Blue Waters Professor allocation report and extension request

Diwakar Shukla
Department of Chemical & Biomolecular Engineering, University of Illinois at Urbana-Champaign.

The allocated computation time (305,000 node hours) on Blue Waters jt3 account for the period of April 1st, 2017 was used for the following projects. We finished the computation time on June 30th, 2017. Therefore, we request an additional ~540,000 node hours for the period March 1, 2018 to Feb 28, 2019.

1. Allosteric control of a plant receptor kinase through S-glutathionylation.

Growing evidence supports the importance of protein S-glutathionylation as a regulatory post-translational modification with functional consequences for proteins. Discoveries of redox-state-dependent protein kinase S-glutathionylation have fueled discussion of redox-sensitive signaling. Following previously published experimental evidence for S-glutathionylation induced deactivation of the Arabidopsis thaliana BRASSINOSTEROID INSENSITIVE 1 (BRI1)-ASSOCIATED RECEPTOR-LIKE KINASE 1 (BAK1), we investigated the consequences of S-glutathionylation on the equilibrium conformational ensemble of BAK1 using extensive all-atom molecular dynamics simulations. Using Markov state models to construct unbiased free energy landscapes of BAK1 in different S-glutathionylation states, we found that glutathionylation of C408 allosterically destabilizes the active-like state of BAK1 while stabilizing an inactive conformation known to recur in protein kinases. Using the Kullback-Leibler divergence between individual residue conformational distributions, we also found that S-glutathionylation of C408 has structural consequences throughout the BAK1 kinase domain mediated through interaction with a key regulatory helix, while glutathionylation of C353 in the N-lobe and C374 near the ATP binding site have few notable effects on BAK1 as compared to the unmodified protein. Our results suggest an allosteric mechanism for inhibition of BAK1 by C408 S-glutathionylation, and more generally, support the notion of protein kinase S-glutathionylation as a means of redox signaling in plant cells.

Publication:

Alexander S. Moffett, Kyle Bender, Steven Huber & Diwakar Shukla. **Allosteric control of a plant receptor kinase through S-glutathionylation**. Biophysical Journal, 2017, 113, 2354-2363

Alexander S. Moffett, Kyle Bender, Steven Huber & Diwakar Shukla. **Molecular dynamics simulations reveal the conformational dynamics of Arabidopsis thaliana BRI1 and BAK1 receptor-like kinases.** J. Biol. Chem, 2017, 292, 12643-12652

Alexander S. Moffett & Diwakar Shukla. **Using molecular simulation to explore the nanoscale dynamics of the plant kinome**. Biochemical Journal, 2018 (*in press*)

2. Understanding protein dynamics using evolutionary couplings.

Understanding of protein conformational dynamics is essential for elucidating molecular origins of protein structure-function relationship. Traditionally, reaction coordinates, i.e., some functions of protein atom positions and velocities have been used to interpret the complex dynamics of proteins obtained from experimental and computational approaches such as molecular dynamics simulations. However, it is nontrivial to identify the reaction coordinates a priori even for small proteins. Here, we evaluate the power of evolutionary couplings (ECs) to capture protein dynamics by exploring their use as reaction coordinates, which can efficiently guide the sampling of a conformational free energy landscape. We have analyzed 10 diverse proteins and shown that a few ECs are sufficient to characterize complex conformational dynamics of proteins involved in folding and conformational change processes. With the rapid strides in sequencing technology, we expect that ECs could help identify reaction coordinates a priori and enhance the sampling of the slow dynamical process associated with protein folding and conformational change.

Publication:
Jiangyan Feng & Diwakar Shukla. **Characterizing conformational dynamics of proteins using evolutionary couplings.** J. Phys. Chem. B, 2018, 122, 1017-1025.

3. Enhanced conformational sampling of proteins using evolutionary couplings.
One of the major challenges in atomistic simulations of proteins is efficient sampling of pathways associated with rare conformational transitions. Recent developments in statistical methods for computation of direct evolutionary couplings between amino acids within and across polypeptide chains have allowed for inference of native residue contacts, informing accurate prediction of protein folds and multimeric structures. In this study, we assess the use of distances between evolutionarily coupled residues as natural choices for reaction coordinates which can be incorporated into Markov state model-based adaptive sampling schemes and potentially used to predict not only functional conformations but also pathways of conformational change, protein folding, and protein-protein association. We demonstrate the utility of evolutionary couplings in sampling and predicting activation pathways of the β_2-adrenergic receptor (β_2AR), folding of the FiP35 WW domain, and dimerization of the *E. coli* molybdopterin synthase subunits. We find that the time required for β_2AR activation and folding of the WW domain are greatly diminished using evolutionary couplings-guided adaptive sampling. Additionally, we could identify putative molybdopterin synthase association pathways and near-crystal structure complexes from protein-protein association simulations.

Publication:
Zahra Shamsi, Alexander S. Moffett & Diwakar Shukla. **Enhanced unbiased sampling of protein dynamics using evolutionary couplings.** Scientific Reports, 2017, 7, 12700.

4. Predicting optimal DEER label positions to study protein conformational heterogeneity.
Once molecular dynamics simulations are performed on a protein of biological interests, the datasets are analyzed resulting in publications and then archived resulting in no further insight obtained from them. Here we address a common problem faced by experimentalists, the choice

of spin-label positions in a protein for DEER spectroscopy experiments, by utilizing all-atom molecular dynamics simulations of protein dynamics. DEER spectroscopy is a powerful experimental technique for understanding the conformational heterogeneity of proteins. It involves attaching nitroxide spin labels to two residues in the protein to obtain a distance distribution between them. However, the choice of residue pairs to label in the protein requires careful thought, as experimentalists must pick label positions from a large set of all possible residue-pair combinations in the protein. Using our developed method, we rank the sets of labeled residue pairs in terms of their ability to capture the conformational dynamics of the protein. Our design methodology is based on the following two criteria: (1) An ideal set of DEER spin-label positions should capture the slowest conformational-change processes observed in the protein dynamics, and (2) any two sets of residue pairs should describe informational that has not already been captured in other choices. We utilize Markov state models of protein dynamics to identify slow dynamical processes and a genetic-algorithm-based approach to predict the optimal choices of residue pairs with limited computational time requirements. We find that the predicted choices of DEER residue pairs determined by our method provide maximum insight into the conformational heterogeneity of the protein while using the minimum number of labeled residues.

Publication:

Shriyaa Mittal & Diwakar Shukla. **Predicting optimal DEER label positions to study protein conformational heterogeneity.** J. Phys. Chem. B, 2017, 121, 9761-9770

Shriyaa Mittal & Diwakar Shukla. **Recruiting Machine Learning Methods for Molecular Simulations of Proteins.** Molecular Simulation, 2018 (*in review*)/

5. Complex dynamics of ABA binding and receptor activation revealed by kinetic network models.

Globally worsening drought conditions pose an increased threat to agriculture productivity and food security. In plants, the abscisic acid (ABA) signaling pathway regulates resistance to water stress, providing a key target for improved control of plant drought tolerance. Recent biochemical and structural studies have explained several aspects of ABA perception, yet the complete dynamic mechanism of ABA binding to the receptor and an understanding of subsequent protein conformational changes have been elusive. Here, we capture the atomic details of ABA binding to the PYL5 and PYL10 receptors and the following receptor activations using computational chemistry approaches. Our results reveal several common and distinct binding intermediate states in PYL5 and PYL10, and a conserved nonproductive binding state which can potentially decrease ABA on-binding rate. Furthermore, ABA binding is necessary but not sufficient condition for receptor activation, confirming the essential role of downstream protein phosphatase PP2C as ABA co-receptor. Our findings demonstrate that the major barrier to ABA binding appears to reflect the substantial dewetting of both ABA and the receptor taken place as ABA binds to the receptor. Moreover, our results indicate that ABA cannot bind to PYL5 after tyrosine nitration of the receptor, which inhibits ABA signaling in plants under conditions in which nitric oxide and reactive oxygen species are both present. Our findings will facilitate

informed genetic manipulation of ABA receptors and improved design of new ABA mimicking agonists.

Publication:
Saurabh Shukla*, Chuankai Zhao* & Diwakar Shukla. **Complex dynamics of ABA binding and receptor activation revealed by kinetic network models.** (in review 2017) *joint first co-authors.

6. Free energy landscape of the complete transport cycle in a key bacterial transporter.
Transport proteins belong to the class of membrane proteins that transfer molecules, often coupled with a proton, against the electrochemical gradient into the cytosol. The substrate binding and transport causes large conformational changes in the transporter resulting in multiple states of the protein; an outward-facing (OF), occluded and inward-facing (OF) state. $PepT_{So}$ belongs to the peptide transporter family (PTR) and transports di/tri peptide molecules. $PepT_{So}$ is a close homolog of human peptide transporter $PepT_1$ and $PepT_2$ that transports dietary peptide molecules from small intestine. Recently obtained $PepT_{So}$ crystal structure in inward-facing state provides the first glimpse of structural topology of these class of proteins. In our study, we performed microseconds long molecular dynamics simulation to identified the key intermediate states (OC and OF) and characterized the complete transport cycle. Using markov state model, we identified the free energy barrier for the transition to various metastable states. Our protein dynamics were validated with available randomly chosen DEER pairwise-distance distribution data a crucial experimental technique employed to capture the conformational changes of the membrane proteins. Using our in-house algorithm, we also predicted an alternative optimal position which has less probes to place nitroxide spin labels for DEER experiments and obtained the distance distributions. Using unbiased simulation, we captured the essential dynamics and the transport cycle for the first time for any membrane transporter protein. We also propose the optimal positions for probes that efficiently capture the possible protein conformational ensembles using optimal probes machine learning approach.

Publication:
Balaji Selvam*, Shriyaa Mittal* & Diwakar Shukla. **Free energy landscape of the complete transport cycle in a key bacterial transporter.** (in review 2017) *joint first co-authors.

7. Reinforcement learning based adaptive sampling: REAPing Rewards by exploring protein conformational landscapes.
One of the key limitations of Molecular Dynamics simulations is the computational intractability of sampling protein conformational landscapes with either large system size or long timescales. To overcome this bottleneck, we present the REinforcement learning based Adaptive samPling (REAP) algorithm that aims to sample a landscape faster than conventional simulation methods by identifying reaction coordinates that are relevant for sampling the system. To achieve this, the algorithm uses concepts from the field of reinforcement learning (a subset of machine learning), which rewards sampling along important degrees of freedom and disregards others that do not facilitate exploration or exploitation. We demonstrate the effectiveness of REAP by comparing

the sampling to long continuous MD simulations and least-counts adaptive sampling on two model landscapes (L-shaped and circular). We also demonstrate that the algorithm can be extended to more realistic systems such as alanine dipeptide and Src kinase. In all four systems, the REAP algorithm outperforms the conventional single long trajectory simulation approach as it can consistently discover more states as a function of simulation time.

Publication:
Zahra Shamsi, Kevin J. Cheng & Diwakar Shukla. **Reinforcement learning based adaptive sampling: REAPing Rewards by exploring protein conformational landscapes** (*in review*) (*https://arxiv.org/abs/1710.00495*)

8. On the transferability of time-lagged independent components between similar molecular dynamics systems

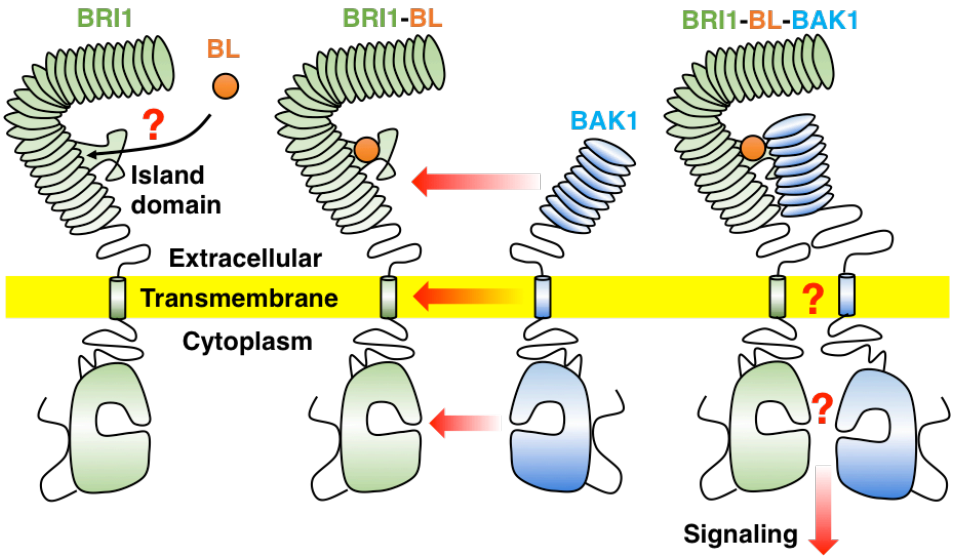
Dimensionality reduction techniques have found great success in a wide range of fields requiring analysis of high-dimensional datasets. Time-lagged independent components analysis (TICA), which finds independent components (TICs) with maximal autocorrelation, is often applied to atomistic biomolecular simulations, where the full molecular configuration can be projected onto only a few TICs describing the slowest modes of motion. Recently, Sultan and Pande have proposed the use of TICs as collective variables for enhanced sampling. However, it is unclear what the best strategy for estimating the TICs of a system is *a priori*. To evaluate the utility of TICs calculated on one system to describe the slow dynamics of similar systems, we develop a methodology for measuring the transferability of TICs and apply it to a wide range of systems. We find that transferred TICs can approximate the slowest dynamics of some systems surprisingly well, while failing to transfer between other sets of systems, highlighting the inherent difficulties of predicting TIC transferability. Additionally, we use two dimensional Brownian dynamics simulations on similar potential surfaces to gain insight into the relationship between TIC transferability and potential surface changes.

Publication:
Alexander S. Moffett & Diwakar Shukla. **On the transferability of time-lagged independent components between similar molecular dynamics systems** *(In review) (https://arxiv.org/abs/1710.00443)*

Total number of publication: 11 (*5 in review*)

Request for computation time for the period March 2018-Feb 2019.

1. **Predicting the full-length structure of the plant brassinosteroid signaling complex (Collaborator: Prof. S.C. Huber, Plant Biology, UIUC).**
   In our previous work on brassinosteroid signaling related to plant growth and development, we reported that the plant steroid receptors could be regulated by controlling the conformation of their kinase domains and by post-translational modifications such as S-glutathionylation. Here, we plan to perform additional investigations of the brassinosteroid receptor complex to computationally predict the structure of the full-length receptor complex as shown in the figure below. We have the structural information about the extracellular complex but the structure of the complex in the transmembrane and cytoplasm is not known.



2. **How bacterial flagellin activates plant immune system (Collaborator: Prof. S. R. Hinds, Crop Sciences, UIUC).**
   Plants and animals detect the presence of potential pathogens through the perception of conserved microbial patterns by cell surface receptors. Certain solanaceous plants, including tomato, potato and pepper, detect flgII-28, a region of bacterial flagellin that is distinct from that perceived by the well-characterized FLAGELLIN-SENSING 2 receptor. Hinds group has identified and characterized the receptor responsible for this recognition in tomato, called FLAGELLIN-SENSING 3. This receptor binds flgII-28 and enhances immune responses leading to a reduction in bacterial colonization of leaf tissues. However, it is not clear how this receptor is able to distinguish between different bacterial flagellin peptides. We intend to computationally investigate how this receptor could be modified to identify different types of bacteria that colonize other crop plants. This study could provide new insights into the plant immune system and transfer of the receptor to other crop plants offers the potential of enhancing resistance to bacterial pathogens that have evolved to evade FLS2-mediated immunity.

3. **Complex dynamics of plant hormone perception.**
   In this project, we plan to perform additional simulations to understand the thermodynamics and kinetics of plant hormone perception. We plan to perform simulations of all 8 plant hormone receptors for which crystal structures are available to quantify the thermodynamic stability (free energy) provided by hormone binding. Typically, plant hormone binding to the receptor leads to the formation of a receptor-coreceptor complex. However, despite the critical role played by plant hormones in regulating a variety of biological processes, the binding mechanism of plant hormones has been elucidated from a thermodynamic or kinetic viewpoint. This study will provide a unique perspective on "How plants sense hormones?".

   List of systems and required computation time for the three future projects outlined above.

#	System	Size	Total simulation time (μs)	Speed (ns/day)	Computation time (node hours)
1.1	Brassinosteroid receptor transmembrane domain association	50000	30	25	28,000
1.2	Brassinosteroid receptor intracellular domain association	100000	20	12	40,000
2.1	Association of Flagellar peptide Pst-T1to FLS3 receptor	90000	30	14	51,500
2.2	Association of Flagellar peptide Pst-T1to FLS3 receptor	90000	30	14	51,500
2.3	Association of Flagellar peptide Pst-T1to FLS3 receptor	90000	30	14	51,500
3.1	Jasmonate receptor	90000	30	14	51,500
3.2	Auxin receptor	90000	30	14	51,500
3.3	Strigolactone receptor	45000	40	28	34,300
3.4	Brassinosteroid receptor	50000	40	25	38,400
3.5	Cytokinin receptor	80000	30	16	45,000
3.6	Gibberellin receptor	80000	40	16	60,000
3.7	ABA receptor	60000	30	20	36,000
	Total				539,200

Total node hours are 539,200 node hours.