MINING THE EVOLUTIONARY DYNAMICS OF PROTEIN LOOP STRUCTURE AND ITS ROLE IN BIOLOGICAL FUNCTIONS

Executive Summary:

Flexible and unstructured regions of protein molecules introduce a source of conformational heterogeneity that is fundamental for their biological function. Here we study this heterogeneity with microsecond-scale molecular dynamics simulations using the NAMD 2.9 platform. Previous studies have shown that the dipeptide make-up of proteins delimits loop regions that are unstructured and flexible, driving the specificity and stereochmistry of the genetic code. A previous Blue Waters allocation confirmed that unstructured regions collapse quickly, but unexpectedly revealed increased intra-chain dynamics of short fragments. Our research provided unprecedented atomistic details of the dynamics of 74 loop regions of protein domains sampled along a timeline of domain history, studying collapse propensities and intrinsic fluctuations of each loop structure. Our study provides insight into how loop flexibility and disorder are linked to the genetic code and primes protein function in evolution.

Methods & Results

In previous work, we discovered that the speed of folding, which correlates with flexibility, is enhanced during the evolution of protein domains [5]. We also discovered that protein structures enriched with flexible loops appeared with the evolutionary unfolding of the genetic code [6]. Since structural flexibility is a conserved feature in the assembly of protein complexes [7], flexibility must be regarded as an emergent property of molecular evolution. Following an exploratory Blue Waters proof-of-concept allocation [8], we used the power of Blue Waters to study the dynamics of proteins loops in 74 protein domains of aminocyl-tRNA synthetases, the enzymes responsible for the specificity of the genetic code. Advanced tools of molecular dynamics (MD) simulations using the NAMD 2.9 platform allowed gathering global parameters (RMSD, radius of gyration) and detailed mapping of motions linked to loop sequence and structure. Each equilibrium all-atom simulation required ~0.005 Mnh. Benchmarking indicates 5 Mnh are needed to cover the target of 1,000 representative loops.

Remarkably, we found a trend of reduced global RMSD and radius of gyration during the entire timeline of protein history that indicates an evolutionary tendency towards conformational order, which we would like to confirm with additional analyses. We also found dynamic heterogeneities in both loops and surrounding regular structures that could be associated with specific functions. Our research links two fields of study that have not yet interfaced in science: (1) the evolution of molecular structure and intrinsic disorder in proteins, and (2) the molecular dynamics of proteins. The former focuses on evolutionary processes spanning billions of years of biological history. The latter looks at molecular change unfolding at nanosecond to microsecond levels. This interdisciplinary exploration is expected to uncover patterns of origin and evolution of the genetic code, protein structures, and functions.

Why Blue Waters?

The results of our initial exploration and benchmarking exercise provided a foundation for a Blue Waters-enabled high-throughput MD simulation study of the dynamics of a massive number of protein loops, which will be indexed with evolutionary, structural, and functional information. The study will yield unprecedented atomistic details of structural and functional evolutionary constraints that are responsible for structuring both proteins and the genetic code.

Publications