

## **Identification and Characterization of the Unfolding Transition State of NTL9**

Christopher N. J. Marai\* and Jin Wang\*\*

\*Graduate Program in Biochemistry and Structural Biology, \*\*Department of Chemistry and Department of Physics

### **Background:**

L9 is a two domain protein of the large ribosomal subunit. The N-terminal domain of L9 forms a small mixed alpha-beta structure (Figure 2).[1-3] The protein fragment consists of a three-stranded antiparallel beta-sheet surrounded by two alpha-helices. The second helix is an extension of the interdomain connector in the intact protein.[4] The three-stranded sheet and the first helix comprise the first 39 residues of the N-terminal domain. This structure constitutes what is the smallest and probably the simplest example of the split beta-alpha-beta fold or ABCD motif. [1,2] Moreover, the split beta-alpha-beta motif is found as substructure in a large number of proteins[5]. The N-terminal domain that constituted 56 residues has been solved by NMR and X-ray crystallography.[1,2] This domain of L9 has been shown to fold independent of the rest of the L9 structure. Finally, NTL9 folds quickly in a 2-state fashion under a wide variety of experimental conditions. [2]

Experimental work on by our collaborators has revealed tremendous structural information about the native state and folding NTL9.[1-4] In spite of this significant experimental progress, atomic level details of the unfolding path remain unknown and have yet to be addressed by computational investigation. Due to its compact size and excellent experimental characterization it is an ideal system to study computationally.

### **Current Progress:**

Current efforts aim to sample the folding pathway of NTL9 with high temperature all-atom MD simulations in implicit solvent. This method has been used successfully in the literature to study the folding and transition state ensembles of other similar sized proteins.[6] We investigated the unfolding path of NTL9 by unfolding the protein independently 100 times for 500ps using all-atom MD implicit solvent. Unfolding was induced by 350K and 400K temperatures. The free energy profile for a series of reactions coordinates was determined from data of 100 unfolding runs and indicates NTL9 folds in a two-state manner. This result is consistent with known experimental evidence of NTL9 folding.[1-4] Moreover, our preliminary calculations indicate specific non-native interactions in the denatured state as suggested by experiment.[7] The current methods were implemented on 10, 500Mhz nodes on the serial node cluster in the department of chemistry (called RAM). Progress has since been limited by CPU time and additional CPU time will be required to uncover some of the questions in our aims as listed below.

**Aims:****What is the 3D structure of the transition state (TS) of unfolding?**

Refolding and mutational experiments have revealed macroscopic details of the transition state ensemble (TSE) of NTL9. In the past such results coupled with molecular dynamics simulations have been used to determine an atomic level description of the transition state ensemble of other proteins.[6] Currently, the experimentally dependent Phi-RMSD method[6] has been implemented in discovering TSE structures. This however provides a somewhat ambiguous description of the TSE by matching best calculated phi-values with experimentally derived phi-values. Structures determined have measured characteristics of the TSE but further analysis is required to determine if they are indeed at the TS.

The free energy surface derived structures in the saddle point region are suggestive of the TSE but are unconfirmed and may be altered by the conditions (high temp) of the simulation. Furthermore, this is highly dependent on the choice of reaction coordinate. Ideally, these results (from many reaction coordinates) should be further refined to determine if the saddle point structures are indeed TS structures, and not an artifact of simulation environment or choice of reaction coordinate.

**Should the denatured state be included in calculation of phi-values?**

On going experimental work by the Raleigh group has highlighted that conventional phi-value analysis maybe limited in its effort to identify the importance of residues on the folding pathway.[7] Specifically, current calculation of phi-values requires mutational analysis of amino acids and considers changes between the native state and the TS. This method, however, fails to take into account any change of the mutation on the unfolded state. Currently, the phi-RMSD method has been implemented with and without the denatured contacts to support the need for denatured contacts. This method is limited however by the identification of the TS. Further work is needed to refine our choice of TS allowing, better phi-RMSD comparisons as well as “ab initio” calculation of phi-values.

**Future Plans:**

Current methods have provided us with a rough idea of the TSE of NTL9, however further work is needed to confirm the TSE. Moreover, our predictions of the TS structures thus far have depended on experimental data. An “ab initio” free energy derived TSE would constitute a predictive methodology for understanding the TSE of other proteins as well as NTL9. Furthermore, this method could also be used to calculate “ab initio” phi-values for comparison to experiment.

We intend to use granted computer time to validate our TSE by P-fold analysis.[6] By this method our predicted TS structures (from both the FE profile and phi-RMSD calculations) will be run at 300K in implicit solvent and let run until folded or unfolded. P-fold is highly dependent upon the number of runs carried out thus, we currently plan to fold our top 10 TS structures, from our phi-RMSD calculations as well as our FE profiles. Each of these top 10 structures will undergo 10 folding simulations of 100ns with the Amber FF03 and GB implicit solvent. It is estimated that each folding simulation of 100ns will take about 2weeks on single CPUs but would also benefit from the parallel setup of the Seawolf cluster. Benchmarking is still necessary however it is estimated about 25000 CPU hours/month will be required.

## References:

1. Kuhlman B, Yang HY, Boice JA, Fairman R, Raleigh DP. An exceptionally stable helix from the ribosomal protein L9: implications for protein folding and stability. *J Mol Biol.* 1997 Aug 1;270(5):640-7.
2. Kuhlman B, Boice JA, Fairman R, Raleigh DP. Structure and stability of the N-terminal domain of the ribosomal protein L9: evidence for rapid two-state folding. *Biochemistry.* 1998 Jan 27;37(4):1025-32.
3. Vugmeyster L, Kuhlman B, Raleigh DP. Amide proton exchange measurements as a probe of the stability and dynamics of the N-terminal domain of the ribosomal protein L9: comparison with the intact protein. *Protein Sci.* 1998 Sep;7(9):1994-7.
4. Lillemoen, J., Cameron, C. S., and Hoffman, D. W. Stability and dynamics of ribosomal protein L9: An investigation of a molecular strut by amide proton exchange and circular dichroism. *J. Mol. Biol.* 1997, 268, 482-493.
5. Orengo, C. A., and Thornton, J. M. Alpha-fold plus beta-fold revisited - some favored motifs. *Structure* 1993 1, 105-120.
6. Settanni G, Rao F, Caflisch A. Epub 2005 Jan 11. Phi-value analysis by molecular dynamics simulations of reversible folding. *Proc Natl Acad Sci U S A.* 2005 Jan 18;102(3):628-33.
7. Luisi DL, Wu WJ, Raleigh DP. Conformational analysis of a set of peptides corresponding to the entire primary sequence of the N-terminal domain of the ribosomal protein L9: evidence for stable native-like secondary structure in the unfolded state. *J Mol Biol.* 1999 Mar 26;287(2):395-407.