High Performance High Throughput Computing of Positioned Nucleosomes on LONI

Rajib Mukherjee[1], Hideki Fujioka[2], Thomas C Bishop[1]

1. Departments of Physics and Chemistry, Louisiana Tech University
2. Center for Computational Science, Tulane University

Chromatin folds the genetic code such that it fits inside a cell nucleus. The basic building block of chromatin is the nucleosome, 147 base pairs of DNA wrapped 1.7 turns around a core of eight histones. The DNA base pairs can be of any sequence, but the affinity of the histone core for different sequences of DNA varies. Thus, histones occupy preferred locations on a length of DNA. This phenomenon is called nucleosome positioning.

In the present work, all atom molecular dynamics simulations are performed to investigate the structure and dynamics of six positioned nucleosomes as a function of DNA sequence. Each system contains hundreds of thousands of atoms and is simulated for 80 ns. A python based tool called ManyJobs is used for managing the simulations, which are run as 80 successive 1 ns tasks. ManyJobs ensures that each task runs on whichever of five LONI supercomputers (Painter, Poseidon, Eric, Louie and Oliver) becomes available first. The total study generates terabytes of data that is stored on a central fileserver, e.g. PetaShare or a local fileserver. Each simulation task retrieves inputs from the central fileserver and deposits outputs to it, while ManyJobs maintains a database of which tasks have completed and any dependencies between tasks. In this way proper data management is performed. For analysis, the data is extracted from the central database. Analysis focuses on essential dynamics of the DNA helical parameters as a means of comparing the trajectories. We relate the formation of kinks in DNA to eigenvectors with only a few active, narrowly distributed components. We also identify several eigenvectors that are present to a large extent in all nucleosomes and that persist for the duration of each simulation. In this manner we have identified what is common and what is unique to this set of positioning sequences.