J|A|C|S
ARTICLES

Published on Web 04/29/2009

Engineered Symmetric Connectivity of Secondary Structure Elements Highlights Malleability of Protein Folding Pathways

Ylva Ivarsson, Carlo Travaglini-Allocatelli, Maurizio Brunori,* and Stefano Gianni

Istituto Pasteur - Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Università di Roma "La Sapienza", Piazzale A. Moro 5, 00185 Rome, Italy

Received January 20, 2009; E-mail: maurizio.brunori@uniroma1.it

Abstract: To understand the role of sequence connectivity in protein folding pathways, we explored by Φ-value analysis the folding pathway of an engineered circularly permuted PDZ domain. This variant has the same sequence connectivity as naturally occurring circularly permuted PDZ domains and displays a symmetrical distribution of secondary structure elements (i.e., $\beta\beta\alpha\beta\beta\alpha\beta\beta$) while maintaining the same tertiary interactions of the well-characterized second PDZ domain from PTP-BL (PDZ2). Reliable Φ values were obtained for both a low-energy intermediate and the late rate-limiting transition state, allowing a description of both early and late events in folding. A comparison with Φ values obtained for wild-type PDZ2 reveals that while the structure of the late transition state is robust and unaffected by circular permutation, the folding intermediate is stabilized by a different nucleus involving residues located at the new N- and C-termini. The results suggest that folding is driven by competing nuclei whose stabilities may be selectively tuned by circular permutation.

Introduction

A powerful strategy for understanding the relationships between sequence information and folding pathways is to study the folding mechanisms of proteins displaying different amino acid compositions but similar folding. Structurally homologous proteins often share a conserved kinetic folding mechanism[1,2] and conserved folding nuclei,[3−9] suggesting that native-state topology is one of the main factors controlling protein folding.[10] The idea that folding pathways are governed by native topology has been challenged by ingenious experimental work using circularly permuted proteins. Early work on engineered circularly permuted variants of the β protein SH3 demonstrated the malleability of the folding process, i.e., that the same native fold may be reached via different transition states upon perturbation of the sequence connectivity if alternative nucleation motifs are available.[11] In contrast, a similar folding nucleus was observed when comparing a circularly permuted variant of chymotrypsin inhibitor 2 with the wild-type protein.[12] More recently, Oliveberg and co-workers[13−16] carried out ambitious studies of the ribosomal protein S6 by systematically altering the sequence connectivity between the six secondary structure elements (two α-helices and four β-strands). They observed that S6 displays a minimal folding nucleus composed of a two-strands-one-helix motif.[14] The precise composition of this motif varied with alterations of the linear arrangement of the secondary structure elements,[13] suggesting that circular permutations tuned the balance between competing nuclei involving the same helix but the docking of different strands.[15]

The postsynaptic density-95/discs large/zonula occludens-1 (PDZ) family displays a typical fold of a six-stranded β-sandwich flanked by two α-helices[17] arranged asymmetrically (β1-β2-β3-α1-β4-β5-α2-β6) (Figure 1). The PDZ domains from bacteria,[18] chloroplasts,[19] and mitochondria[20] may be considered

(1) Chi, C. N.; Gianni, S.; Calosci, N.; Travaglini-Allocatelli, C.; Engstrom, Å.; Jemth, P. *FEBS Lett.* **2007**, *581*, 1109–1113.
(2) Travaglini-Allocatelli, C.; Gianni, S.; Brunori, M. *Trends Biochem. Sci.* **2004**, *29*, 535–541.
(3) Friel, C. T.; Capaldi, A. P.; Radford, S. E. *J. Mol. Biol.* **2003**, *326*, 293–305.
(4) Gianni, S.; Guydosh, N. R.; Khan, F.; Caldas, T. D.; Mayor, U.; White, G. W.; DeMarco, M. L.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13286–13291.
(5) Zarrine-Afsar, A.; Larson, S. M.; Davidson, A. R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 42–49.
(6) Clarke, J.; Cota, E.; Fowler, S. B.; Hamill, S. J. *Structure* **1999**, *7*, 1145–1153.
(7) Chiti, F.; Taddei, N.; White, P. M.; Bucciantini, M.; Magherini, F.; Stefani, M.; Dobson, C. M. *Nat. Struct. Biol.* **1999**, *6*, 1005–1009.
(8) Martínez, J. C.; Serrano, L. *Nat. Struct. Biol.* **1999**, *6*, 1010–1016.
(9) Riddle, D. S.; Grantcharova, V. P.; Santiago, J. V.; Alm, E.; Ruczinski, I.; Baker, D. *Nat. Struct. Biol.* **1999**, *6*, 1016–1024.
(10) Baker, D. *Nature* **2000**, *405*, 39–42.
(11) Viguera, A. R.; Serrano, L.; Wilmanns, M. *Nat. Struct. Biol.* **1996**, *3*, 874–880.
(12) Otzen, D. E.; Fersht, A. R. *Biochemistry* **1998**, *37*, 8139–8146.
(13) Hubner, I. A.; Lindberg, M.; Haglund, E.; Oliveberg, M.; Shakhnovich, E. I. *J. Mol. Biol.* **2006**, *359*, 1075–1085.
(14) Lindberg, M.; Tangrot, J.; Oliveberg, M. *Nat. Struct. Biol.* **2002**, *9*, 818–822.
(15) Lindberg, M. O.; Oliveberg, M. *Curr. Opin. Struct. Biol.* **2007**, *17*, 21–29.
(16) Lindberg, M. O.; Tångrot, J.; Otzen, D. E.; Dolgikh, D. A.; Finkelstein, A. V.; Oliveberg, M. *J. Mol. Biol.* **2001**, *314*, 891–900.
(17) Morais Cabral, J. H.; Petosa, C.; Sutcliffe, M. J.; Raza, S.; Byron, O.; Poy, F.; Marfatia, S. M.; Chishti, A. H.; Liddington, R. C. *Nature* **1996**, *382*, 649–652.
(18) Krojer, T.; Garrido-Franco, M.; Huber, R.; Ehrmann, M.; Clausen, T. *Nature* **2002**, *416*, 455–459.
(19) Liao, D. I.; Qian, J.; Chisholm, D. A.; Jordan, D. B.; Diner, B. A. *Nat. Struct. Biol.* **2000**, *7*, 749–753.

10.1021/ja900438b CCC: $40.75

Figure 1. Structure of PDZ2 (PDB entry 1GM1), with the altered sequence connectivity of cpPDZ2 indicated by dotted lines. The numbering of the structural elements is according to the wild-type structure. Positions probed by mutagenesis in the current study are indicated in the stick representation.

as circularly permuted variants of their metazoan counterparts, since they share the same overall fold but have the N- and C-termini located at different positions along the sequence. Hence, this protein family represents an ideal system for investigating the relationships between sequence connectivity and protein topology.

Canonical PDZ domains fold via a conserved three-state mechanism, proceeding through two sequential transition states (TS1 and TS2) and an intervening high-energy intermediate.[1,21−24] The structural features of TS1 and TS2 have been addressed by Φ-value analysis and restrained molecular dynamic simulations of two members of the PDZ domain family, namely, the second PDZ repeat of murine protein tyrosine phosphatase BL (PDZ2) and the third PDZ domain from PSD-95 (PDZ3).[22,25] In the case of PDZ2, TS1 is stabilized by a three-stranded β-sheet (β1-β4-β6) docked with the major α-helix 2,[22] suggesting that as observed for other proteins,[26] interactions between the N- and C-terminal strands are crucial for productive folding. On the other hand, the structure of the late transition state TS2 resembles a distorted version of the native state, with much secondary structure formed and a quasi-nativelike topology. While the TS2's for PDZ2 and PDZ3 are nearly superimposable, TS1 has very different structural features in the two proteins, suggesting that native bias is weak at the early stages of folding but exerts a strict control on the folding mechanism close to the native state.[25]

We have recently characterized the folding kinetics of both engineered and naturally occurring circularly permuted PDZ domains displaying similar sequence connectivites.[27,28] In particular, in the engineered variant, we shifted the N-terminal strand of PDZ2 to the C-terminus, producing a circularly permuted PDZ2 (designated as cpPDZ2) with the linear connectivity β2-β3-α1-β4-β5-α2-β6-β1 and examined its effect on the folding kinetics.[28] This organization of secondary structure elements parallels that observed in naturally occurring circularly permuted PDZ domains, such as D1pPDZ (the PDZ domain from the D1 C-terminal processing protease of *Scenedesmus obliquus*). D1pPDZ populates a folding intermediate that was shown to be an off-pathway trap.[27] This observation contrasted with the behavior of canonical PDZ domains, for which the high-energy intermediate represents an on-pathway productive species. As reported below, by comparing the folding kinetics of wild-type PDZ2 and cpPDZ2, we conclude that circular permutation accelerates the early events of folding and stabilizes an intermediate but destabilizes the native state. Unfortunately, in the case of cpPDZ2, the (un)folding kinetics of the intermediate and native states are uncoupled, making it impossible to ascertain whether the intermediate is an on- or off-pathway species. In this work, we have unveiled the structural details of the folding pathway of cpPDZ2 by Φ-value analysis and probed not only the transition-state structure but also the interactions formed in the intermediate state. The data reveal that the early events of PDZ2 folding are altered by the circular permutation, with the interactions formed in the intermediate largely being contacts within the new N- and C-termini. In contrast, the late folding events, as probed by comparing the TS2 structures, remain basically unaffected. The implications of these results, in light of the previous work on circularly permuted proteins, are discussed below.

Materials and Methods

Protein Engineering and Purification. The gene encoding for cpPDZ2 was obtained from GENEART and subcloned into the pET28c (+) expression vector, as described in ref 28. The circular permutant was designed on the basis of a multiple structural alignment of different PDZ domains present in the PDB database. This procedure was facilitated by the presence of naturally occurring circularly permuted PDZ domains. In particular, cpPDZ2 was designed by cleaving between D22 and G23 (numbered according to the native PDZ2 structure, PDB entry 1GM1) and linking residues G99 and G11 together. In fact, D22 and G23 occur in the structurally variable solvent-accessible loop between the structurally conserved regions including β-strands 2 and 3 (β1 and β2 in cpPDZ2). Cleavage within this loop is therefore not expected to significantly affect the PDZ2 structure. In the structure of the PDZ domain from DegS (PDB entry 1TE0), a natural circular permutant, residues D338 and D339 (structurally equivalent to G99 and G11, respectively, in PDZ2) are linked together, and the main-chain conformation of all structurally equivalent secondary structure elements between the two domains are conserved. As documented in ref 28, the permissibility of the circular permutation was tested by performing ligand-binding experiments. Site-specific mutants, numbered according the wild-type PDZ2 sequence 1GM1, were obtained using the QuikChange site-directed mutagenesis kit (Stratagene). All of the mutations were confirmed by DNA sequencing, and the DNA was transformed into the *Escherichia coli* strain BL21(DE3) (Invitrogen) for protein expression. Protein expression was induced with 1 mM IPTG (OD_{600} = 0.6); after induction, cells were grown for 24 h at 20 °C. The resulting hexahis-

(20) Zhang, Y.; Appleton, B. A.; Wu, P.; Wiesmann, C.; Sidhu, S. S. *Protein Sci.* **2007**, *16*, 1738–1750.
(21) Gianni, S.; Calosci, N.; Aelen, J. M.; Vuister, G. W.; Brunori, M.; Travaglini-Allocatelli, C. *Protein Eng., Des. Sel.* **2005**, *18*, 389–395.
(22) Gianni, S.; Geierhaas, C. D.; Calosci, N.; Jemth, P.; Vuister, G. W.; Travaglini-Allocatelli, C.; Vendruscolo, M.; Brunori, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 128–133.
(23) Ivarsson, Y.; Travaglini-Allocatelli, C.; Jemth, P.; Malatesta, F.; Brunori, M.; Gianni, S. *J. Biol. Chem.* **2007**, *282*, 8568–8572.
(24) Jemth, P.; Gianni, S. *Biochemistry* **2007**, *46*, 8701–8708.
(25) Calosci, N.; Chi, C. N.; Richter, B.; Camilloni, C.; Engstrom, Å.; Eklund, L.; Travaglini-Allocatelli, C.; Gianni, S.; Vendruscolo, M.; Jemth, P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19241–19246.
(26) Krishna, M. M.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1053–1058.
(27) Ivarsson, Y.; Travaglini-Allocatelli, C.; Brunori, M.; Gianni, S. *J. Biol. Chem.* **2008**, *283*, 8954–8960.
(28) Ivarsson, Y.; Travaglini-Allocatelli, C.; Morea, V.; Brunori, M.; Gianni, S. *Prot. Eng., Des. Sel.* **2008**, *21*, 155–160.

tagged proteins formed inclusion bodies, which were resuspended in 50 mM sodium phosphate (pH 6.3) at a saturating concentration of urea. The protein solution was loaded on a S-Sepharose ion exchange column (Amersham Biosciences), washed with 50 mM sodium phosphate (pH 6.3), and eluted using a gradient of 0−2 M sodium chloride.

(Un)Folding Kinetics. Kinetic (un)folding experiments were performed using a Pi-star stopped-flow apparatus (Applied Photophysics, Leatherhead, U.K.) in 50 mM sodium phosphate (pH 7.2) at 25 °C in the presence of 0.4 M sodium sulfate. The engineered tryptophane (Y43W) was excited at 280 nm, and the (un)folding reaction was followed by the change in intrinsic fluorescence emission using a 320 nm cutoff filter. An 11-fold dilution of native or denatured protein in the appropriate buffer initiated unfolding or refolding.

Peptide-induced refolding experiments were carried out under the same conditions as above but by mixing the denatured protein against a urea buffer solution (0−4.5 M urea, depending on the mutant used) containing different concentrations of the peptide EQVSAV (50−8000 μM). Binding of the peptide to the natively folded molecules drains the folding equilibrium toward the native state, even under highly denaturing conditions.[23]

As described previously,[29,30] continuous-flow experiments were carried out on an in-house-built capillary mixer of design and methodology similar to that published by Shastry and Roder.[31]

Data Analysis. In a two-state mechanism, the dependence of the rate constants for folding (k_f) and unfolding (k_u) on denaturant (urea) concentration is defined by eqs 1 and 2, respectively:

$$k_f = k_f^{H_2O} \exp(-m_{D-TS}[\text{urea}]/RT) \quad (1)$$

$$k_u = k_u^{H_2O} \exp(m_{TS-N}[\text{urea}]/RT) \quad (2)$$

The parameters m_{D-TS} and m_{TS-N} are the slopes of the dependence of log k_f and log k_u, respectively, on urea concentration. The observed rate constant k_{obs} is the sum of k_f and k_u.

As described in the Results and in ref 28, the folding mechanism of cpPDZ2 involves the accumulation of an intermediate. Furthermore, since the fast and slow refolding phases are kinetically uncoupled, we fitted the data by assuming a fast pre-equilibrium between the intermediate and denatured states. Under such conditions, the observed kinetics are described by the following equation:

$$k_{obs} = k_u + \frac{k_f}{1 + \frac{1}{K_{D-I}}} \quad (3)$$

where K_{D-I} is the equilibrium constant for the rapid equilibrium between the denatured and intermediate states. The linear dependence of K_{D-I} on the urea concentration is described by

$$K_{D-I} = K_{D-I}^{H_2O} \exp(m_{D-I}[\text{urea}]/RT) \quad (4)$$

where m_{D-I} is the associated m value.

The dependence of k_{obs} on urea concentration for a three-state folder obtained from the classical (un)folding experiments is thus described by eq 5:

$$k_{obs} = k_u^{H_2O} \exp(m_{TS-N}[\text{urea}]/RT) + \frac{k_f^{H_2O} \exp(-m_{I-TS}[\text{urea}]/RT)}{1 + \frac{1}{K_{D-I}^{H_2O} \exp(m_{D-I}[\text{urea}]/RT)}} \quad (5)$$

It may be noted that such an analysis cannot distinguish between an on-pathway productive intermediate and an off-pathway kinetic trap. In fact, while eq 5 may be derived by assuming the presence of an on-pathway intermediate,[32] a model implying an off-pathway species would return a nearly identical solution. A more complete comparison between the two models may be found in refs 33 and 34.

The ligand-induced refolding experiments provide refolding rate constants independently of the unfolding rate constants, and these data are fitted to eq 5 with the unfolding part omitted. The data from the conventional (un)folding experiments and the ligand-induced refolding experiments were globally fitted with shared k_f and K_{D-I} values and with the m values kept constant and assumed to be equal to those of cpPDZ2 ($m_{I-TS} = 0.44$, $m_{TS-N} = 0.23$, and $m_{D-I} = 0.48$; values obtained from ref 28).

For the late transition state, Φ values were calculated using the expression

$$\Phi_{TS} = \frac{\Delta\Delta G_{D-TS}}{\Delta\Delta G_{D-N}} = 1 - \frac{\Delta\Delta G_{TS-N}}{\Delta\Delta G_{D-N}} \quad (6)$$

where

$$\Delta\Delta G_{TS-N} = \Delta G_{TS-N}^{mut} - \Delta G_{TS-N}^{WT} = RT \ln\left(\frac{k_u^{mut}}{k_u^{WT}}\right) \quad (7)$$

and $\Delta\Delta G_{D-N}$ is the difference between the free energies of unfolding for the denatured and native states upon mutation:

$$\Delta\Delta G_{D-N} = \Delta G_{D-N}^{mut} - \Delta G_{D-N}^{WT} \quad (8)$$

where

$$\Delta G_{D-N} = \Delta G_{D-I} + \Delta G_{I-N} = RT \ln\left[\frac{(k_f/k_u)}{K_{D-I}}\right] \quad (9)$$

For the intermediate, Φ values were calculated using the equation

$$\Phi_I = \frac{\Delta\Delta G_{D-I}}{\Delta\Delta G_{D-N}} \quad (10)$$

where

$$\Delta\Delta G_{D-I} = \Delta G_{D-I}^{mut} - \Delta G_{D-I}^{WT} = -RT \ln\left(\frac{K_{D-I}^{WT}}{K_{D-I}^{mut}}\right) \quad (11)$$

In all cases, kinetic $\Delta\Delta G_{D-N}$ were used for the calculations.

The Tanford β values (β_T) for the intermediate state and the late transition state were calculated using the following formulas:

$$\beta_T^I = \frac{m_{D-I}}{m_{D-N}} \quad (12)$$

$$\beta_T^{TS} = \frac{m_{D-TS}}{m_{D-N}} = 1 - \frac{m_{TS-N}}{m_{D-N}} \quad (13)$$

Results

To test how the protein folding free-energy landscape is shaped by the connectivity of the secondary structure elements, we performed a Φ-value analysis on an engineered circularly permuted variant of PDZ2 and compared the results with those previously obtained for the wild type. The circularly permuted PDZ2 was engineered so the N-terminal strand was shifted to

(29) Gianni, S.; Ivarsson, Y.; Bah, A.; Bush-Pelc, L. A.; Di Cera, E. *Biophys. Chem.* **2007**, *131*, 111–114.
(30) Gianni, S.; Walma, T.; Arcovito, A.; Calosci, N.; Bellelli, A.; Engström, Å.; Travaglini-Allocatelli, C.; Brunori, M.; Jemth, P.; Vuister, G. W. *Structure* **2006**, *14*, 1801–1809.
(31) Shastry, M. C.; Roder, H. *Nat. Struct. Biol.* **1998**, *5*, 385–392.
(32) Parker, M. J.; Spencer, J.; Clarke, A. R. *J. Mol. Biol.* **1995**, *253*, 771–786.
(33) Capaldi, A. P.; Shastry, M. C.; Kleanthous, C.; Roder, H.; Radford, S. E. *Nat. Struct. Biol.* **2001**, *8*, 68–72.
(34) Jemth, P.; Gianni, S.; Day, R.; Li, B.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6450–6455.

Table 1. Kinetic Folding Parameters for cpPDZ2 and Site-Directed Variants

	k_f (s^{-1})	k_u (s^{-1})	K_{D-I}	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})	Φ_I	Φ_{TS}	Φ_{TS2}[a]
cpPDZ2	122 ± 4	5.3 ± 0.2	0.032 ± 0.005	–	–	–	
V16A	150 ± 40	41.6 ± 0.9	0.6 ± 0.3	2.8 ± 0.3	0.6 ± 0.1	0.57 ± 0.07	0.62 ± 0.05
L25A	110 ± 10	44 ± 2	0.08 ± 0.05	1.9 ± 0.4	0.3 ± 0.2	0.33 ± 0.07	0.3 ± 0.1
I27V	94 ± 5	28.1 ± 0.6	0.09 ± 0.01	1.7 ± 0.1	0.34 ± 0.08	0.43 ± 0.04	0.43 ± 0.04
I42V	130 ± 20	11.9 ± 0.5	0.19 ± 0.06	1.5 ± 0.2	0.7 ± 0.2	0.7 ± 0.1	0.61 ± 0.06
V44A	122 ± 4	11.4 ± 0.3	0.062 ± 0.007	0.8 ± 0.1	0.5 ± 0.1	0.46 ± 0.07	0.39 ± 0.03
A46G	160 ± 20	13.8 ± 0.4	0.18 ± 0.04	1.4 ± 0.2	0.7 ± 0.1	0.6 ± 0.1	0.67 ± 0.05
A52G	150 ± 10	46 ± 2	0.06 ± 0.02	1.5 ± 0.3	0.2 ± 0.1	0.17 ± 0.03	0.20 ± 0.03
A53G	210 ± 20	30 ± 1	0.07 ± 0.03	1.2 ± 0.3	0.4 ± 0.2	0.13 ± 0.03	0.12 ± 0.03
I59V	83 ± 3	14.3 ± 0.3	0.019 ± 0.008	0.5 ± 0.3	n.d.[b]	n.d.[b]	0.41 ± 0.06
L66A	11 ± 1	17.4 ± 0.4	0.09 ± 0.04	2.7 ± 0.3	0.2 ± 0.1	0.74 ± 0.08	0.81 ± 0.07
A81G	21 ± 2	12.2 ± 0.2	0.08 ± 0.02	2.0 ± 0.2	0.25 ± 0.09	0.76 ± 0.08	0.91 ± 0.04
V92A	37 ± 3	9.8 ± 0.3	0.08 ± 0.02	1.6 ± 0.2	0.35 ± 0.09	0.78 ± 0.09	0.72 ± 0.09

[a] Values for PDZ2 from ref 22. [b] Not determined because the change in stability induced by mutation was too small to calculate a reliable Φ value.

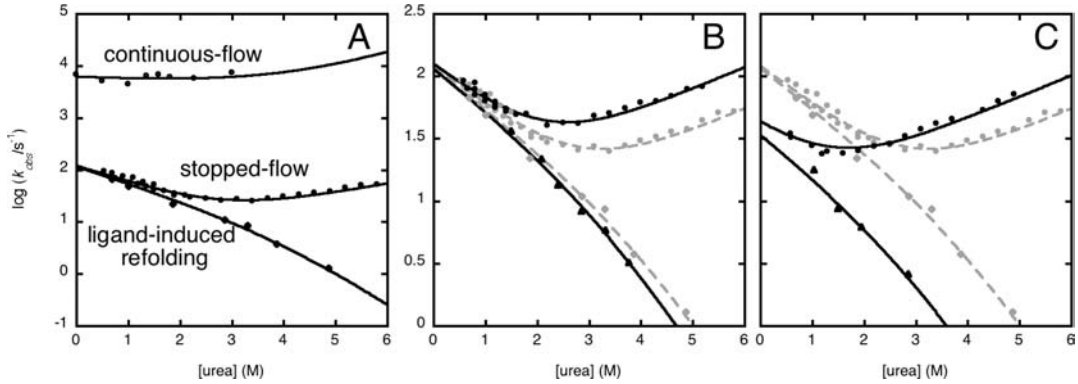

Figure 2. Chevron plots for cpPDZ2 and the representative site-directed mutants. (A) Folding kinetics of cpPDZ2 were obtained by continuous-flow (top), stopped-flow (middle), and ligand-induced refolding (bottom) experiments. Ligand-induced refolding experiments were performed by mixing denatured cpPDZ2 with different concentrations of ligand, allowing direct measurement of the folding rate constants as discussed in refs 23 and 36. As described previously,[29,30] continuous-flow experiments were carried out on an in-house-built capillary mixer of design and methodology similar to that published by Shastry and Roder.[31] The folding of cpPDZ2 involves a rapid pre-equilibrium between the denatured and intermediate states, as demonstrated by the ultrafast continuous-flow kinetics (top curve). The data were globally fitted to the solution for a three-state mechanism. (B, C) Chevron plots for the two cpPDZ2 mutants (B) V44A and (C) V92A, as obtained by stopped-flow (black ●) and ligand-induced refolding (black ▲) experiments. The corresponding data and fits for wild-type cpPDZ2 are also given as gray symbols and lines.

the C-terminus (Figure 1); this was done by linking G99 with G11 and creating new N- and C-termini by splitting the sequence between residues D22 and G23.[28] To facilitate comparison between the wtPDZ2 and cpPDZ2 results, the mutants and structural elements are numbered according to the wtPDZ2 structure (PDB entry 1GM1).[35]

We constructed 19 site-specific mutants of cpPDZ2 (Figure 1 and Table 1). The choice of positions was guided by the Φ-value analysis previously performed on wtPDZ2.[22] Five of the 19 mutants (L18A, V29A, V65A, L85A, and L96A) were excluded from the analysis because they expressed poorly or not at all. The remaining 14 mutants were subjected to kinetic folding experiments. At each monitored urea concentration, the folding and unfolding time courses were satisfactorily fitted to single-exponential decays. The observed rate constants were plotted on a semilogarithmic scale against the urea concentration in the so-called chevron plots, and representative plots are shown in Figure 2. In some cases (I42A, V44A, A46G, I47V, I59V, and V68A), the chevron plots displayed well-defined folding and unfolding branches, while in others, the folding arms were poorly defined as a result of the destabilizing effect of the mutations.

A careful analysis of the cpPDZ2 equilibrium and kinetic (un)folding experimental data revealed the presence of a fast event lost in the dead time of the stopped-flow apparatus. Using ultrafast continuous flow kinetics (Figure 2A, top curve), we directly detected this (un)folding event, which is attributed to a low-energy folding intermediate.[28] However, it was not feasible to perform continuous-flow experiments on the mutants of cpPDZ2 presented in this study because of the large amounts of protein required to carry out such experiments vis-à-vis the low expression yields of the mutants. In an effort to characterize the folding intermediate of cpPDZ2, we resorted to measuring the folding arm of the chevron plot by employing a ligand-binding-induced folding experiment.[23] In such an experiment, a large excess of a ligand, in this case the peptide EQVSAV, is added to the denatured protein. When (i) the ligand interacts only with the native protein and (ii) the binding reaction is much faster than the folding reaction, it is possible to obtain the folding rates unaffected by the unfolding rates, even under strongly denaturing conditions.[23,36] This method was particularly useful for studying the marginally stable mutants of cpPDZ2. The ligand-induced folding experiment was performed on 12 mutants at urea concentrations ranging from 0 to 4.5 M and using peptide concentrations in the range 50−8000 μM, as K_D for the cpPDZ2−ligand interaction is on the order of 20 μM.[28] The two mutants I47A and V68A were excluded from the ligand-induced folding experiments, as their (un)folding kinetics were

(35) Walma, T.; Spronk, C. A.; Tessari, M.; Aelen, J.; Schepens, J.; Hendriks, W.; Vuister, G. W. *J. Mol. Biol.* **2002**, *316*, 1101–1110.
(36) Sanz, J. M.; Fersht, A. R. *FEBS Lett.* **1994**, *344*, 216–220.

identical to that of cpPDZ2 (data not shown). Examples of data for the ligand-induced refolding experiments are shown in Figure 2 along with the conventional chevron plots.

While the conventional (un)folding experiments resulted in apparently V-shaped chevron plots, the presence of a folding intermediate could be discerned from the curvature of the folding limb obtained from the ligand-induced refolding experiment (see Figure 2). The kinetics of a two-step reaction should in theory be fitted to the two roots of a quadratic equation.[37] However, in many cases, only one relaxation rate may be monitored. Since in the case of cpPDZ2 we previously showed that the fast and slow folding phases are kinetically uncoupled,[28] we analyzed the data assuming a fast pre-equilibrium between the denatured state (D) and the intermediate (I):

$$\mathrm{D} \overset{K_{\mathrm{D-I}}}{\rightleftharpoons} \mathrm{I} \underset{k_{\mathrm{u}}}{\overset{k_{\mathrm{f}}}{\rightleftharpoons}} \mathrm{N} \qquad (14)$$

Under such conditions, according to eq 3, the apparent folding rate constant is slowed by a factor equal to $(1 + K_{\mathrm{D-I}})$, where $K_{\mathrm{D-I}}$ is the equilibrium constant for the rapid pre-equilibrium between the denatured and intermediate states. In order to reduce the standard errors, the data for each variant were globally fitted to a three-state equation with m values assumed to be the same as for cpPDZ2 ($m_{\mathrm{I-TS}} = 0.44$, $m_{\mathrm{TS-N}} = 0.23$, and $m_{\mathrm{D-I}} = 0.48$), providing values of k_{f}, k_{u}, and $K_{\mathrm{D-I}}$ (Table 1). The m values for cpPDZ2 were obtained by globally fitting the data from the continuous-flow, stopped-flow, and ligand-induced refolding experiments, as described in ref 28. From the folding kinetics we calculated the stability change due to each mutation as well as the Φ values of the intermediate state (Φ_{I}) and the late transition state (Φ_{TS}). Fitting the data set with uncontrained m values gave larger standard errors but did not alter the distribution of the Φ values (data not shown). Reliable Φ values were calculated for 11 of the initial 19 mutants (Table 1), as described in Materials and Methods.

The compactness of the late transition states of the wtPDZ2 and cpPDZ2 folding are similar (with β_{T} values of 0.89 and 0.80, respectively) justifying a comparison of the Φ values of the two states. Each point in Figure 3A represents the Φ_{TS} values for corresponding positions and mutations in wtPDZ2 and cpPDZ2. The correlation between the two data sets is almost perfect with a unit slope ($R = 0.96$), indicating a highly conserved structure for the late transition state. The distribution of Φ values is indicated in the wtPDZ2 structure in Figure 4.

To compare the structure of the rate-limiting TS to that of the intermediate state, we plotted the Φ_{I} values against the Φ_{TS} values for each position (Figure 3B). The most of the values fall on a line with unit slope, indicating that most of the interactions are formed to the same extent in the two states. Three positions have significantly larger Φ values in the late transition state TS, one located in α2 (A81G) and the other two in the main β-sheet (L66A in β4 and V92A in β6) (Figure 4). At only one position (A53G in α1) is Φ_{TS} clearly smaller than Φ_{I}, suggesting that the intermediate is more structured than the TS in the region of the helix α1.

The large Φ values for the intermediate state (Φ_{I}) are clustered in the terminal β-hairpins, indicating two seemingly independent folding nuclei in these regions (Figure 4). The largest Φ values for the N-terminal β-hairpin are found in β3 (I42A, V44A, and A46G) extending toward β2 (L25A and I27V), whereas the largest Φ values for the C-terminal β-hairpin are found in β1 (V16A) radiating toward β6 (V92A). Additionally, a relatively large Φ value was noted for A53G in α1. Taken together, the results suggest that the intermediate state is constituted by (partially) formed terminal β-hairpins docked with each other and α1.

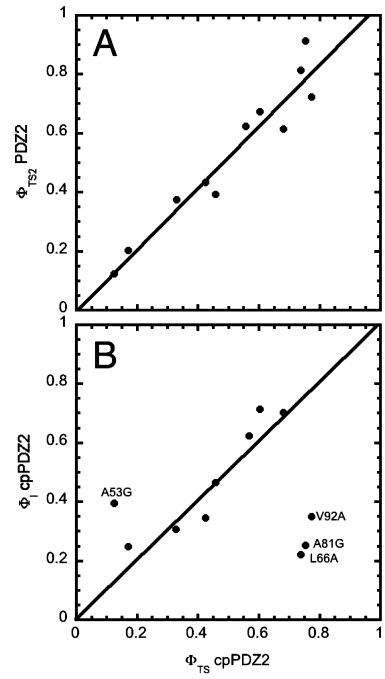

Figure 3. Correlation between (A) corresponding Φ_{TS2} values for wtPDZ2 and cpPDZ2 and (B) Φ values for the intermediate state and late transition state in cpPDZ2 folding. The high correlation between the Φ_{TS2} values in (A) suggests that the structure of the late transition state TS2 is largely conserved. The comparison between Φ_{TS2} and Φ_{I} in (B) reveals that a large part of the interactions are formed to similar extents in the two states; the four exceptions are labeled with amino acid numbers and mutations.

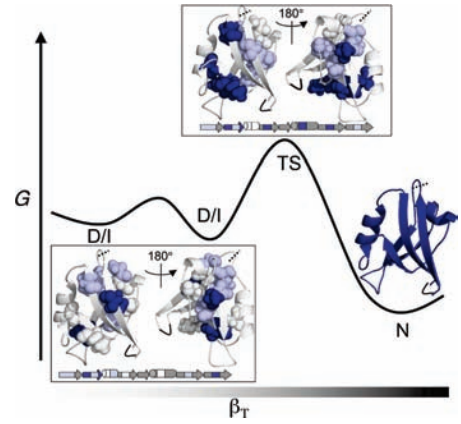

Figure 4. Summary of the folding of cpPDZ2. The folding is characterized by a fast pre-equilibrium between the denatured state (D) and a populated intermediate state (I). As discussed in the text, the data do not allow a determination of whether I is productive or nonproductive. The Φ values of I and the transition state (TS) are mapped on the wtPDZ structure. The Φ values are grouped into three different classes: small values (<0.3) are shown in white, intermediate values ($0.3 < \Phi < 0.6$) in light-blue, and large values (>0.6) in deep-blue. The Φ values are further mapped in the schematic representations of the secondary structure elements underneath the structures, highlighting that the large Φ_{I} values are clustered around the new N- and C-terminal β-hairpins.

Discussion

Much of our current knowledge of the protein folding reaction has been achieved by extensively characterizing the folding mechanisms of simple proteins.[38] In view of the complexity of

the folding problem, an informative approach has been to study the folding of different members of the same family in order to unveil general correlations between amino acid composition and folding pathways. In fact, such an approach has shown that the overall folding mechanism is generally conserved within a protein family[6–9] and that hidden common features may be unveiled even when apparently different folding mechanisms are observed.[2–4] In this context, the study of topologically perturbed proteins, such as circular permutants, may represent a critical step forward. In these cases, while the chain topology is altered, the amino acid composition and overall structure are the same. In this work, we have addressed the effect of circular permutation on the folding of a PDZ domain by employing a Φ-value analysis and comparing the structures of the rate-limiting transition state and a folding intermediate with that of the wild-type protein and other members of this family.

A very useful parameter in analyzing the nature of a protein folding transition state is the Tanford β value (β_T), which is an estimate of the degree of buried surface exposure in the transition state for unfolding relative to the denatured and native states. The folding pathway of wtPDZ2 is characterized by two sequential transition states TS1 and TS2 displaying β_T values of 0.62 ± 0.04 and 0.89 ± 0.05, respectively. As outlined above, we observed that the β_T and Φ values obtained for the rate-limiting TS in the folding of cpPDZ2 are similar to those of TS2 in PDZ2 (Figure 4); this indicates that the structure of the late TS is marginally affected by circular permutation. Therefore, we conclude that the late events in the folding of PDZ2 are stabilized by a robust folding nucleus that resists topological mutations. The insensitivity of the late transition state in PDZ2 folding to the altered sequence connectivity recalls what was observed previously for the circularly permuted variant of CI2.[12]

The early events of protein folding are often characterized by the formation of local structures such as isolated α-helices and β-hairpins.[39] The preorganization of these isolated structural elements typically depends on their inherent stabilities.[4,40] The two-strands-one-helix motif has been proposed as a basic unit (or "foldon") in the folding of α/β proteins. This foldon has been observed in the transition states of a variety of α/β proteins, such as CI2,[41] human procarboxypeptidas A2,[42] U1A,[43] CheY,[44] acyl phosphatase,[45] and the POB domain.[46] Furthermore, it has been proposed that the malleability of the folding pathways of α/β proteins stems from the multiplicity of ways the two-strands-one-helix motif can be established.[15] By circular permutation, PDZ2 acquires a symmetrical distribution of the α and β elements (β2-β3-α1-β4-β5-α2-β6-β1) along the midpoint of the amino acid sequence, with each half of the sequence containing a potential two-strands-one-helix foldon. Such a symmetrical arrangement has been proposed by theoretical studies[47] and shown by experiments[48,49] to promote multiple folding pathways. The N- and C-terminal β-strands of cpPDZ2 may adopt β-hairpin motifs involving only local contacts; furthermore, the N-terminal part of the protein can form a two-strands-one-helix folding unit solely by local contacts. Indeed, the Φ values for the intermediate (Figure 4) indicate that this loosely formed structure is constructed from the terminal β-hairpins, which are consolidated by docking against each other and the minor α1.

The effect of circular permutation appears to be different for diverse proteins. In particular, the transition state for folding of CI2 is not affected by circular permutation,[12] whereas the SH3 and S6 domains display structural rearrangements upon loop entropy perturbations,[11,14] the latter protein displaying more pronounced changes than the former. A rational explanation of such differences has been proposed to arise from the presence or absence of multiple nuclei driving the folding reaction,[15] with the number of accessible pathways being linked to the number of nucleation motifs contained within the native structure. From this viewpoint, the effect of permuting CI2 is probably small because there are no such alternative nuclei. On the other hand, in the case of S6, the presence of multiple foldons has been inferred by probing the folding pathways of systematically permutated variants. In this context, it is interesting to analyze the folding pathway of cpPDZ2 in comparison with previously characterized PDZ domains. In fact, by performing a comparative Φ-value analysis on two different PDZ domains, we have recently shown that the folding pathway of these proteins displays a weak bias toward the native topology at the early stages of folding, whereas alternative pathways converge at the late stages.[25] It is then not surprising that while circular permutation stabilizes alternative nuclei in the early events of folding, as probed by the structure of the cpPDZ2 folding intermediate presented in this work, the major rate-limiting transition state is less sensitive to loop entropy perturbations and similar to that of PDZ2.

We previously showed that circular permutation in PDZ2 stabilizes an intermediate in the folding pathway.[28] In this context, it is worth comparing the folding of cpPDZ2 with that in the naturally occurring circular permutant D1pPDZ.[27] In fact, at variance with what was observed for canonical PDZ domains, both of these circular permutants display a low-energy intermediate. Furthermore, in the case of D1pPDZ, the observed kinetics allowed an unambiguous demonstration that the intermediate is an off-pathway species competing with productive folding. While at this stage it is impossible to assign a kinetic role to the folding intermediate in cpPDZ2, it is tempting to assume that this intermediate is also an off-pathway kinetic trap, in analogy with the D1pPDZ off-pathway folding intermediate. On the basis of this view, the Φ-value analysis of cpPDZ2 suggests that circular permutation allows the formation of two competing folding nuclei, one being a two-strands-one-helix foldon composed of the N-terminal β-hairpin and the structurally and sequentially adjacent α1 (β2-β3-α1) and the other involving contact formation in the C-terminal β-hairpin (β6-β1) forming part of the main β-sheet; in the folding of wtPDZ2, the latter β-hairpin forms the early transition state TS1 when docked with β4 and α2. The kinetic partitioning between these two main folding nuclei would then dictate the fraction of protein being

(37) Fersht, A. R. *Structure and Mechanism in Protein Science*; Freeman: New York, 1999.
(38) Jackson, S. E. *Folding Des.* **1998**, *3*, R81–R91.
(39) Eaton, W. A.; Munoz, V.; Thompson, P. A.; Chan, C. K.; Hofrichter, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 10–14.
(40) Daggett, V.; Fersht, A. R. *Trends Biochem. Sci.* **2003**, *28*, 18–25.
(41) Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. *J. Mol. Biol.* **1995**, *254*, 260–288.
(42) Villegas, V.; Martínez, J. C.; Aviles, F. X.; Serrano, L. *J. Mol. Biol.* **1998**, *283*, 1027–1036.
(43) Ternstrom, T.; Mayor, U.; Akke, M.; Oliveberg, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14854–14859.
(44) Lopez-Hernandez, E.; Serrano, L. *Folding Des.* **1996**, *1*, 43–55.
(45) Taddei, N.; Chiti, F.; Fiaschi, T.; Bucciantini, M.; Capanni, C.; Stefani, M.; Serrano, L.; Dobson, C. M.; Ramponi, G. *J. Mol. Biol.* **2000**, *300*, 633–647.
(46) Sharpe, T. D.; Ferguson, N.; Johnson, C. M.; Fersht, A. R. *J. Mol. Biol.* **2008**, *383*, 224–237.
(47) Klimov, D. K.; Thirumalai, D. *J. Mol. Biol.* **2005**, *353*, 1171–1186.
(48) Kim, D. E.; Fisher, C.; Baker, D. *J. Mol. Biol.* **2000**, *298*, 971–984.
(49) McCallister, E. L.; Alm, E.; Baker, D. *Nat. Struct. Biol.* **2000**, *7*, 669–673.

trapped in the intermediate state before proceeding to the native state. Remarkably, we did not observe any unusual Φ values for the cpPDZ2 intermediate, suggesting that such a partitioning may be due to pure topological frustration.

Conclusions

Energy landscape theory suggests that proteins may reach their native conformation via multiple pathways and alternative folding nuclei.[50] Previous work on circularly permuted proteins suggested that the dominance of such nuclei may be tuned by loop entropy mutations. In this work, we have compared the folding of wtPDZ2 with that of the circularly permuted variant, cpPDZ2. While the late transition state is robust and apparently unaffected by circular permutation, the Φ values for the folding intermediate reveals that alternative folding nuclei are operative and may be selectively stabilized.

Acknowledgment. Y.I. was supported by a fellowship from the Wenner-Gren Foundations (Sweden). This work was partially supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (2007B57EAB_004, 20074TJ3ZB_005, and RBRN07BMCT_007).

JA900438B

(50) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167–195.