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

Allocation: Illinois/0.19 Mnh
PI: Gustavo Caetano-Anollés[1]
Co-PI: Frauke Gräter[2]
Collaborator: Fizza Mughal[1]

[1]University of Illinois at Urbana-Champaign
[2]Heidelberg Institute for Theoretical Studies, gGmbH

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 stereochemistry 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.

INTRODUCTION

Proteins sustain life on our planet, from major biogeochemical cycles necessary for planetary stability to crucial signaling in brain activities important for cognition and behavior. Protein misfolding results in aggregation and diseases such as Alzheimer's or Creutzfeldt–Jakob. Their challenge and deregulation causes pathogenesis and cancer. Despite proteins' importance, many aspects have yet to be uncovered; among them are how and why this sophisticated molecular machinery carries biological functions, the rationale for its change in time, the mysterious origin of the "vocabulary" that shapes genetics (the genetic code) and the evolutionary drivers of the structure of proteins.

Traditionally, the structure of proteins is defined by a "regular" structure, the 3D arrangement of helical (typically α-helices) and strand (typically β-strands) segments resulting from local and long-range interactions of the protein backbone, respectively [1]. In recent years, however, there has been growing recognition that "non-regular" structural regions that connect helices and strands play more important roles at functional and regulatory levels. These intra-chain regions are compact but lack fixed internal hydrogen bonding interactions. They include loops and intrinsically disordered regions. Protein loops, which are often present in active sites of enzymes or often enable molecular mechanics, have been of particular interest [2]. Their flexibility may provide a link between protein structure and function. Information about them has been collected in numerous databases, which index their topology, supporting fold structures, and associated gene ontology definitions of molecular functions [e.g., 3,4].

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 aminoacyl-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.

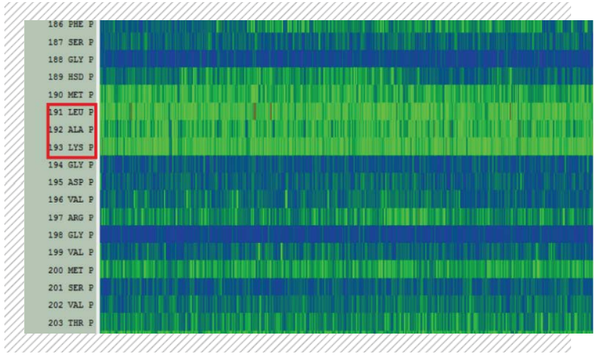

FIGURE 1: Root mean square fluctuation (RMSF) plot showing a heat map of dynamic fluctuations of each amino acid residue of loop 1JMZ_A_182, embedded in an OB-fold Myf domain, over time from left to right. Note the increased dynamics of the loop region (red square encompassing Leu, Ala, and Lys residues) and the dynamic heterogeneities of the van der Waals lock surrounding it.

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

Caetano-Anollés, G., et al., The dynamics of protein disorder and its evolution: Understanding single-molecule FRET experiments of disordered proteins. *Blue Waters Annual Report* (Urbana, Illinois, 2014), pp. 120–121.