J|A|C|S
COMMUNICATIONS

Published on Web 02/21/2007

Folding Intermediate in the Villin Headpiece Domain Arises from Disruption of a N-Terminal Hydrogen-Bonded Network

Jana Khandogin,[†] Daniel P. Raleigh,[‡] and Charles L. Brooks, III*,[†]

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and Department of Chemistry, State University of New York at Stony Brook, Stony Brook, New York 11794

Received December 12, 2006; E-mail: brooks@scripps.edu

A complete understanding of protein folding requires knowledge of the structure and energetics of various states on the free energy landscape and of how specific interactions, as well as environmental effects, modulate their relative stabilities and the mechanism for interconversion. In a traditional view, small globular proteins are thought to fold cooperatively in which only the native (folded) and denatured (unfolded) states are populated at equilibrium. Advances in NMR techniques have allowed direct measurement of transiently and sparsely populated intermediates for small fast-folding proteins under native conditions,[1,2] thus providing evidence for a more complex picture of protein folding.

The villin headpiece domain is a 67-residue protein (HP67) composed of the N-terminal (residues 10−42; the first nine residues can be removed without affecting its stability or function) and C-terminal (43−76) subdomains. The N-terminal subdomain is mainly made up of turns and loops around a hydrophobic core of residues F16, L21, L29, P30, V33, K38, and H41, while the C-terminal subdomain is an autonomous folding unit comprising three helices surrounding a hydrophobic core of residues F47, F51, M53, F58, and L69.[3,4] The two subdomains are stabilized via an interfacial hydrophobic cluster consisting of residues L18, L42, V50, and L75 and by a buried salt bridge between E39 and K70. Under native conditions at pH 7.0, folding of the intact HP67 has been recently shown to involve a marginally populated intermediate, in which the N-terminal subdomain is nearly random coil but the C-terminal subdomain retains a native fold.[5,6] At pH below 7, the N-terminal subdomain unfolds because of the protonation of H41. Interestingly, the pK_a of the native HP67 was estimated to be at least 1.7 units below the value of the intermediate state.[6] Thus, the titration behavior of H41 offers a unique probe for the conformational state of HP67. Here we describe an atomistic simulation study of two conformational states of HP67 (Figure 1), referred to as **1**, derived from an X-ray crystal structure (PDB ID: 1YU5[4]), and **2**, derived from a minimized average solution NMR structure (PDB ID: 1QQV[3]). We provide evidence that while state **1** represents the native fold of HP67, state **2** more closely resembles an intermediate.

We have recently developed a novel molecular simulation technique, continuous constant pH molecular dynamics (CPHMD), which enables a microscopic coupling between protein conformational dynamics and protonation equilibria in a macroscopic solvent representation under a specified solution pH condition.[7,8] Combined with the replica−exchange (REX) enhanced conformational sampling protocol, we demonstrated that short-time CPHMD simulations can robustly reproduce experimental pK_a values in proteins to within 1 pK_a unit.[9] Here we performed 1-ns REX-CPHMD simulations for HP67 starting from the X-ray and NMR structures at various pH values to determine the titration behavior of H41 (Supporting Information). We obtain a pK_a of 2.2 and 5.6 for state **1** and **2**, respectively (Figure 2a). While the former value is consistent with the estimated pK_a (< 4) for the native state,[6] the latter matches remarkably well the measured value (5.7) for a sparsely populated intermediate.[6] The significantly different titration behavior of H41 in **1** and **2** and the excellent agreement with the NMR experiment[6] led us to hypothesize that the environment near H41 in **2**, in fact, more closely resembles the intermediate than the native state. We tested this hypothesis by characterizing its structural and dynamic properties using CPHMD simulations. These simulations provide unique mechanistic insights into the unfolding of the

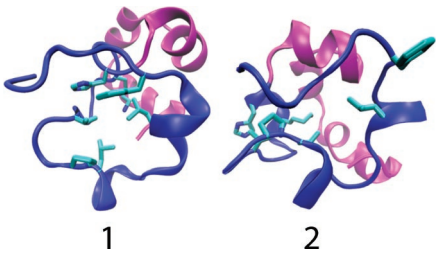

Figure 1. Representative conformations for state **1** and **2** obtained as centroids of the largest conformational cluster from the REX-CPHMD simulations at pH 7. The N- and C-terminal subdomains are colored blue and magenta, respectively. The side chains of the N-terminal hydrophobic cluster are shown in liquorice.

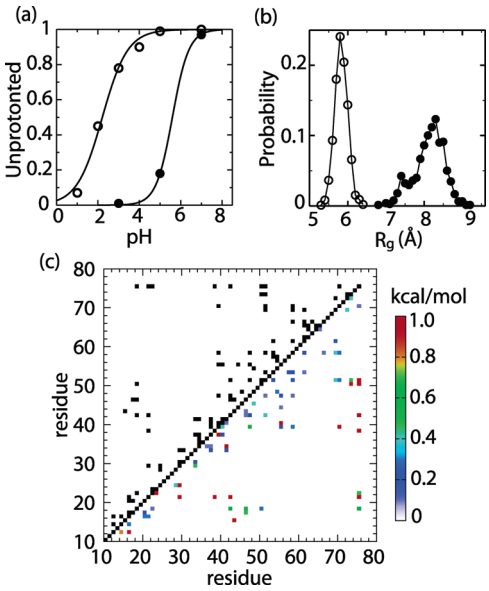

Figure 2. (a) The unprotonated fractions of H41 in state **1** (open) and **2** (filled) are plotted against pH. (b) Probability density of the radius of gyration (R_g) for the N-terminal hydrophobic cluster in **1** (open) and **2** (filled). (c) In the upper triangle, a contact map for **1** is given. Two residues are considered in contact (black), if $P_{ij}^1 > 0.4$, where P_{ij}^1 is the probability for the minimum distance between side-chain heavy atoms to be within 4.5 Å. The lower triangle shows a map of contact destabilization energies for **2**, defined as $-RT \ln(P_{ij}^2/P_{ij}^1)$ with P_{ij}^2 being the contact probability in **2**. Only contacting pairs in 1 are considered. The contacts with $P_{ij}^2 < 0.1$ are colored red.

[†] The Scripps Research Institute.
[‡] State University of New York at Stony Brook.

10.1021/ja0688880 CCC: $37.00

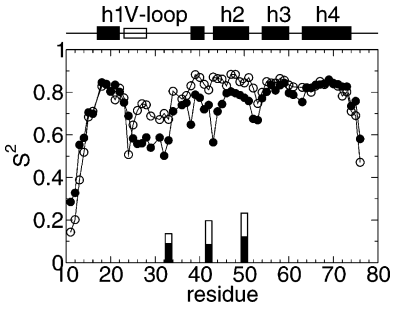

Figure 3. Computed NMR order parameters (S^2) are plotted against amino acid sequence for state **1** (open) and **2** (filled symbols). The backbone N−H order parameters are represented as circles while the methyl C−H parameters are represented as bars. For the latter, only those within the N-terminal and interface hydrophobic clusters displaying significant changes (>25%) are shown.

N-terminal subdomain which are not accessible by conventional molecular dynamics simulations.

We carried out 5-ns CPHMD simulations at 298 K and pH 7 starting from the X-ray and NMR structures, respectively. The total energy remains constant throughout the former trajectory but decreases rapidly in the first 500 ps before stabilizing in the latter (Figure S2) indicating that the NMR structure does not represent a local energy minimum in the context of the force field applied (Supporting Information). To rule out the possibility that **2** represents a local conformational state closely related to **1**, we tested whether a trajectory initiated with the NMR structure led to the X-ray structure. After a rapid initial increase in rmsd to above 7 Å, consistent with the time evolution of the total energy, the rmsd stabilizes at about 7.5−8 and 7−7.5 Å with respect to the NMR and X-ray structures of the intact HP67, respectively (Figure S2). Thus, state **2** does not appear to be related to **1**. To investigate the internal dynamics of HP67 in **1** and **2**, we calculated the backbone N−H order parameters (Figure 3). In state **1**, the N-terminal subdomain is more flexible than the C-terminal one; the residues preceding helix 1 (17−22) as well as T24 display significant mobility. Overall, except for a more mobile C-terminus, the trend in the computed order parameters for **1** is in good agreement with the NMR data for the native state.[6] In **2**, by contrast, residues 25−33 in the V-loop (22−28) and the adjacent turn region as well as K38, S43, and G52, M53 of helix 2 exhibit backbone motions with large amplitude. These residues encompass the N-terminal and inter-subdomain hydrophobic cores, thus suggesting a possible partial unfolding. To further probe the dynamics of the N-terminal and inter-subdomain hydrophobic clusters in **2**, we computed their side-chain methyl C−H order parameters relative to **1**. The most significant change is observed at V33, L42, and V50, further suggesting a partial unfolding of the interface (Figure 3).

To examine the structural properties of state **1** and **2** that give rise to the internal dynamics as discussed above, we performed 4-ns REX-CPHMD simulations at pH 7 initiated from the X-ray and NMR structures, respectively. The REX protocol was employed to ensure better conformational convergence (Figure S1). In comparison to state **2**, the radius of gyration (Figure 2b) and solvent accessible surface area in **1** (Figure S3) show a significant increase for the N-terminal hydrophobic cluster but remain unchanged for the C-terminal one. Figure 2c shows the contact map for state **1** and the destabilization energies of these contacts in **2**. While the majority of inter-residue contacts are weakened, the interactions between L21 and K38, L21 and L29, T24 and L29, and L21 and L42, which contribute to the stability of the N-terminal hydrophobic core, are completely disrupted. Also, the inter-subdomain hydrophobic contacts between T15 and S43, L21 and L75, K38 and L75, L42 and L75, and V50 and L75 are lost, which is most likely the cause for the increased mobility of helix 2 relative to helix 3 and 4 in the C-terminal subdomain (Figure 3). Furthermore, the interfacial salt bridge between E39 and K70 is disrupted. Thus, **2** may represent an intermediate state with a fully folded C-terminal but partially unfolded N-terminal subdomain. It is not quite clear whether this state is the same as the sparsely populated intermediate found under native conditions, which has a random-coil like N-terminal subdomain.[5,6] However, our results are consistent with the NMR data, which suggest that unfolding of the N-terminal subdomain is non-cooperative, whereby disruption of the interface occurs on a faster time scale than that of the hydrophobic core.[6] They are also in agreement with the finding that disruption of the interfacial salt bridge through mutation at E39 or K70 leads to (partial) unfolding of the N-terminal subdomain (DPR, unpublished data).

What makes the structural and dynamics properties of **2** so much different from the native state considering that it is derived from a structure that deviates (rms) by merely 1.7 Å from the X-ray structure? A closer examination of the two static structures (PDB ID: 1YU5 and 1QQV) reveals a subtle difference in the local environment of H41. In the X-ray structure, H41 forms two hydrogen bonds via Nδ and Nϵ with the backbone carbonyl oxygen of E14 and amide nitrogen of D34, respectively. In the NMR structure, however, the former hydrogen bond does not exist. As a consequence, H41 is significantly more mobile (Figure 3), leading to disruption of the hydrogen bond with D34 and the ability to protonate on both δ and ϵ sites as seen in the simulation of **2**. By contrast, in the simulation of **1**, only Nδ is protonated, in agreement with the NMR data of the native state.[6] We propose that the loss of the hydrogen-bonding network around H41 results in an increased flexibility in the turn region (see 31−33 in Figure 3), which propagates to induce the disruption of the interface contacts as well as partial unfolding of the N-terminal hydrophobic cluster. It also results in stabilization of the doubly protonated charged form of H41, as the nonspecific electrostatic interaction with the side chains of E14 and D34 becomes dominant, which explains its increased pK_a value in state **2**. Interestingly, the loss of the hydrogen bond network also accompanies the acid-induced unfolding of the N-terminal subdomain in a REX-CPHMD simulation at pH 2 initiated from the X-ray structure (data not shown). Taken together, our results highlight the importance of subtle interactions in modulating the protein folding landscape.

Acknowledgment. Financial support from the National Institute of Health (Grants GM57513 and GM48807 to C.L.B.) and from the NSF (Grant MCB 614365 to D.P.R.) is greatly appreciated.

Supporting Information Available: Computational details and supporting data. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Vugmeyster, L.; Kroenke, C. D.; Picart, F.; Palmer, A. G., III; Raleigh, D. P. *J. Am. Chem. Soc.* **2000**, *122*, 5387−5388.

(2) Roder, H.; Maki, K.; Cheng, H. *Chem. Rev.* **2006**, *106*, 1836−1861.

(3) Vardar, D.; Buckley, D. A.; Frank, B. S.; McKnight, C. J. *J. Mol. Biol.* **1999**, *294*, 1299−1310.

(4) Meng, J.; Vardar, D.; Wang, Y.; Guo, H.-C.; Head, J. F.; McKnight, C. J. *Biochemistry* **2005**, *44*, 11963−11973.

(5) Tang, Y.; Grey, M. J.; McKnight, J.; Palmer, A. G., III; Raleigh, D. P. *J. Mol. Biol.* **2006**, *355*, 1066−1077.

(6) Grey, M. J.; Tang, Y.; Alexov, E.; McKnight, C. J.; Raleigh, D. P.; Palmer, A. G., III. *J. Mol. Biol.* **2006**, *355*, 1078−1094.

(7) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. *Proteins* **2004**, *56*, 738−752.

(8) Khandogin, J.; Brooks, C. L., III. *Biophys. J.* **2005**, *89*, 141−157.

(9) Khandogin, J.; Brooks, C. L., III. *Biochemistry* **2006**, *45*, 9363−9373.

JA0688880