analysis of structure-reactivity relationships in the transition state and unfolding pathway of barnase: The structural basis of Hammond and anti-Hammond effects. *J Am Chem Soc* 120:12740–12754.
35. Fersht AR (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. *Proc Natl Acad Sci USA* 97:1525–1529.
36. Piotto M, Saudek V, Sklenar V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661–665.
37. Scouras AD, Daggett V (2011) The dynameomics rotamer library: Amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. *Prot Sci* 20:341–352.
38. Beck DA, Armen RS, Daggett V (2005) Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. *Biochemistry* 44:609–616.
39. Beck DA, Daggett V (2004) Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods* 34:112–120.
40. Kell GS (1967) Precise Representation of volume properties of water at one atmosphere. *J Chem Eng Data* 12:66–69.
41. Levitt M, Hirshberg M, Sharon R, Daggett V (1997) Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J Phys Chem B* 101:5051–5061.
42. Levitt M, Hirshberg M, Sharon R, Daggett V (1995) Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Comput Phys Commun* 91:215–231.
43. Beck DAC, Alonso DOV, Daggett V (2000-2011) *In lucem molecular mechanics (ilmm)* (University of Washington, Seattle).
44. Armen RS, Bernard BM, Day R, Alonso DO, Daggett V (2005) Characterization of a possible amyloidogenic precursor in glutamine-repeat neurodegenerative diseases. *Proc Natl Acad Sci USA* 102:13433–13438.