



MECHANICAL UNFOLDING OF DDFLN4 STUDIED BY COARSE-GRAINED KNOWLEDGE-BASED CABS MODEL

MAKSIM KOUZA, MICHAL JAMROZ, DOMINIK GRONT,
SEBASTIAN KMIETEK AND ANDRZEJ KOLINSKI

*Faculty of Chemistry, University of Warsaw
Pasteura 1, 02-093 Warsaw Poland*

(Paper presented at the CBSB14 Conference, May 25–27, 2014, Gdansk, Poland)

Abstract: Mechanical unfolding of the fourth domain of *Distyostelium discoideum* filamin (DDFLN4) was studied using a CABS – coarse-grained knowledge-based protein model. Our study demonstrates that CABS is capable of reproducing the unfolding free energy landscape of protein unfolding and highlights an important role of non-native interactions in the protein unfolding process. The obtained three peaks in the force-extension profile suggest a four-state picture of DDFLN4 protein unfolding and correspond reasonably to the results of the all-atom simulation in explicit solvent.

Keywords: CABS, coarse-grained modeling, lattice model, mechanical unfolding of proteins

1. Introduction

Many proteins, including those in the muscle and cytoskeleton, are functional while being subjected to a wide range of mechanical forces. One of these proteins, the fourth domain of *Distyostelium discoideum* filamin (DDFLN4) (a key player in the cytoskeleton reorganization process) has been extensively studied theoretically [1, 2] and experimentally [3, 4]. The results of the Atomic Force Microscopy (AFM) experiments have shown that the force-extension profile of DDFLN4 displays two peaks located at ~ 12 nm and ~ 22 nm, suggesting a three-state scenario of its mechanical unfolding [3, 4]. A detailed interpretation of AFM experiments might be provided using molecular simulation techniques [5–7].

One of the most popular simulation approaches to study protein unfolding are $G\bar{o}$ -type reduced models, which are based solely on the topology of the native state [8–11]. In the case of DDFLN4, the $G\bar{o}$ -model simulations did not give a force-extension profile peak at ~ 22 nm extension, which is observed in the experiment, but predicted an additional peak at ~ 2 nm extension [2]. In contrast, the force-extension profile obtained by all-atom molecular dynamics simulations in



explicit solvent displays three peaks at ~ 2 nm, ~ 12 nm and ~ 22 nm, respectively, indicating that mechanical unfolding of DDFLN4 follows a four-state scenario [1]. This study has also pointed to the importance of the non-native interactions which lead to an additional intermediate state in the mechanical unfolding of DDFLN4 [1]. Unfortunately, at present, all-atom molecular dynamics in explicit solvent remains computationally challenging due to the huge system size required to study a complete process of mechanical unfolding. Therefore, more efficient, but still accurate methods of simulation are desirable. In the present paper, we test the effectiveness of the coarse-grained CABS model to study the protein unfolding process. With the help of the CABS model, we obtained a similar pattern of the force-extension profile as in simulations of all-atom molecular dynamics [1], but using much less computational power.

2. Method and Results

Inspired by the AFM experiments [3, 4] and molecular dynamics simulations on DDFLN4 protein [2, 1], we conducted the constant-velocity pulling simulation using the CABS model. Importantly, spatial resolution of protein models generated by CABS allows physically realistic models to be reconstructed in all-atom representation [12, 13] (the details of the CABS model are given in Refs. [14, 15]). In the constant-velocity pulling scheme, a protein is pulled apart by a force ramped linearly with time, while monitoring the dependence of the force (or mechanical resistance) on time. The movement of the pulled termini causes an extension of the protein and the total force can be measured by $F = k(vt - x)$, where v is the constant velocity, k is a spring constant of an AFM cantilever and x is the displacement of the pulled atom from its original position. The resulting force is computed for each time step to generate a force extension profile, which has peaks showing the most mechanically stable regions of protein.

The resulting force-extension profile for DDFLN4, obtained by the CABS model is shown in Figure 1. The presence of the three peaks implies a four-state free energy landscape of the DDFLN4 unfolding process. This result is in agreement with the all-atom simulation results [1], but not with the G \bar{o} -model simulations (the first two peaks at ~ 2 nm and ~ 12 nm detected) [2] and AFM experiments (the last two peaks at ~ 12 nm and ~ 22 nm detected) [3]. The structures of the first and second transition states are shown on the upper snapshots of Figure 1. The first transition state remains native-like with all β -strands kept formed, whereas the second transition state is formed by the core of native C–F β -strands, with detached A and G strands. Contacts between strands C and F are visible on maps (obtained by the BioShell package [16, 17]) in Figure 3(a)–(c), where the probability of contact for each pair of residues (calculated over a relevant part of a trajectory) was plotted for the native-like population (panel a) and conformations from the two first peaks in Figure 1 (panels b, c). It is also clearly evident from Figure 1 that the structures corresponding to the first two peaks are mainly stabilized by native interactions. This is perhaps why both peaks located

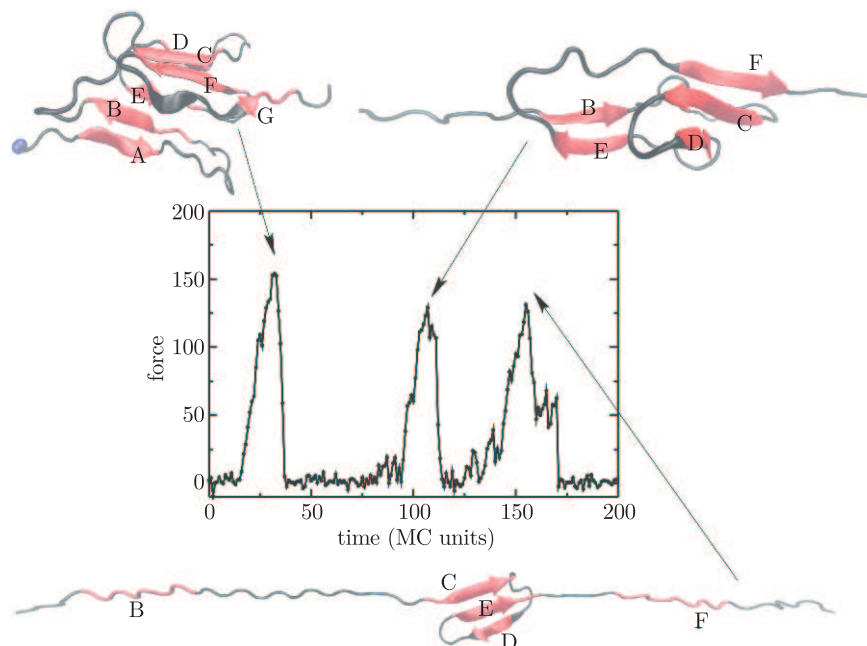


Figure 1. An example of force-time profile for DDFLN4 protein obtained using coarse-grained force-field; arrows refer to positions of three peaks which are the signatures of the transition states

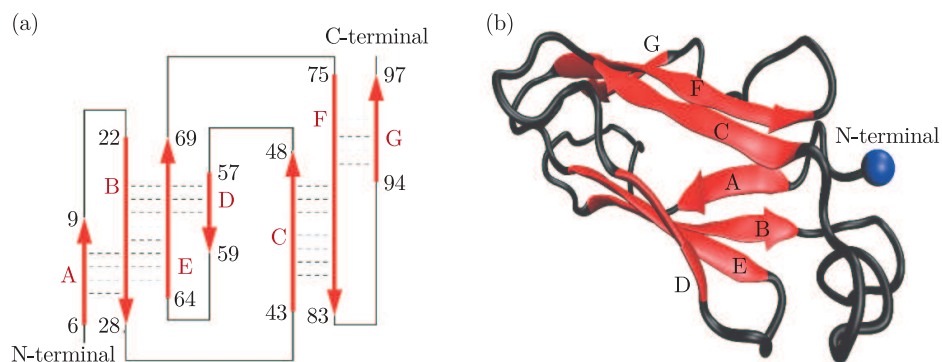


Figure 2. (a) Schematic plot showing native state backbone contacts (dashed lines) between β -strands; (b) cartoon representation of native state of DDFLN4 protein (PDB code: 1ksr) with seven β -strands labeled: A(6-9), B(22-28), C(43-48), D(57-59), E(64-69), F(75-83), G(94-97); N-terminal is marked by blue circle

at ~ 2 nm and ~ 12 nm might be encountered by the simple G \bar{o} -model, which is based solely on the topology of the native conformation. The reason behind the inability of the G \bar{o} -model to detect the third peak (or third transition state) located at ~ 22 nm is discussed below.

A contact map (Figure 3d) as well as a typical snapshot (Figure 1, bottom) corresponding to the last peak demonstrate that the structure of the third

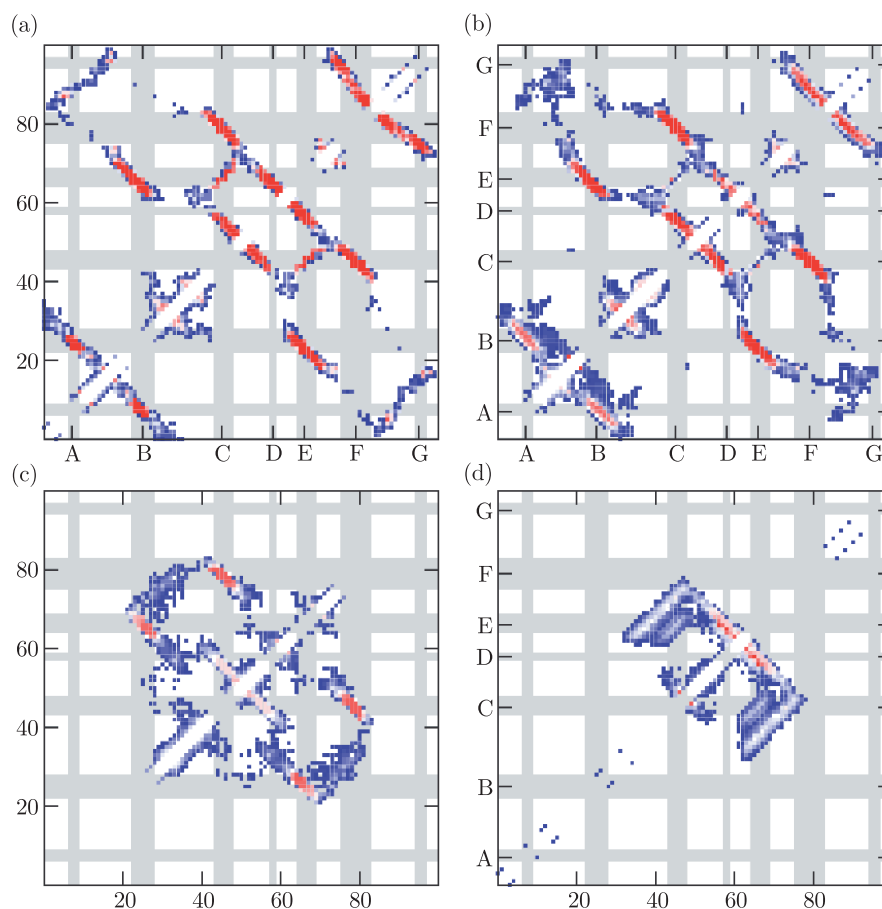


Figure 3. Contact probability maps (low to high marked as blue and red, respectively) for (a) the native population and the three peaks observed in the Figure 1 (b to d, respectively). Gray stripes mark residues that belong to a one of the β -strands, marked from A to G, respectively.

transition state involves the formation of hydrogen bonds between strands C and E, as well as E and D. Since the hydrogen bonds between strands C and E are not present in the native conformation (Figure 2a), the third peak comes from non-native interactions. This result is in agreement with the previously published all-atom simulations results [1] and explains the inability of the $G\bar{o}$ -model (where non-native interactions are omitted) to reproduce a complete picture of free energy landscape. Namely, the third intermediate state is not reproduced by the $G\bar{o}$ -model simulations due to the lack of non-native interactions. Furthermore, CABS the knowledge based force-field predicts the existence of the peak close to the native state. However, it is not clear why this maximum was not detected by AFM experiments [3] and further experimental studies are required to explain this issue. Interestingly, in case of I27 titin, the first intermediate state was skipped in one of the first AFM experiments [18]. A “hump” on the force-

extension profile which is a signature of the mechanical unfolding intermediate was then predicted by molecular dynamics simulations [5], confirmed experimentally later [19].

3. Conclusions

The CABS model has been already extensively tested in simulations of protein dynamics including: protein folding mechanisms from denatured to native state [20–23], near-native dynamics of globular proteins [24, 25] or a mechanism of folding and binding of intrinsically disordered protein [26]. These studies have shown that the CABS model is capable of reproducing complex details of protein folding dynamics. Importantly, a knowledge-based CABS force-field retains both native and non-native interactions (unlike in structure-based models, where only native interactions are taken into account [27, 28]). Therefore, sometimes the important role of non-native interactions, for example in protein folding [29, 30], unfolding [10, 31] and binding processes [32] is not neglected in CABS-based simulations.

In present paper we have tested a coarse-grained CABS model in the investigation of mechanical unfolding behaviour of a DDFLN4 protein. Using a CABS model we have shown that the non-native interactions have led to an additional intermediate state along a mechanical unfolding pathway, which was previously detected in the AFM experiments [3, 4] and all-atom MD simulation in explicit solvent [1], but not in a Gō-model [2]. The prediction of the first peak close to the native state by the CABS model is in agreement with the Gō-model and all-atom simulations, but not with the AFM experiment. Since the existence of the first peak does not depend on the choice of a force-field, we speculate that our results may describe the physics that governs the unfolding process. However, the partial disagreement between three different qualitatively different simulation techniques and the experiment calls for further investigation. Nonetheless, the presented simulation approach is a promising and efficient alternative to mechanical unfolding studies.

Acknowledgements

We would like to acknowledge the support from the Foundation for the Polish Science TEAM project (TEAM/2011-7/6) cofinanced by the European Regional Development Fund operated within the Innovative Economy Operational Program and from Polish Ministry of Science and Higher Education Grant No. IP2012 016872. DG has been supported by the Polish National Science Centre (NCN), Grant No. DEC-2011/01/D/NZ2/07683.

References

- [1] Kouza M, Hu C K, Zung H and Li M S 2009 *J. Chem. Phys.* **131** 215103
- [2] Li M S and Kouza M 2009 *J. Chem. Phys.* **130** 145102
- [3] Schwaiger I, Kardinal A, Schleicher M, Noegel A A and Rief M 2004 *Nat. Struct. Mol. Biol.* **11** 81
- [4] Schwaiger I, Schlierf M, Noegel A A and Rief M 2005 *EMBO Rep.* **6** 46

- [5] Lu H, Isralewitz B, Krammer A, Vogel V and Schulten K 1998 *Biophys. J.* **75** 662
- [6] Paci E and Karplus M 2000 *Proc. Natl. Acad. Sci. USA* **97** 6521
- [7] Arad-Haase G, Chuartzman S, Dagan S, Nevo R, Kouza M, Mai B, Nguyen H, Li M S and Reich Z 2010 *Biophys. J.* **99** 238
- [8] Yew Z T, Krivov S and Paci E 2008 *J. Phys. Chem. B* **112** 16902
- [9] Sulowska J I, Kloczkowski A, Sen T, Cieplak M and Jernigan R 2008 *Proteins* **71** 45
- [10] Kouza M, Hu C K, Li M S and Kolinski A 2013 *J. Chem. Phys.* **139** 65103
- [11] West D, Olmsted P and Paci E 2006 *J. Chem. Phys.* **124** 154909
- [12] Gront D, Kmiecik S and Kolinski A 2007 *J. Comput. Chemistry* **28** 1593
- [13] Kmiecik S, Gront D, Kouza M and Kolinski A 2012 *J. Phys. Chem. B* **116** 7026
- [14] Kolinski A 2004 *Acta Biochimica Polonica* **51** 349
- [15] Jamroz M, Kolinski A and Kmiecik S 2014 *Protein Structure Prediction (3rd edition), Methods in Molecular Biology*, Kihara D (ed.) **1137** 235
- [16] Gront D and Kolinski A 2006 *Bioinformatics* **22** 621
- [17] Gront D and Kolinski A 2008 *Bioinformatics* **24** 584
- [18] Rief M, Gautel M, Oesterhelt F, Fernandez J M and Gaub H E 1997 *Science* **276** 1109
- [19] Marszalek E P, Lu H, Li H, Carrion-Vazquez M, Oberhauser A F, Schulten K and Fernandez J M 1999 *Nature* **402** 100
- [20] Kmiecik S and Kolinski A 2007 *Proc. Natl. Acad. Sci. USA* **104** 12330
- [21] Kmiecik S and Kolinski A 2008 *Biophysical Journal* **94** 726
- [22] Kmiecik S and Kolinski A 2011 *J. Am. Chem. Soc.* **133** 10283
- [23] Wabik J, Kmiecik S, Gront D, Kouza M and Kolinski A 2013 *Int. J. Mol. Sc.* **14** 1
- [24] Jamroz M, Orozco M, Kolinski A and Kmiecik S 2013 *J. Chem. Theory Comput.* **9** 1
- [25] Jamroz M, Kolinski A and Kmiecik S 2014 *Bioinformatics* **30** 15
- [26] Kurcinski M, Kolinski A and Kmiecik S 2014 *J. Chem. Theory Comput.* **10** 2224
- [27] Go N 1983 *Ann. Rev. Biophys. Bioeng.* **12** 183
- [28] Clementi C, Nymeyer H and Onuchic J N 2000 *J. Mol. Biol.* **298** 937
- [29] Lewandowska A, Oldziej S, Liwo A and Scheraga H A 2010 *Biopolymers* **95** (3) 469
- [30] Faisca P F N, Nunes A, Travasso R D M and Shakhnovich E I 2010 *Protein Science* **19** 2196
- [31] Forman J R, Yew Z T, Qamar S, Sandford R N, Paci E and Clarke J 2009 *Structure* **17** (12) 1582
- [32] Yongqi H and Zhirong L 2010 *PLoS ONE* **5**, e15375