Title: developing an implicit solvent model based on the normal mode analysis.

Proteins are large and complex molecules that play many critical roles in living organisms. Understanding the structures and dynamics of proteins is one of the first step to understand their functional mechanisms, and computational studies provide valuable insights for understanding the dynamics at the molecular and atomic level. Even though proteins function in solvated environment, it is often preferred not to explicitly involve solvent molecules in computation since this can reduce computational cost and therefore speed up conformational sampling. Using implicit solvent models, such as generalized born surface area model, is one way to include the solvent effect without using solvent molecules. The implicit solvent models have been designed to approximate the average behavior of solvent molecules by applying the potential of mean force. Those models are useful, while may be not accurate enough to model the dynamics of hydrated proteins.

In this study, the PI plans to develop a new implicit solvent models that is designed to approximate the dynamics of hydrated protein, in order to use in computational biology studies. Our approach is motivated to use normal mode analysis techniques, which determines intrinsic dynamics of a protein from its Hessian matrix (the second derivatives of potentials). To this end, first, the potentials of proteins and water molecules will be approximated up to the second order, and the benchmark explicit solvent model will be obtained by removing water molecules by projecting their components onto proteins. Then, our implicit solvent model will be developed to approximate the benchmark explicit solvent models. Fig. 1 illustrates the concept of our approach. Our implicit solvent model will be developed through the following sequence of aims:

A1. Develop the benchmark explicit solvent model that provides the *accurate* dynamics of hydrated proteins. This model use information about the explicit solvent molecules, which will be eliminated in the following aims;
A2. Determine factors that significantly contribute to the hydrated protein's dynamics, by inspecting how the solvent molecules affects the benchmark explicit solvent model;
A3. Develop our implicit solvent model that approximates the benchmark model in A1 with using the significant factors in A2 but without explicitly using solvent molecules;
A4. Develop another implicit solvent model to use with the elastic network models, by further simplifying the developed model in A3.

This study will be performed with one or two graduate students in computer science, who are interested in continuing their studies in a Ph.D. program. There are several strengths for computer science people to work on computational biology studies, such as overcoming computational hurdles using different data structures and having different opinions/approaches when solving problems that are not popular in physics/chemistry/biology, in my experience. The PI will strongly encourage and seek out women to involve in this project.

Intellectual Merit

From this project, it is expected to improve the accuracy of describing the intrinsic dynamics of solvated proteins. The developed implicit solvent model is designed to approximate the potentials

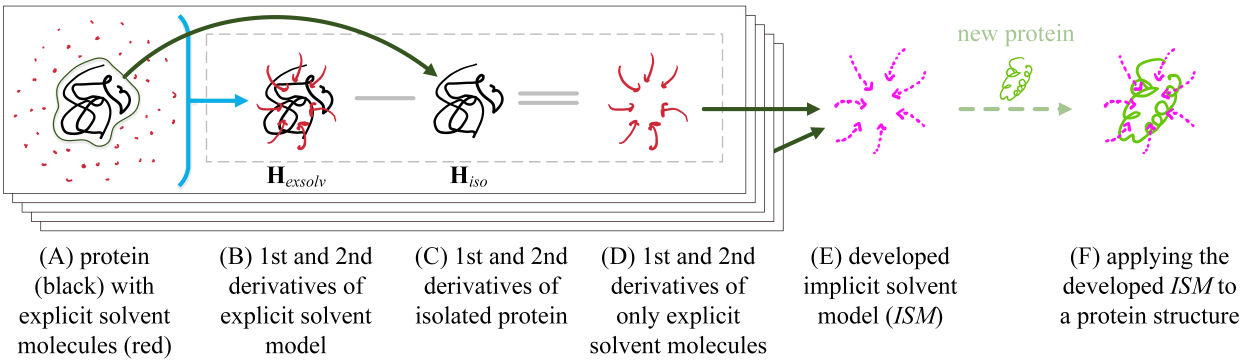

Figure 1: **Concept of the proposed approach.** From a given protein structure with explicit solvent molecules in (A), the benchmark explicit solvent model in (B) and the derivatives of isolated protein structure in (C) are determined. Then the derivatives containing to only solvent component in (D) are determined by subtracting (B) and (C). New implicit solvent model in (E) is developed to approximate data set (D)s. (F) illustrates applying our model to other proteins.

up to the second order, while traditional models are designed to approximate the potentials up to the first order (potential mean force). Additionally, since our approach approximates the empirical force fields and water models, such as CHARMM (chemistry at harvard macromolecular mechanics) and TIP3P, we expect that the computational results using our model will be congruent with the the molecular dynamics simulation results, and improve the accuracy of estimating free energy.

It is expected to enhance our understanding of how solvent mediates the interaction between proteins, which is essential for the computer-aided drug design. The goal of A2 is to identify factors that significantly contribute hydrated protein dynamics. From its results, we expected to find factors that have been overlooked when studying functional mechanisms of proteins. Especially, the importance of involving solvent molecules connecting proteins can be emphasized.

To reduce computational cost in this study, the PI plan to find the size/thickness of water layer (hydration shell) surrounding proteins, which directly influence the protein intrinsic dynamics in A1. From this, we could provide the recommended size of water box to perform the molecular dynamics simulation with given a protein.

The elastic network models are variations of the normal mode analysis, which is widely used to study the functional mechanism of biomolecules. However, solvents are not included in the elastic network model since there is no proper solvent model. The last model in A4 will increase our understanding of how proteins function in solvated environment.

Broader Impacts

The proposed research activities will be integrated with education activities, and research outcomes will be disseminated to scientific communities through several ways. The project will be performed with several graduate students in computer science who are interested in continuing their studies in Ph.D. programs; they will be knowledgeable about the strength and limitation of our models and framework, and will continue to disseminate the outcome to other communities. A web application to conveniently use our framework will be developed by encouraging undergraduate students in their capstone projects; through this, the web application will be provided to scientific communities, and undergraduate students will be motivated to be involved in scientific activity. The research results will be broadly disseminated through journal publications and conference presentations, and the developed computational framework will be publicly shared to the scientific community.