CTMS Fellowship Report

Xiancheng Zeng[*], Piotr E. Marszalek[†‡], Weitao Yang[*‡]

[*]Department of Chemistry, [†]Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708
[‡]Principle investigators

Project 1. Equilibrium Sampling for Biomolecules under Mechanical Tension
(Published on *Biophysical Journal*, 98, **2010**, 733–740; Presented in oral form in 238th ACS national meeting in 2009, Washington D.C.)

In the studies of force-induced conformational transitions of biomolecules, the large time-scale difference from experiments presents the challenge of obtaining convergent sampling for molecular dynamics simulations. To circumvent this fundamental problem, an approach combining the replica-exchange method[1] and umbrella sampling (REM-US) is developed to simulate mechanical stretching of polysaccharides under equilibrium conditions. Equilibrium properties of conformational transitions can be obtained directly from simulations without further assumptions. To test the performance, we carried out
REM-US simulations of atomic force microscope (AFM) stretching and relaxing measurements on the polysaccharide pustulan, a (1→6)-β-D-glucan, which undergoes well-characterized rotameric transitions in the backbone bonds.[2,3] With significantly enhanced sampling convergence and efficiency, the REM-US approach closely reproduced the equilibrium force-extension curves measured in AFM experiments. Consistent with the reversibility in the AFM measurements, the new approach generated identical force extension curves in both stretching and relaxing simulations, an outcome not reported in previous studies, proving that equilibrium conditions were achieved in the simulations. REM-US may provide a robust approach to modeling of mechanical stretching on polysaccharides and even nucleic acids.

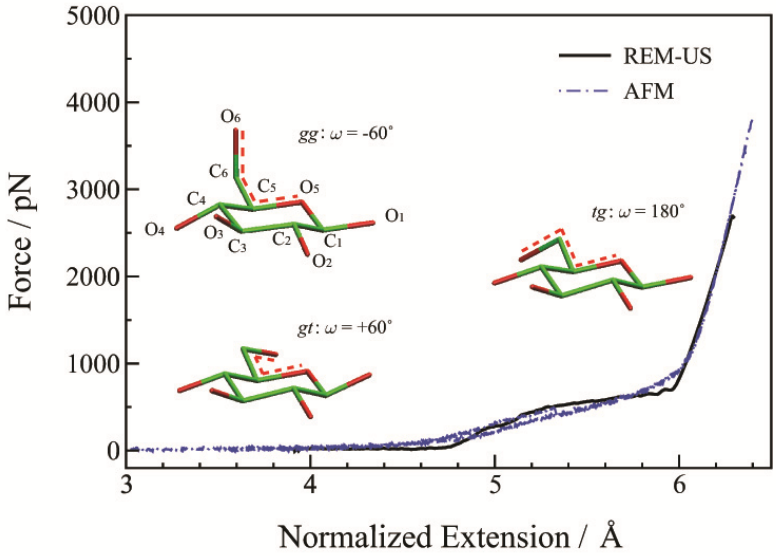

Figure 1. Comparison of force spectra obtained from the REM-US simulations (black solid line) and AFM experiments (blue dash-dot lines). Three major conformations involved in the stretching process are shown in the plot, corresponding to the rotamers: gauche-gauche (gg), with the dihedral angle O6-C6-C5-O5 (denoted by the red dash lines in the molecules) ω = −60°; gauche-trans (gt), with ω = +60°; and trans-gauche (tg), with ω = 180°.

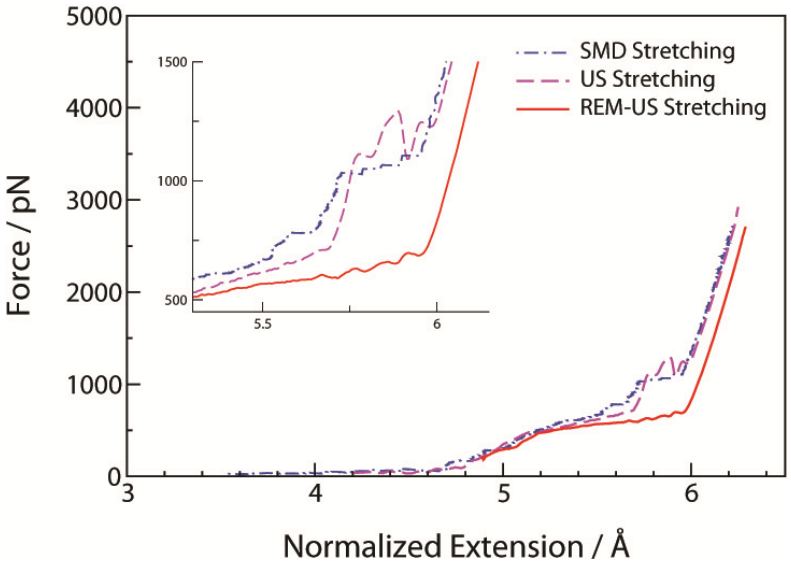

Figure 2. Force spectra of the stretching process obtained using SMD (blue dash-dot line), umbrella sampling (magenta dash line) and REM-US (red solid line). The inset is the zoom-in of the region 5.5-6.0 Å, to better demonstrate the discrepancies of the force among different methods.

Project 2. Full Reconstruction of a Vectorial Protein Folding Pathway: Insights into the Co-translational Folding of the Nascent Polypeptide Chain (Submitted; Accepted as an oral presentation in 240th ACS national meeting in 2010, Boston.)

Another promising application of AFM experiment is its ability to probe protein folding pathway. As recently suggested by Cabrita et al.[4], the vectorial character of co-translational folding, is in a way mimicked by force-induced unfolding experiments. During mechanical unfolding and refolding, the N- and C-termini of the polypeptide chain are constrained to the pulling axis, which limits the conformational space of the chain in a vectorial fashion. In order to examine the conformational transitions in detail, we need to perform simulations of the force-induced protein folding and unfolding. However, due to the extraordinary vast conformational space of protein, the all-atom force field is no longer affordable to describe efficiently the large-amplitude protein motions. Although attempts have been made to study similar systems with all-atom details[5], the sampling convergence remains a major concern in all-atom simulations of protein folding. Therefore, various approaches have been developed to model proteins at a coarse-grained (CG) level in order to efficiently deal with the large-amplitude motions in protein folding[6-10]. The CG methods simplify the protein structures and interactions (protein-protein and protein-solvent) and use one or few particles to describe each amino residue instead of all-atom details, thus, it greatly reduced the dimension of the conformational space.

By combining CG model and steered molecular dynamics simulations to unfold and refold single proteins under mechanical control, we examine protein folding in the presence of 1D constraint, similar to those imposed on the NPC. As shown in Fig. 3, the simulations exquisitely reproduced the experimental unfolding and refolding force spectra, and led to the full reconstruction of the

vectorial folding pathway of a large polypeptide, the 253-residue consensus ankyrin repeat protein, NI6C.

The remarkable, hitherto unattained accuracy, with which SMD simulations recreated the experimental force spectrograms allowed us to confidently reconstruct the vectorial folding pathway of NI6C, which is summarized in Fig. 4. First, our results indicate that the folding of NI6C under geometrical constraints is hierarchical and does not follow a simple two-state process indicated in bulk folding measurements[11]. Second, the results from SMD simulations clearly demonstrate that the folding rate limiting step in the vectorial folding process of the entire NI6C protein involves the nucleation of three N terminal repeats. We propose that the vectorial folding pathway reconstructed here for NIC6 may be representative of the co-translational folding of thousands of ankyrin repeat domains and other repeat proteins that form similar extended structures composed of stacked alpha helical repeats. The combination of AFM measurements with structure-based CG simulations presents a powerful tool for advancing our understanding of co-translational protein folding.

Acknowledgements

All the above work would not have been possible without the support from CTMS. Both the encouragement in cross-disciplinary research and the sponsorship helped me working on these exciting projects, which integrate theories and experiments, combining chemistry, biophysics and mechanical engineering.

Comments on "Market Place of Ideas"

It have been a great fun for me to join the "Market Place of Ideas" lunch seminars. The great variety of topics introduces many new ideas and thoughts to the community; however, the broadness sometimes leads to weak attractions to people working in a complete different area. I am thinking if CTMS could organize the lunch seminars in a series talks with different sessions. One session (per month, for instance) could contain few seminars with broad but still related topics. Of course, I fully understand that this will cause additional works and burden to the organizers and may not be feasible at all.

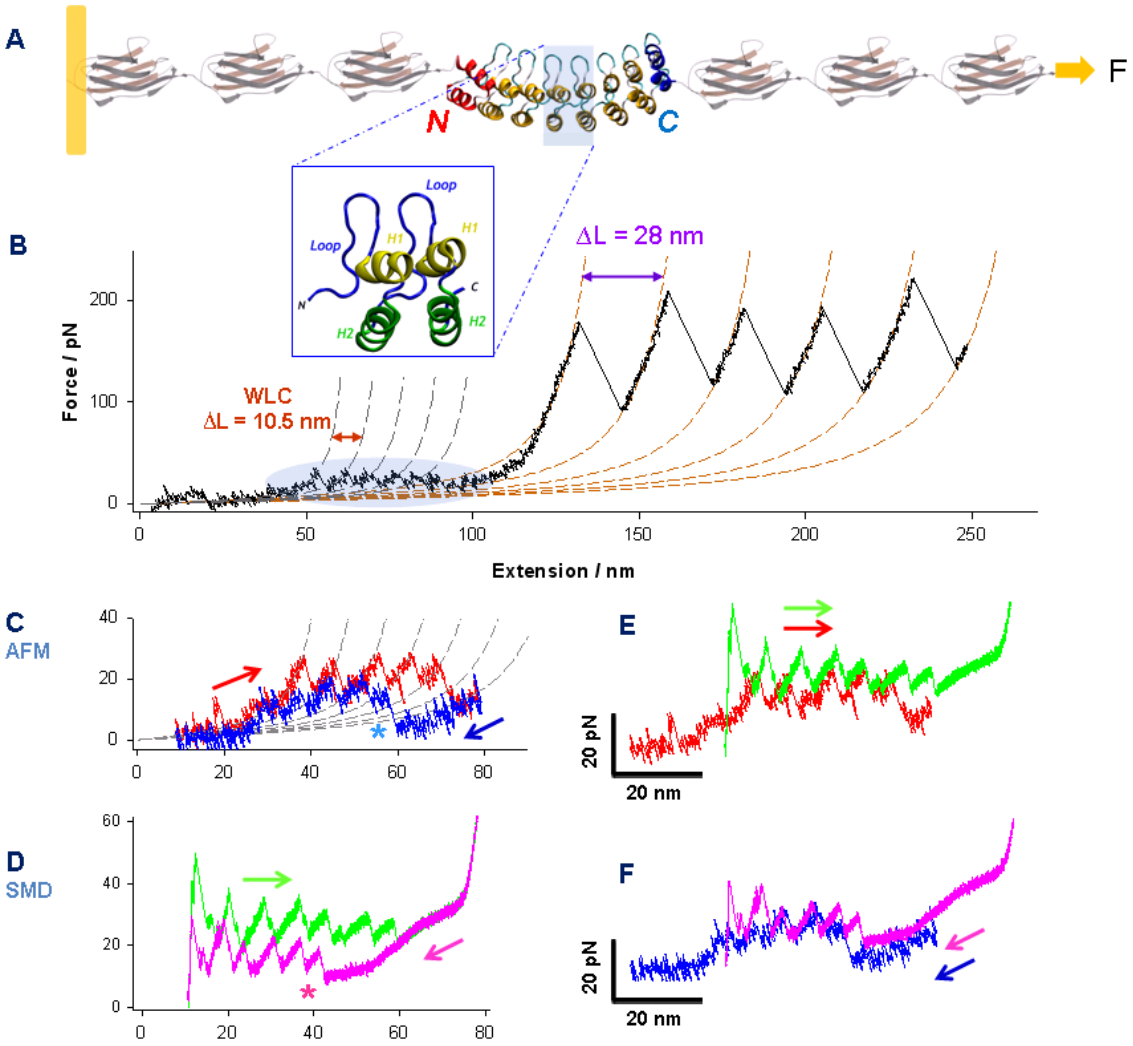

Fig. 3. Complete mechanical unfolding and refolding of the NI6C-I27 chimeric protein.
(*A*) The structure of the chimeric polyprotein designed for the AFM experiments. NI6C contains 8 ARs[12]. Two internal repeats are shown in the blue box, each composed of two α-helices (H1 and H2) and a loop. The full sequence is shown in Fig. S1*D*. (*B*) A representative unfolding trace of NI6C-I27 obtained at a stretching speed of 100 nm/s. The AFM data is fitted to two families of WLC curves, one with a contour length increment ΔL = 10.5 nm and a persistence length p = 0.78 nm (gray dash lines) and the other with ΔL = 28 nm and p = 0.36 nm (orange dash lines). These values of ΔL are consistent with the fully stretched lengths of one AR and one I27 domain, respectively; hence the two families of peaks correspond to the unfolding of individual ARs of NI6C and of I27 domains, respectively. (*C*) The measured unfolding (red) and refolding (blue) force-extension traces of NI6C at 30 nm/s, fitted to a family of WLC curves (grey dash lines) with ΔL = 10.5 nm and p = 0.86 nm. Note that following the complete unfolding of all ARs, the first refolding force peak appears only after the molecule was partially relaxed (star). (*D*) The simulated unfolding (green) and refolding (pink) force-extension traces of NI6C. Similar to the AFM data, the first SMD refolding force peak occurs only after the molecule has been significantly relaxed (star). (*E*) A comparison of the unfolding force-extension traces by SMD (green) and AFM (red). The SMD trace is shifted to the right by 20 nm to compensate for the initial length of I27 modules that contribute to the extension in the AFM measurements but are absent in the SMD simulations. (*F*) A comparison of the refolding force-extension traces by SMD (pink) and AFM (blue) following the unfolding of NI6C. The same 20 nm shift is applied to the SMD trace.

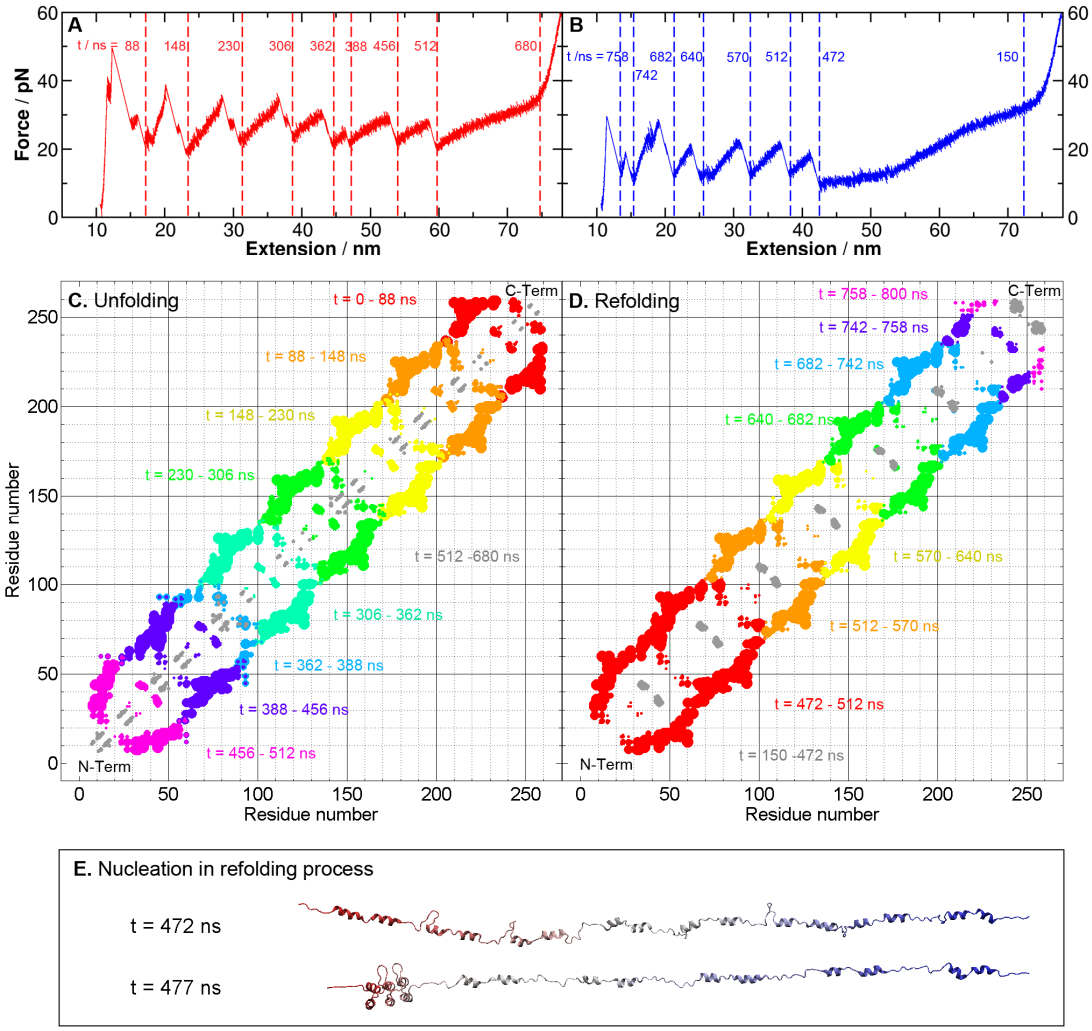

Fig. 4. Unfolding and refolding sequences of NI6C.
(*A*) The simulated complete unfolding force-extension traces of NI6C with timestamps marked between each major force peaks, indicating the breaking of a tertiary structure. (*B*) The simulated complete refolding force-extension traces of NI6C, with timestamps. (*C*) Changes of the distance between native contacts during the complete unfolding process, with color code from red to gray representing different time-spans defined in *A*. For each contact, a varying size of dot is used. The size, *S*, of a dot is proportional to the change in the distance between the residues of an original native contact within a certain time-span, i.e., $S=c[r(t_2)-r(t_1)]$, where $r(t)$ is the residual distance at time t, and c is a scaling constant. This change in contact distance clearly shows that the breakup of native contacts during the mechanical unfolding proceeded from the C- to N-terminus. The gray regions represent the unfolding of local α-helices (H1 and H2), corresponding to the plateau between 512 ns to 680 ns in *A*. (*D*) Changes of native contact distance during the complete refolding process, with color coded time-spans. From t = 150 ns to 472 ns, local structures were formed (shown by the gray regions), which were followed by the simultaneous folding of three repeats at the N-terminus. This nucleation event produced the first force peak that appeared between t = 472 ns to 512 ns in *B*. (*E*) The snapshots of the NI6C structures before and after the nucleation event with timestamps. Three N-terminal repeats folded within 5 ns and formed a nucleation core, which facilitated the folding of the rest of the polypeptide chain.

References:

1 Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141-151, (1999).

2 Kuttel, M., Brady, J. W. & Naidoo, K. J. Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *J. Comput. Chem.* **23**, 1236-1243, (2002).

3 Lee, G., Nowak, W., Jaroniec, J., Zhang, Q. & Marszalek, P. E. Molecular Dynamics Simulations of Forced Conformational Transitions in 1,6-Linked Polysaccharides. *Biophys. J.* **87**, 1456-1465, (2004).

4 Cabrita, L. D., Hsu, S. T. D., Launay, H., Dobson, C. M. & Christodoulou, J. Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 22239-22244, (2009).

5 Sotomayor, M., Corey, D. P. & Schulten, K. In Search of the Hair-Cell Gating Spring: Elastic Properties of Ankyrin and Cadherin Repeats. *Structure* **13**, 669-682, (2005).

6 Go, N. THEORETICAL-STUDIES OF PROTEIN FOLDING. *Annu. Rev. Biophys. Eng.* **12**, 183-210, (1983).

7 Liwo, A. *et al.* A united-residue force field for off-lattice protein-structure simulations .1. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.* **18**, 849-873, (1997).

8 Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins. *J. Mol. Biol.* **298**, 937-953, (2000).

9 Izvekov, S. & Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **109**, 2469-2473, (2005).

10 Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111**, 7812-7824, (2007).

11 Mello, C. C. & Barrick, D. An experimentally determined protein folding energy landscape. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 14102-14107, (2004).

12 Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K. & Plückthun, A. Folding and Unfolding Mechanism of Highly Stable Full-Consensus Ankyrin Repeat Proteins. *J. Mol. Biol.* **376**, 241-257, (2008).