



## A Non-Native Helix Extension Channels Folding

S. Mohanty, U. H. E. Hansmann

published in

*From Computational Biophysics to Systems Biology (CBSB08)*,  
Proceedings of the NIC Workshop 2008,  
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,  
Walter Nadler, Olav Zimmermann (Editors),  
John von Neumann Institute for Computing, Jülich,  
NIC Series, Vol. 40, ISBN 978-3-9810843-6-8, pp. 329-332, 2008.

© 2008 by John von Neumann Institute for Computing

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume40>

# A Non-Native Helix Extension Channels Folding

Sandipan Mohanty<sup>1</sup> and Ulrich H. E. Hansmann<sup>1,2</sup>

<sup>1</sup> John v. Neumann Institute for Computing (NIC), Research Centre Jülich, 52425 Jülich, Germany  
*E-mail: s.mohanty@fz-juelich.de*

<sup>2</sup> Dept. of Physics, Michigan Technological University, Houghton, MI 49931, USA

Unbiased parallel tempering Monte Carlo simulations of a 49 residue protein starting from random conformations, reveal a non-trivial path followed by the molecule to the native state. The molecule (PDB id: 2GJH) consists of an  $\alpha$ -helix and a 3 stranded  $\beta$ -sheet, in which two of the adjacent strands straddle the other secondary structure elements along the sequence. In the course of folding, one of the strands making sequence non-local contacts is seen to be "cached" as a non-native extension of the native  $\alpha$ -helix. After the other secondary structure elements have formed and assembled in their proper tertiary arrangement, the cached segment is released and it changes its secondary structure to a strand as it attaches to a  $\beta$ -hairpin to complete the native structure. The study is based on a physics based implicit water all-atom interaction potential called the Lund force field.

## 1 Introduction

While many protein structures contain  $\beta$ -sheets with complex arrangement of  $\beta$ -strands, the folding mechanisms giving rise to such structures are largely unclear. Most successful folding simulations to date have been with  $\alpha$ -helical proteins, which are dominated by sequence local interactions. The hydrogen bonds are formed between residues  $i$  and  $i + 4$  along the sequence, which, even for the most stretched out conformation of the chain, are in the spatial neighbourhood of each other. Local interactions quickly lead such proteins into their folded structures. While  $\beta$ -sheets consisting of one or more sequence adjacent  $\beta$ -hairpins require somewhat longer range contacts along the sequence, they are still local structures. Typically they arise through a zipper-like mechanism starting from the turn regions and sequentially forming hydrogen bonds away from the turns. An important feature of these sequence local structures is that they do not interfere with the formation of other similar structures elsewhere along the chain.

When two neighbouring strands in a  $\beta$ -sheet come from regions of the sequence separated by a large number of residues, formation of contacts between them is no longer independent of the structure of the intervening segment. Premature formation of such contacts creates large steric barriers and can hinder the proper folding of other secondary structure elements. The protein is then lead into a deep local minimum and can only fold by first breaking the prematurely formed long distance native  $\beta$ -contacts. Such considerations suggest that proteins with complex  $\beta$ -sheets should on average fold slower than  $\alpha$ -helices. This is indeed consistent with experimental observations. The so called "contact order", the average sequence separation of residues in contact, bears a striking correlation with the folding rate over a huge range of folding rates.  $\beta$ -sheets with complex strand arrangements have high contact orders, and are seen to fold slowly. But is it possible that some proteins with relatively high contact order might have evolved tricks to avoid the deep local minima and fold much faster than others of the same complexity? Plots of the folding rates versus

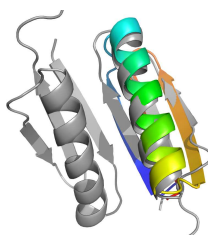


Figure 1. Comparison of the global energy minimum (colour) with the PDB structure 2GJH.pdb (grey).

contact orders suggest such a possibility, as there is a large fluctuation in the folding rates for proteins with intermediate contact orders. The exact nature of such fold accelerating mechanisms are unknown. From all-atom Monte Carlo simulations of a 49 residue protein, Top7-CFR (PDB id: 2GJH, residues 2–50) with both helical and  $\beta$ -sheet structures and a non-trivial  $\beta$ -sheet geometry, we propose one possible mechanism.

## 2 Methods

Our model represents all atoms of the protein chains, including all hydrogen atoms, but an implicit treatment of the solvent molecules through an effective interaction potential. The model assumes constant bond lengths, bond angles and peptide-bond torsion angles of 180 degrees. Each protein molecule has only the Ramachandran backbone torsion angles and the side chain torsion angles as its degrees of freedom. The effective force field contains terms to account for excluded volume repulsion, local backbone electrostatics, hydrogen bonds and hydrophobic interactions. The model and the force field have been described in detail elsewhere<sup>1,2</sup>, where we also show that the force field describes the folding and thermodynamics of a range of short peptides with both  $\alpha$ -helical and  $\beta$ -sheet structures. Sampling of protein conformations is carried out using replica exchange Monte Carlo techniques with 32 replicas. For this work, we have used the protein folding software package PROFASI<sup>3</sup>, version 1.1.2. The results presented here are based on  $1.4 \times 10^{10}$  elementary Monte Carlo updates of the protein chain per replica. All simulations were initialised with random values for all degrees of freedom and different random number seeds.

## 3 Results

The simulated molecule folds to the native state with a backbone RMSD (all residues) of about 1.8 Å. The global energy minimum found in the simulations has a backbone RMSD of 1.7 Å, and is shown superimposed on the PDB structure in Fig.1. This minimum energy structure shares all the hydrogen bonds and  $C_{\alpha}$  contacts with the native state.

More interesting than the fact that the molecule folds is the observed manner of formation of  $\beta$ -sheet contacts between the N-terminal strand and the C-terminal hairpin<sup>4</sup>. When the molecule folds from random conformations, the first structures to emerge are the native helix and the C-terminal hairpin. These are strong structural elements consisting of sequence local contacts and fold to the same structures as excised segments in simulations.

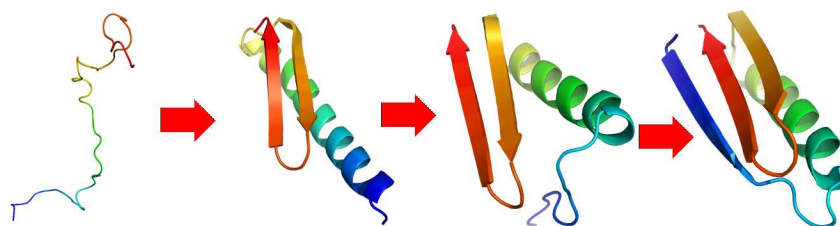


Figure 2. The folding events follow a systematic pattern, and indicate that folding is facilitated by the chameleon behaviour of the N-terminal segment.

But we observe that the N-terminal strand initially folds simply as a continuation of the native helix.  $\beta$ -strands are stabilised by inter-strand interactions. Therefore, initially when the C-terminal hairpin is absent, the N-terminal region does not have any stabilising interactions as a  $\beta$ -strand. The helix, which forms first, provides a good template for the N-terminal strands, and absorbs them. The helix, even with its non-native extension, is a structure that folds and unfolds easily. The  $\beta$ -hairpin forms independently, and subsequently makes hydrophobic contacts with the helix. Upon the formation of hydrophobic contacts between the helix and the hairpin, both the structures are stabilised. The non-native extension of the helix, containing the N-terminal strand residues, does not benefit from the hydrophobic contacts with the hairpin and eventually unfolds. Unlike the situation for an entirely unfolded molecule, when the N-terminal residues are freed with both the helix and hairpin in place, they do have other  $\beta$ -strands to bind to, which turns out to be lower in energy. Hence, the N-terminal strands join it with a larger probability.

## 4 Conclusions

Using all-atom Monte Carlo simulations starting from random initial conformations, we find that the molecule Top7-CFr folds to within 1.7 Å of its native state, following a non-trivial folding pathway. The observed mechanism of formation of sequence non-local  $\beta$ -sheet contacts depends on the chameleon behaviour of the N-terminal strand. We believe that such caching of  $\beta$ -strands in neighbouring helices is one mechanism for accelerating the formation of complex  $\beta$ -sheet structures. More detailed results are published in Refs. 4, 5.

## References

1. A. Irbäck and S. Mohanty *Biophysical Journal* **88** (2005) 1560–1569.
2. S. Mohanty and U.H.E. Hansmann, *Biophysical Journal* **82** (2006) 3573.
3. A. Irbäck and S. Mohanty *Journal of Computational Chemistry* **27**, (2006) 1548. <http://cbbp.thep.lu.se/activities/profasi/>.
4. S. Mohanty, J. H. Meinke, O. Zimmermann and U.H. E. Hansmann, *Proc. Natl. Acad. Sci. U.S.A.* (2008) Published online. Doi: 10.1073/pnas.0708411105.
5. S. Mohanty and U.H.E. Hansmann, *Caching of a Chameleon Segment facilitates Folding of a Protein with End-to-End  $\beta$ -sheet*, submitted for publication.

