Protein Folding

DOI: 10.1002/ange.200603415

De Novo Folding of the DNA-Binding ATF-2 Zinc Finger Motif in an All-Atom Free-Energy Forcefield**

*Srinivasa M. Gopal and Wolfgang Wenzel**

Zinc fingers are among the most abundant proteins in eukaryotic genomes and occur in many DNA-binding domains and transcription factors.[1] They function in DNA recognition, RNA packaging, transcriptional activation, protein folding, and assembly and apoptosis. Many zinc fingers contain a Cys_2His_2 binding motif that coordinates the Zn ion in an αββ framework.[2–4] Much effort has been directed towards the engineering of novel zinc fingers.[5]

All-atom folding studies of the classical TFIIIa αββ motif may help elucidate the structural and functional properties of zinc finger proteins, but they tax the presently available all-atom folding methods. Recent improvements in simulation methods and forcefields have permitted the folding of several small helical proteins, hairpins, and β sheets from the unfolded ensembles.[6–14] The reproducible folding of proteins with mixed secondary structures, however, remains a significant challenge for the accuracy of the all-atom forcefield and the simulation method.[15]

Here we use the all-atom free-energy forcefield PFF02 to predictively fold a segment (amino acids 23–51) of the N-terminal subdomain of ATF-2 (Protein Data Bank (PDB) file 1BHI),[8,16] a 29 amino acid peptide that contains the basic leucine zipper motif. This peptide folds into the classical TFIIIa αββ conformation found in many zinc-finger subdomains. The fragment contains all the conserved hydrophobic residues (Phe25, Phe36, and Leu42) of the classical zinc-finger motif and the Cys_2His_2 zinc-binding pattern.

Starting with a completely unfolded conformation with no secondary structure (16-Å backbone root mean square deviation (bRMSD) from the native peptide), we performed 200 cycles of an evolutionary algorithm,[17] which speeds up the search of the conformational space by evolving an entire population (population size = 20), instead of a single conformation. The total simulation is split into many small steps, during which many closely related conformations are explored with fictitious dynamics to accelerate the search of the free-energy surface. New conformations are admitted into the active population according to a set of criteria that balance the free-energy gain and the structural diversity of the active population.

The distribution of the bRMSD versus the energy of all accepted conformations (Figure 1) demonstrates that the simulation explores a wide variety of conformations, with regard to their free-energy and their deviation from the native conformation. Among the 10 energetically lowest conformations, six fold into near-native conformations with bRMSDs of 3.68–4.28 Å, while four fold into conformations with a larger bRMSD. The three energetically best conformations are all near-native in character; an overlay with the experimentally determined conformation (Figure 2, top left) illustrates that the helix, the β sheet, and both turns are correctly formed. The hydrophobic residues, which determine the packing of the β sheet against the helix, are illustrated in orange in the figure. The helical section (Glu39–Glu50) and the β sheet (Phe25–Leu26 and Arg35–Phe36) deviate individually by bRMSDs of 1.6 and 2.4 Å, respectively, from the experimental counterparts. The overall deviation between the experimental and simulated folded conformations arises in the relative arrangement of the β sheet and the helix, which is controlled by unspecific hydrophobic interactions. All conserved hydrophobic side chains are also buried in the folded structure. The zinc-coordinating cysteine residues (Cys27 and Cys32) are within 2 Å of their native positions and are available for association with the Zn ion. We find many low-energy conformations in which much of the structure of the zinc finger is preformed. Binding of the metal ion is ultimately required to provide an important enthalpic contribution for the stabilization of the native conformation. These data demonstrate that the forcefield and simulation methodology are free of bias with respect to the formation of a particular type of the secondary structure and are able to describe the complex folding dynamics of at least small peptides/proteins with mixed secondary structures.

The partially folded low-energy conformations among the 10 energetically best (squares in Figure 1) all have a fully

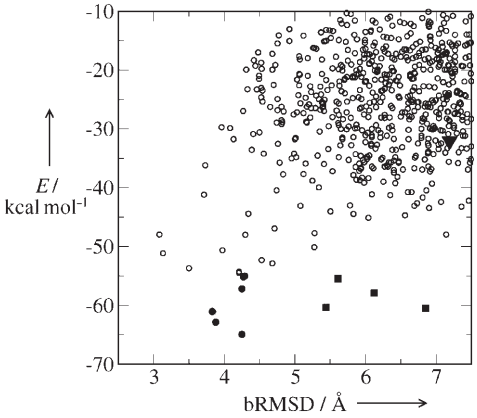

Figure 1. Internal free energy versus bRMSD of all accepted conformations in the simulation. Full circles/squares indicate the near-native and partially folded fractions of the 10 best conformations in the final population. The large triangle (▼) indicates the folding intermediate illustrated in the bottom right panel of Figure 2.

[*] S. M. Gopal, Dr. W. Wenzel
Institute for Nanotechnology
Research Centre Karlsruhe
Post Box 3640, 70621, Karlsruhe (Germany)
Fax: (+49) 724-7826434
E-mail: wenzel@int.fzk.de

[**] This work was funded by the Deutsche Forschungsgemeinschaft (WE 1863/10-2) and the Bode Foundation. We acknowledge the use of the computational facilities of the KIST Supercomputational Materials Lab.

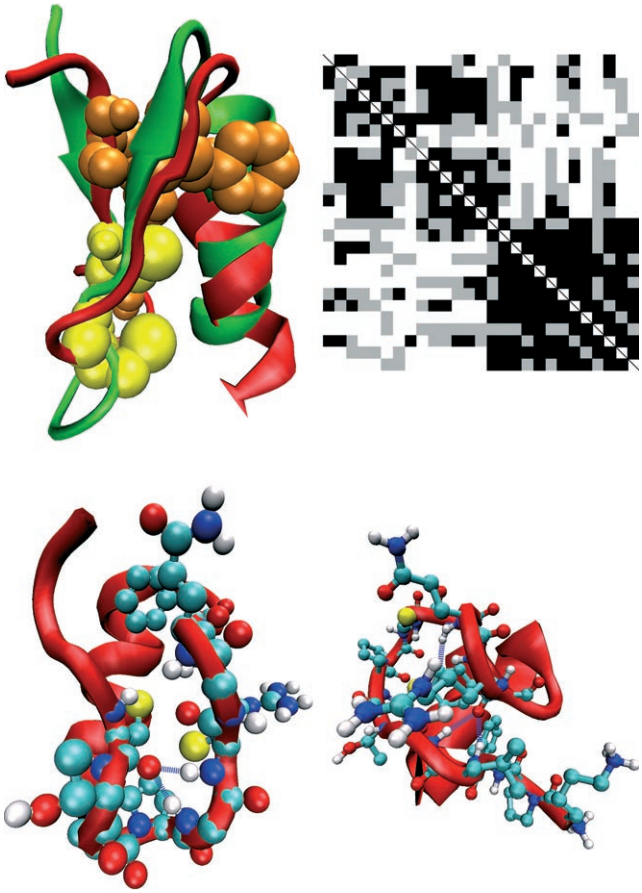

Figure 2. Top left: Overlay of the native (green) and folded (red) conformations of the peptide with the conserved hydrophobic residues shown in orange and the zinc-binding cysteine residues shown in yellow. Top right: C_β–C_β distance matrix of the folded conformation, which indicates the presence of 67% of the native contacts. Bottom left: Folding intermediate at the bottom of the free-energy surface with buried cysteine and stabilizing backbone hydrogen bonds (blue dashed lines; see text). Bottom right: Intermediate conformation with a partially formed helix and β sheet. Protein structures were drawn with VMD.[18] Red: backbone and oxygen; cyan: carbon; blue: nitrogen; yellow: sulfur; white: hydrogen.

formed helical section in the right region of the amino acid sequence, a turn near Thr28–Pro30, and significant β content. The bottom left panel of Figure 2 shows the conformation furthest from the native one, with a partially unzipped β sheet. The hydrogen bonds near the turn are still present, while the native H-bonds at the end of the zipper have not yet formed.

Figure 3 shows the convergence of the energy: after about 120 steps per population member (3.5×10^8 function evaluations), the population converged with the native ensemble. According to the funnel paradigm for protein folding,[19] the tertiary structure forms as the protein slides downhill on the free-energy surface from the unfolded ensemble towards the native conformation. Here, each member of the active population of the evolutionary algorithm represents one point on this high-dimensional free-energy surface. Each cycle (simulated annealing with geometric cooling, number of steps $= 10^4 \sqrt{\text{cycle number}}$) generates a small perturbation on the existing conformation, which averages to a 0.5 Å

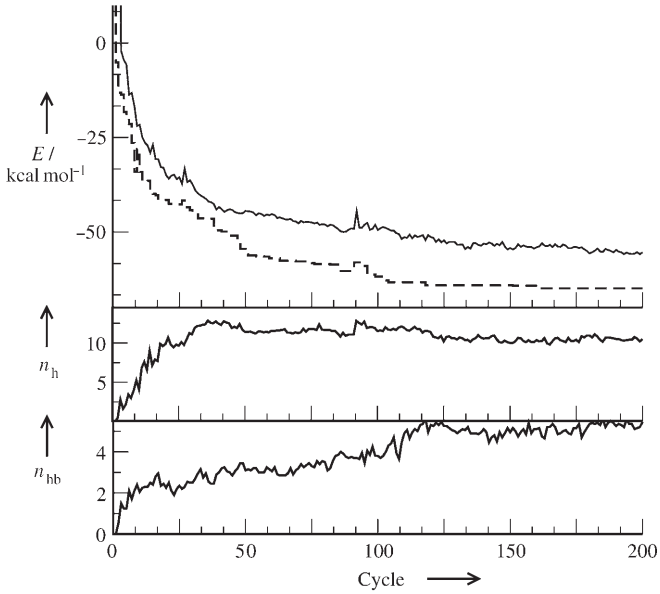

Figure 3. Average (solid line) and best (dashed line) energies, average number of amino acids (n_h) in a helical conformation (as computed by DSSP[20]), and number of hydrogen bonds (n_{hb}) as a function of the cycle number.

bRMSD change (with a maximum of 3 Å initially). As new low-energy conformations replace old conformations, the population slides down the funnel of the free-energy landscape.

Ensemble averages as a function of free-energy over the moving population are thus associated with different stages of the structure-formation process, even though our method does not generate any kinetic information. The lower two panels of Figure 3 show the average helical content and the number of β-sheet H-bonds as a function of the cycle number. After a rapid collapse to a compact conformation, the helix forms first, and this is followed by formation of the β sheet. An analysis of the folding funnel upwards in energy illustrates that the lowest energy metastable conformations correspond to a partial unzipping of amino acids Phe25–Arg35, while the conserved cysteine residues are still buried (Figure 2, bottom left). Even much higher on the free-energy funnel (triangle in Figure 1), we find many structures that have a lot of residual secondary structure but few long-range native contacts. Figure 2 (bottom right) illustrates one such conformation where the preformed sheet region is stabilized by H-bonds (Leu26–Cys27, Arg35) and packs at a right angle to a broken helix. Here, the hydrophobic residues are only partially buried.

As the algorithm tracks the development of the population, it is possible to reconstruct a folding pathway by starting with the converged conformation and moving backwards to the completely unfolded conformation. Crucial steps along the continuous folding pathway are illustrated in Figure 4. (Note that there is no quantitative mapping onto the time axis.) The early folding process is characterized by helix nucleation and concurrent collapse into a globular conformation with a radius of gyration that is comparable to that of the native conformation. The simulation then explores confor-

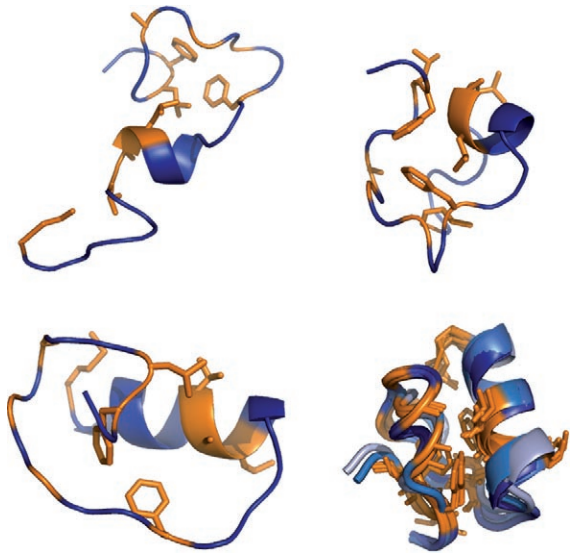

Figure 4. Key events in the folding process: helix nucleation (top left) with concomitant collapse into a globular conformation (top right), followed by helix growth. The fully formed helix (bottom left) acts as a template against which many partially formed beta-sheet conformations can pack (bottom right). The protein backbone is shown in blue, and the hydrophobic residues including side chains are shown in orange. In the bottom right panel, the shading of the backbone becomes darker as the simulation approaches the native conformation.

mations of the same spatial extent with increasing helical, but no β-sheet content. Lower on free-energy funnel, the simulation samples conformations in which partially formed β sheets pack against the helix.

De novo folding of the zinc-finger domain permits a direct sampling of the relevant low-energy portion of the free-energy surface as a first step towards the elucidation of the structural mechanisms involved in DNA binding.[5,21] On the basis of the free-energy estimate, conformations without the helix (at least 8 kcal mol^{-1} higher than the native conformation) cannot participate in DNA binding and transcription. By using the global optimization approach, which uses fictitious dynamics, we can rapidly explore the relevant free-energy region of the biologically active ensemble. Met51 packs in all low-energy conformations against the combined αββ scaffold and acts as a closure of the DNA-binding motif. It may thus provide an enthalpic contribution to a nonstandard helix-capping motif that differs from the TGEKP linker sequence observed in multifinger domains[1] in several zinc fingers.

We have thus demonstrated predictive all-atom folding of the DNA-binding zinc-finger motif in a free-energy force-field. This investigation offers the first unbiased characterization of the low-energy free-energy surface of the zinc-finger motif, which is unattainable in coarse-grained, knowledge-based models. We find that the helix forms first along the folding path and acts as a template against which a variety of near-native β-sheet backbone arrangements can pack. There are many zinc fingers with RMS deviations of less than 2 Å relative to the conformation illustrated in PDB file 1BHI;[16] thus, this investigation provides one important step in the theoretical understanding of zinc-finger formation and function.

Received: August 21, 2006
Published online: October 24, 2006

Keywords: evolutionary algorithms · protein folding · simulations · structure prediction · zinc-finger proteins

[1] J. H. Laity, B. M. Lee, P. E. Wright, *Curr. Opin. Struct. Biol.* **2001**, *11*, 39–46.
[2] M. S. Lee, G. P. Gippert, K. V. Soman, P. E. Wright, *Science* **1989**, *245*, 635–637.
[3] N. P. Pavletich, C. O. Pabo, *Science* **1991**, *252*, 809–817.
[4] S. A. Wolfe, L. Nekludova, C. O. Pabo, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 183–212.
[5] a) Y. Choo, M. Isalan, *Curr. Opin. Struct. Biol.* **2000**, *10*, 411–416; b) F. D. Urnov, J. C. Miller, Y.-L. Lee, C. M. Beausejour, J. M. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory, M. C. Holmes, *Nature* **2005**, *435*, 646–651.
[6] A. Schug, A. Verma, T. Herges, K. H. Lee, W. Wenzel *ChemPhysChem* **2005**, *6*, 2640–2646.
[7] a) J. Lee, H. A. Scheraga, S. Rackovsky, *J. Comput. Chem.* **1997**, *18*, 1222–1232; b) D. J. Wales, H. A. Scheraga, *Science* **1999**, *285*, 1368–1372; c) J. M. Carr, D. J. Wales, *J. Chem. Phys.* **2005**, *123*, 234901.
[8] a) T. Herges, W. Wenzel, *Biophys. J.* **2004**, *87*, 3100–3109; b) A. Verma, W. Wenzel, unpublished results.
[9] J. Chen, W. Wen, C. Brooks III, *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.
[10] a) C. Simmerling, B. Strockbine, A. Roitberg, *J. Am. Chem. Soc.* **2002**, *124*, 11258–11259; b) J. A. Vila, D. R. Ripoll, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 14812–14816; c) M. Khalili, A. Liwo, H. A. Scheraga, *J. Mol. Biol.* **2006**, *355*, 536–547.
[11] a) A. Schug, T, Herges, W. Wenzel, *Phys. Rev. Lett.* **2003**, *91*, 158102; b) T. Herges, W. Wenzel, *Phys. Rev. Lett.* **2005**, *94*, 018101; c) A. Schug, W. Wenzel, *J. Am. Chem. Soc.* **2004**, *126*, 16736–16737.
[12] a) C. D. Snow, L. Qiu, D. Du, F. Gai, S. J. Hagen, V. S. Pande, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4077–4082; b) P. G. Bolhuis, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12129–12134; c) D. A. Evans, D. J. Wales, *J. Chem. Phys.* **2004**, *121*, 1080–1090.
[13] W. Wenzel, *Europhys. Lett.* **2006**, *76*, 156–162.
[14] P. Ferrara, A. Caflisch, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10780–10785.
[15] A. Abagyan, M. Totrov, *J. Comput. Phys.* **1999**, *151*, 402–412.
[16] A. Nagadoi, K. Nakazawa, H. Uda, K. Okuno, T. Maekawa, S. Ishii, Y. Nishimura, *J. Mol. Biol.* **1999**, *287*, 593–607.
[17] A. Schug, W. Wenzel, *Biophys. J.* **2006**, *90*, 4273–4280.
[18] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.
[19] J. N. Onuchic, Z. L. Schulten, P. G. Wolynes, *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
[20] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577–2637.
[21] J. H. Laity, H. J. Dyson, P. E. Wright, *J. Mol. Biol.* **2000**, *295*, 719–727.