TN

BW

MOLECULAR DYNAMICS BINDING FREE ENERGY CALCULATIONS OFFER A WINDOW TO UNDERSTAND PROTEIN–PROTEIN BINDING SPECIFICITY

Allocation: NSF PRAC/1,330 Knh
PI: Benoit Roux[1]
Collaborators: Chris Boughter[1], Prithviraj Nandigrami[1], Jonathan Thirman[1]

[1]University of Chicago

EXECUTIVE SUMMARY

Extracellular domains of cell surface receptors and ligands mediate cell–cell communication, adhesion, and initiation of signaling events. In order to understand highly organized cellular systems, it is necessary to study how intricate dynamical networks arise from specific protein–protein interactions. Thus, the research team has used a free energy computational methodology to quantitatively examine the basis for protein binding specificity in two subfamilies of recently identified genes within the immunoglobulin superfamily found in the model organism *Drosophila melanogaster* (the common fruit fly).

RESEARCH CHALLENGE

The living cell can be pictured as a collection of macromolecules that are carrying out a number of well-defined tasks. Thanks to tremendous progress made in X-ray crystallography, our knowledge of the three-dimensional structure of individual proteins present in the cell has been greatly expanded in the last decades. However, as scientists seek to comprehend the function of highly organized cellular systems, it is becoming increasingly clear that this will not be possible without understanding how intricate dynamical networks arise from specific protein–protein interactions. For example, the intercellular communications needed for the morphogenesis of the central nervous system is essentially mediated by protein–protein interactions and recognition processes involving cell surface receptors and ligands.

Some of the most intriguing of such interactions are presented by the set of recently identified Dpr–DIP complexes in the common fruit fly (*Drosophila melanogaster*), which has a total of 20 Dpr genes. The Dpr (defective proboscis extension response) and DIP (Dpr-interacting proteins) genes belong to the immunoglobulin superfamily and associate to form a molecular complex (Fig. 1). Most members of the DIP subfamily cross-react with several members of the Dprs and vice versa.

The Dpr–DIP binding specificity plays a critical role in neuronal development and synaptogenesis. The research team's objective is to use a free energy computational methodology to quantitatively explain the basis for Dpr–DIP protein binding specificity. To understand how cell surface receptors control developmental morphogenesis of the nervous system as well as its function, researchers need to answer questions such as how the distinct structural features of cell adhesion complexes, including possible sets of highly homologous proteins, instruct the formation of synaptic networks. In other words, scientists must decipher the molecular code that governs the specific association of these proteins and understand why some of them bind together but not others despite their high structural similarities.

Predicting with quantitative accuracy when and why proteins can specifically associate and bind must begin with the statistical mechanics concept of binding free energy. For a binary protein complex, the binding affinity is determined by the equilibrium dissociation constant and the binding free energy. Based on insight gained by categorizing and observing many known protein complexes, several models have been proposed to predict the experimental binding affinities using the features discussed above. Although some have been very successful on small training sets, the published models have performed considerably less well on larger sets and their predictive value remains poor. From this point of view, a computational approach based on atomistic molecular dynamics (MD) simulation and free energy methodology is advantageous because it does not rely on any particular empirical assumptions about the binding; the approach is applicable to any protein complex and does not suffer from the limitation displayed by empirical bioinformatics/coevolution or docking/scoring methods, which often perform poorly in large-scale benchmark testing. But this raises several important issues, such as whether the atomic force fields used in the simulations are sufficiently accurate, or whether one can design an efficient computational strategy to overcome the challenges presented by conformational sampling. Another important question is whether one can truly design an effective and scalable computational strategy to tackle the processes occurring over long timescales that is well adapted for a leadership-class computer such as Blue Waters. Unbiased MD trajectories, while very valuable, can be limited. However, advanced free energy methodologies can help overcome these limitations.

METHODS & CODES

The solution to this computational problem proposed by the research group is to break down the binding process into a large number of physically meaningful steps and express the binding free energy as a sum of free energies associated with each step.

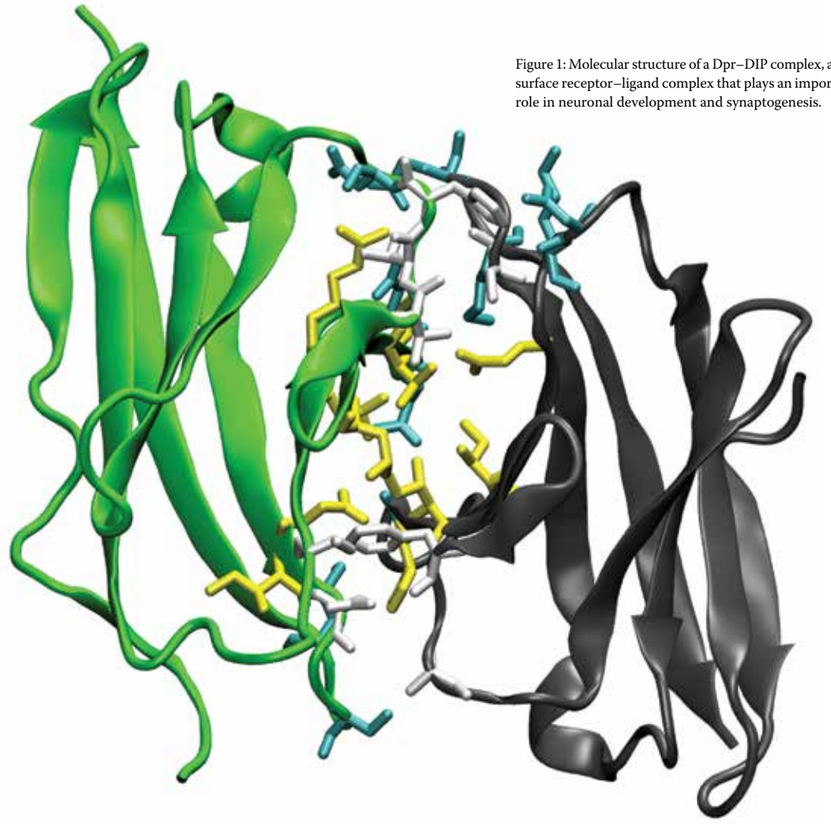

Figure 1: Molecular structure of a Dpr–DIP complex, a cell surface receptor–ligand complex that plays an important role in neuronal development and synaptogenesis.

This requires carrying out a large number of simulations with different biasing restraints acting on the system. The biased data are harvested from a large collection of copies or replicas of the molecular system via a replica-exchange algorithm. Such multiple copy algorithms (MCAs) offer a general and powerful strategy to enhance the sampling efficiency of conventional MD simulations. Further, the simulation program NAMD fully supports extremely scalable and efficient parallel active messaging interface-level simulations with MCAs on leadership computers. This makes it possible to carry out extremely scalable MD simulations on Blue Waters using the team's MD replica-exchange free energy strategy.

RESULTS & IMPACT

The research team has developed and tested a novel theoretical framework for binding free energy calculations, leaning on the optimal curvilinear minimum free energy path determined from the string method. Fundamentally, this curvilinear pathway strategy is based on the reversible spatial separation of two binding macromolecules, which is clearly the method of choice to quantitatively characterize the affinity of large molecular complexes in solution. Nonetheless, previous implementations based on the potential of mean force for the separation of two proteins along a predefined rectilinear path led to MD calculations that converged too slowly. The proposed methodology was validated by comparing the results obtained using both rectilinear and curvilinear pathways for a prototypical host–guest complex formed by cucurbituril binding benzene and for the barnase–barstar protein complex. The researchers found that the calculations following the traditional rectilinear pathway and the string-based curvilinear separation pathway agree quantitatively, but convergence is faster with the latter.

WHY BLUE WATERS

The multiple-copy algorithms require a very large number of nodes to be effective. Blue Waters is a unique platform that makes it possible to fully exploit the power of this advanced sampling methodology.