**In-situ Data Analysis of Protein Folding Trajectories**

Travis Johnston\(^1\), Boyu Zhang\(^2\), Adam Liwo\(^3\), Silvia Crivelli\(^2\), Michela Tauffer\(^1\)

\(^1\)U. Delaware Global Computing Lab, \(^2\)U. Gdansk, Poland, \(^3\)Lawrence Berkeley National Lab

---

**Motivation**

The transition from petascale to exascale computing will be accompanied by substantial changes in computer architectures and technologies. The research community relying on computational simulations is being forced to revisit the algorithms for data generation and analysis due to various concerns, such as higher degrees of concurrence, deeper memory hierarchies, substantial I/O and communication constraints. Simulations today typically save all data for post simulation analysis. Simulations at the exascale will require us to analyze data as it is generated and save only the results that enhance our scientific understanding. The analysis of this data will need to primarily be accomplished in-situ, i.e. executed sufficiently fast locally, using very limited amounts of memory and disk space, and avoiding large data movement.

**Increasing Peak Performance**

**Stagnant I/O Bandwidth**

---

**Method for In-situ Data Analysis**

1. Discretize the protein by extracting the positions of α-Carbons and β-Carbons
2. Build a Euclidean Distance Matrix, D, (single structure) or a Bipartite distance matrix (two structures)
3. Associate the largest eigenvalue of D with the protein conformation as metadata
4. Use the metadata to identify stable and transition states during the simulation

---

**Algorithm**

- **1EOL: Single β-Strand**
  - Meta-stable Transition
  - Eigenvalues computed with only single frame information
  - RMSD computed with a priori knowledge of folded structure
  - Formation of α-helix

- **1BDD: Single α-Helix**
  - Entire Protein
  - Coarse change is protein conformation
  - Individual α-helices stable

- **1BDD: Pair of α-Helices**

---

**Conclusions**

We propose a novel method for in-situ data analysis of protein folding trajectories. We validate our metadata mapping method by applying it to two 40k frame trajectories: one trajectory of 1BDD (containing 3 β-strands). Our metadata mapping enabled us to observe metastable states, transition states, the formation of an individual substructure and the repositioning of substructure relative to others.

---

**Acknowledgements:**

NSF Grant #1318445