



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LDRD Final Report (08-ERD-037): Important Modes to Drive Protein MD Simulations to the Next Conformational Level

B. Sadigh

April 19, 2011

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Final Report for the ER-LDRD – FY2008

Important Modes to Drive Protein MD Simulations to the Next Conformational Level

Principle Investigator:

Babak Sadigh, (PLS)

Co-investigators:

Richard Law, Brian Bennion and Felice Lightstone (PLS)

Budget: \$180K, yr 1; \$90K, yr 2, \$100K, yr3.

This document's information falls within DUSA BIO.

Abstract

Every action in biology is performed by dynamic proteins that convert between multiple states in order to engage their functions. Often binding to various ligands is essential for the rates of desired transitions to be enhanced. The goal of computational biology is to study these transitions and discover the different states to fully understand the protein's normal and diseased function, design drugs to target/bias specific states, and understand all of the interactions in between. We have developed a new methodology that is capable of calculating the absolute free energy of proteins while taking into account all the interactions with the solvent molecules. The efficiency of the new scheme is an order of magnitude greater than any existing technique. This method is now implemented in the massively parallel popular MD program package NAMD. This now makes it possible to calculate the relative stability of different conformational states of biological macromolecules as well as their binding free energies to various ligands.

Background and Motivation

Biological operations are performed by proteins as they switch from inactive/closed states, through transition states, to transient active/open conformations. This general statement is true of proteins as diverse as soluble enzymes to transmembrane receptors and ion channels, e.g. potassium (K^+) channels (Fig 1). Many K^+ channel structures have been solved in the closed state (Doyle et al., 1998; Kuo et al., 2003; Zhou et al., 2001) because ion channel open states are very transient and are almost impossible to obtain structures experimentally. Potassium channels have been implicated in a host of neuromuscular, cardiac (Kv channels) and other diseases such as type-2 diabetes (the Kir6.2 channel). These channels are common in excitable cells, such as nerve cells and are the target of many anesthetics, pesticides and natural toxins. There are both ligand and voltage gated forms. The neuronal voltage gated forms are the channels responsible for the progression of an action potential along a nerve. Kv4.2-type channels are implicated in pain detection and are an important target for analgesic drugs. Drug design to target potassium channels as well as improved mechanistic understanding of channel operation and events such as toxin interactions require knowledge of the open conformation. Experimental groups have tried to solve structures for open states of K^+ channels of different types but the extreme techniques used have led to a series of what are now seen as incorrect structures, especially, for instance, for the Kv channel and its voltage sensing domain (Long et al., 2005).

Classical molecular dynamics (MD) based on empirical force fields is currently the best balance between an accurate representation of the chemistry involved and the ability to sample the complex energy landscapes encountered in molecular biology. However, there is a practical limit to how much information can be derived from MD, and this limit is well below the timescales and energetics required to search much more than local minima. The maximum simulation times accessible to current MD simulations are on the order of several nanoseconds. Yet, many biological phenomena of interest occur on time scales of microseconds and longer. This time-scale disparity can be traced to certain topographical features of the energy landscape of macromolecules in biological systems.

MD naturally operates on the time scale of typical high frequency atomic vibrations. The system spends most of its time in deep energy basins surrounded by energy barriers that are many times larger than its thermal energy. Only very occasionally, thermal fluctuations allow transitions over the barriers and movement to the neighboring basins. Given that the transition rate decreases exponentially with increasing

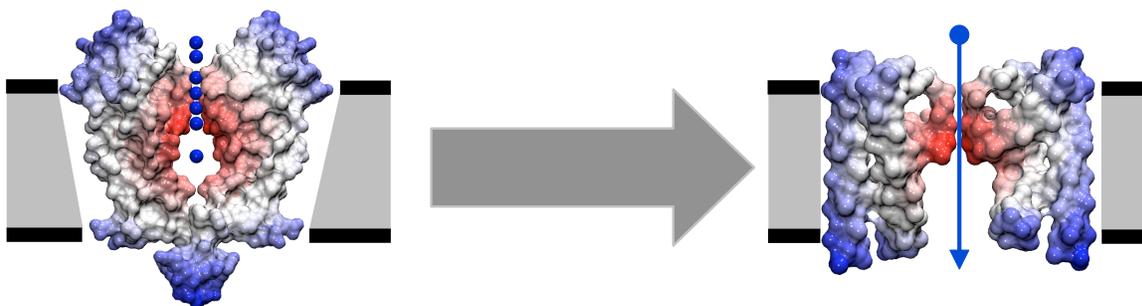


Fig. 1: The transition of a transmembrane potassium ion channel from closed to open state.

barrier height, such transitions represent rare events on the time scale of atomic scale molecular motion. Consequently, attempts to simulate them directly using MD are very inefficient.

During the past two decades, Normal Mode Analysis (NMA) has emerged as a powerful and universal tool for characterizing the dynamics of biological molecules (Brooks et al., 1983; Tirion, 1996). The NMA technique has often been conjectured to be able to predict the important degrees of freedom that may be responsible for steering the long-time dynamics of biological macromolecules. However, direct validation of the NMA forecasts is impossible. The time-scales covered by MD simulations are many orders of magnitude too short for this task.

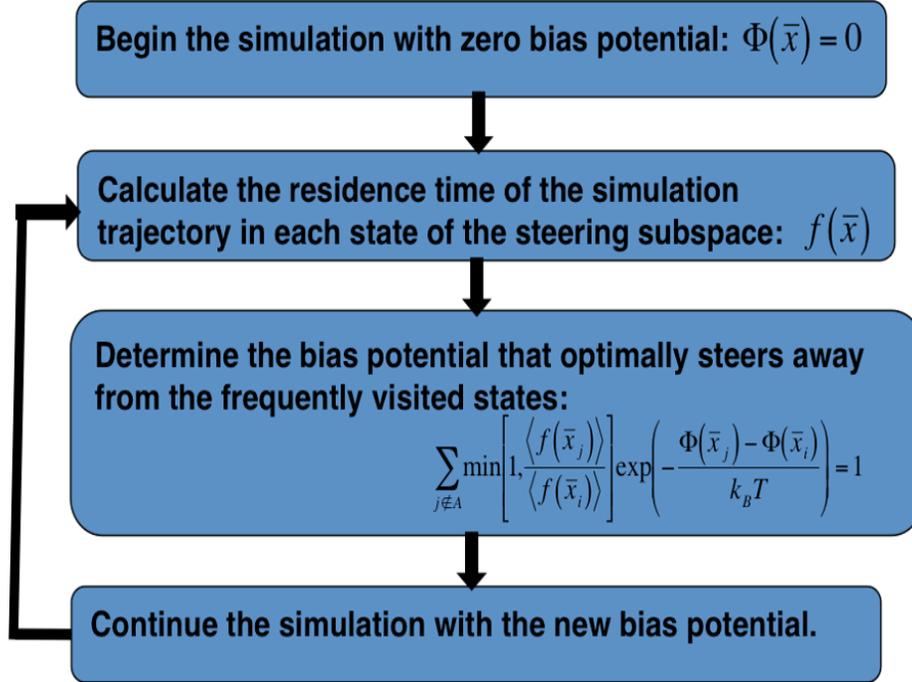
Achievements in the first year

In its first year, the project was funded at \$180K, and it was lead by Richard Law. The focus was the conformational transitions in the potassium channel. Our hypothesis was that the slow dynamics of the biological molecules are driven by only a few important reaction coordinates, characterized by low-frequency normal modes of the system as suggested by the NMA of the short-time dynamics. If true this hypothesis is able to drastically reduce the apparent complexity of the biological function. **The goal was to devise a Monte Carlo algorithm in the projected space of the slow normal modes in order to drive a closed K^+ channel to its open state. If successful, an important unsolved problem in biology would be resolved.**

Our method, described below is somewhat reminiscent of the methods of hyper-dynamics and the very popular method of meta-dynamics. Just like meta-dynamics, our method appeals to the notion of coarse-grained dynamics actively pursued in the physics and applied mathematics communities. The simulation is performed on a reduced (coarse) subspace of order parameters or steering coordinates chosen with the knowledge of the problem at hand, e.g. the subspace of a few low lying Normal Mode Coordinates (NMC). Let us denote the reduced NMC space by $\{\bar{X}\}$ and the full atomistic space by $\{\bar{x}\}$. Each microstate \bar{x} corresponds to one and only one steering state \bar{X} , while a single steering state vector \bar{X} generally corresponds to numerous microstates. In the most straightforward case, e.g. NMC, the microstates are mapped onto the steering coordinates through a linear projection operation. However, it is also possible to construct other mappings that are not described in terms of linear projections.

Our method accelerates the microscopic Monte Carlo simulation using a bias potential $\Phi[\bar{X}]$ defined on a reduced subspace of steering coordinates $\{\bar{X}\}$. During the course of one running simulation, an optimal bias potential is progressively deduced and used to steer the dynamics of the macromolecule towards the final conformation. The bias potential modifies the potential energy function and thus alters the transition rates and dramatically increases the probability of generating successful transitions. The key characteristic of our method is that despite the distortion to the stochastic dynamics produced by steering, the “smart” simulation finds the same transition pathways and with the same relative probabilities as the “honest” Monte Carlo simulation described above. Starting from scratch, the bias potential accumulated in the course of a simulation contains information on the rates at which the observed transitions would have taken place in an honest simulation as well as the variance between the “smart” and the

“honest” Monte Carlo simulations. Thus, statistical errors induced by steering can be quantified and controlled. Below we show a schematic representation of our Monte Carlo algorithm.



Hence in practice, one starts by selecting a set of continuous steering coordinates and defining a regular grid around the starting position in the sub-space of $\{\bar{X}\}$. The bias potential $\Phi[\bar{X}]$ will be defined on this grid. Now an unbiased simulation is performed for a few time steps. Over these initial moves, the residence time of simulated microstates $\bar{x}(t)$ in the steering subspace is collected. This is proportional to the probability $f(\bar{X})$ of finding the microscopic system in state \bar{X} . In the following we refer to $f(\bar{X})$ as the occupation function. An optimal bias potential can be obtained by solving in the subspace of the steering coordinates the following matrix equation

$$\sum_{j \in A} \min \left[1, \frac{\langle f(\bar{X}_j) \rangle}{\langle f(\bar{X}_i) \rangle} \right] \exp \left(- \frac{\Phi[\bar{X}_j] - \Phi[\bar{X}_i]}{k_B T} \right) = 1, \quad (1)$$

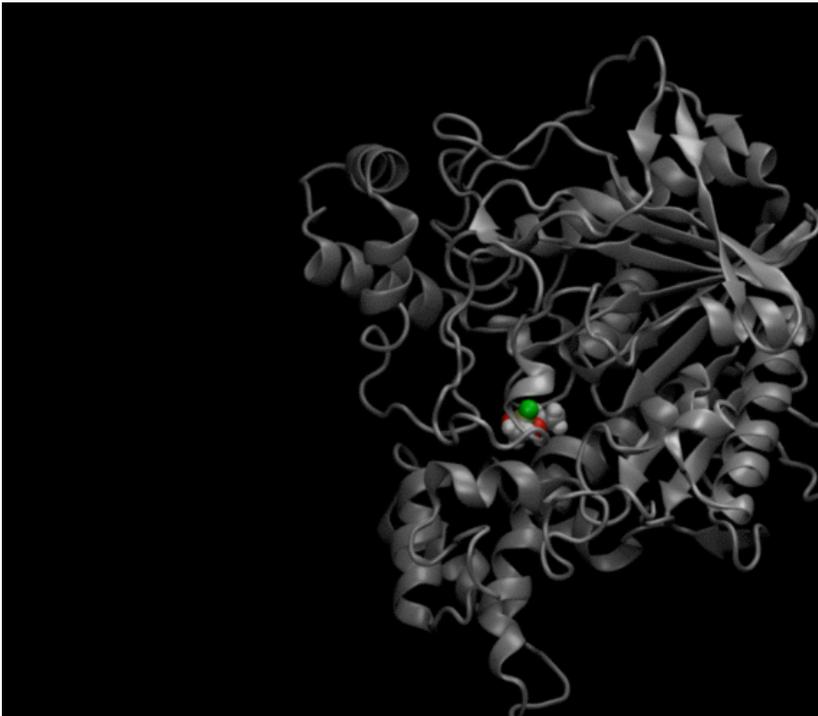
It can be shown that the bias potential obtained by solving equation (3) is optimal in the sense that it can be used to steer a subsequent MC simulation to still unexplored and poorly explored microstates so as to minimize the variance of the generated occupation function [de Koning et al.].

Given the optimal bias potential obtained from Eq. (3), the MC simulation can now continue where we last left off but now with the transition probabilities computed from the potential energy modified by the optimal bias potential. The difference in the value of the bias potential between any two points in the steering sub-space is proportional to the enhancement of the transition rate between these points as compared to the unbiased (honest) simulations. Again, the residence times of the simulated microstates $\bar{x}(t)$ in the steering subspace are accumulated. For simulations that have been

steered with a non-zero bias potential, the occupation function $f(\bar{X})$ is no longer proportional to the observed residence time, but must be weighted by $\exp\left[\frac{\Phi[X]}{k_B T}\right]$ and normalized.

Achievements in the second year

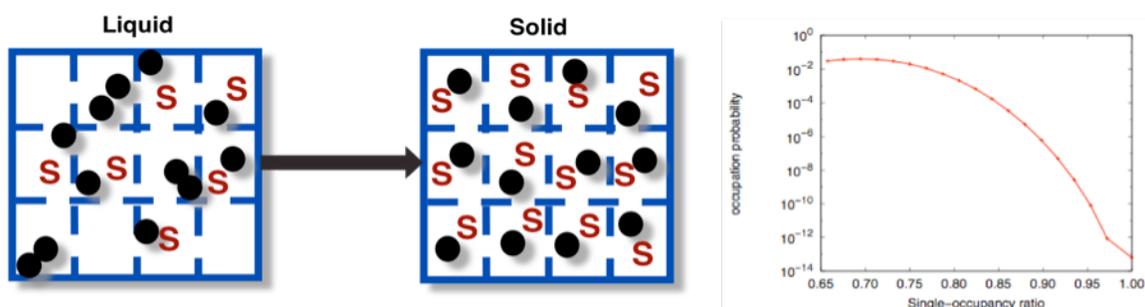
In the second year, Richard Law left the laboratory, and Babak Sadigh took over as PI of the project; its funding was cut into half of the original, mere \$90 K for the year. The method development could continue as had been planned originally, but the biological focus would now change from the study of the conformational changes of the potassium channel from closed to open, to investigating the binding of the Chemical Weapon Agent (CWA) Sarin to the active site of the protein BChE. This reorientation of the project was made in order to be able to collaborate with Brian Bennion, another staff computational biologist in Felice Lightstone's group. Understanding the binding energy as well as activation energy for the enzymatic reaction is important for obtaining predictive capability of the human toxicity to CWAs.



In the second year, we applied the method developed in the first year, described in the previous section, to two test cases: (i) direct calculation of the absolute entropy of a hard-sphere liquid by simulating the nucleation of an fcc crystal in the liquid, and (ii)

calculating the free energy of binding of the CWA sarin in the active site of BChE protein.

Application (i) above was chosen to showcase the power of the steered Monte Carlo technique, by performing a direct calculation of the free energy difference between a liquid and a solid phase. Such a calculation is known to be IMPOSSIBLE using the conventional free integration techniques. This is because of the discontinuities in the free energy derivatives across a first order phase transition. However, using our steered Monte Carlo this problem can be solved since the guided technique only required continuous bias potentials, while the underlying free energy landscape can be as arbitrarily complicated and fractal.



It is however, important to realize that a guided simulation is only as good as the chosen steering directions. In the case of the liquid/solid free energy difference calculation in particular, the choice of the steering subspace is crucial. For this application, we chose the steering coordinate to be the total number of single-occupancy sites (denoted by S in the Figure above) in the system. This coordinate is determined by decomposition of space into Voronoi cells about the perfect solid (fcc) sites. The total single-occupancy is the number of Voronoi cells containing one and only one particle. In the solid phase the total single occupancy is equal to the total number of particles in the system. The calculated probability of various single-occupancies is shown in the Figure above. It is found that the probability of generating an fcc configuration in a hard-sphere liquid with packing fraction of 0.44 containing 108 particles is about 6×10^{-14} , a very small probability that can be calculated with good statistical accuracy via our developed methodology. This is a very impressive achievement and can be extended to protein reactions in solvent. We are in the process of writing this achievement up for publication.

Achievements of the third year

As was found out from our endeavor in the second year that the main hurdle to obtaining rates of protein reactions with ligands or relative stability of various protein conformations from atomistic calculations is due to the large changes in the structure at e.g. the binding site that can occur along the transition pathway, which may require very long simulation times. These calculations become prohibitively expensive when explicit interaction with the surrounding solvent medium needs to be taken into account. In other words, if an efficient method could be devised to calculate the absolute solvation entropy

of an arbitrary protein conformation, all other problems relating to protein interactions with ligands or conformational transition rates can be easily solved.

Our experience from the second year of the project lead us to the understanding that novel sampling techniques can only speed up the calculations by about a factor of two or three. However, for these calculations to become practical at least an order of magnitude speed-up is necessary. This can be achieved through clever choices of steering coordinates, as was showcased in the second year by the liquid/solid free energy difference calculations.

Along these lines we made a significant innovation in the third year. We found a novel pathway that can allow for calculation of the free energy of freezing of protein degrees of freedom in solvent. This involves a non-trivial renormalization of the potential energy landscape for the frozen degrees of freedom. To describe this, let us denote by $\{\bar{x}\}$ the set of the protein degrees of freedom to be frozen into $\{\bar{x}_0\}$ and by $\{\bar{y}\}$ the set of solvent degrees of freedom. The potential energy landscape along the path to freezing renormalizes as follows

$$E_{\tau}[\{\bar{x}\},\{\bar{y}\}] = E[\{(1-\tau)\bar{x}_0 + \tau\bar{x}\},\{\bar{y}\}] + \frac{\kappa}{2}(\bar{x} - \bar{x}_0)^2 \quad (2)$$

where $E[\{\bar{x}\},\{\bar{y}\}]$ is the regular complex atomistic potential energy landscape, τ is the fraction of the protein temperature to the solvent temperature (room temperature). We have implemented this technique into the massively parallel popular MD code (written in C++) for the biology community, i.e. NAMD. We have shown that this method works without a problem for freezing the active sites in BChE where the ligand binds in solvent. Currently, we are concluding a study of the full absolute entropy of the BChE protein in water using this methodology. If successful, this will be a small revolution in the computing protein function in solvent. NAMD is a very large and complicated code. We are currently tackling a few remaining issues in our implementation. We are hopeful that within a short time, we will start preparing a high-impact paper on this innovation.

Relevance to LLNL and DOE

DOE OSCR has also identified the importance for continuing the advancement of US dominance in computations innovation by being able to simulate across multiple scales. The use of computer simulation to solve a biological problem clearly fits this goal. This study represents an important component in the future multiscale computing effort within the Computations, PLS Directorates and would increase LLNL's impact lab in the area of computational biology and new materials.

Exit Plan

We are very hopeful that we soon can present our new methodology as implemented in the NAMD program package to the scientific community as a breakthrough in atomistic calculations of protein reactions. This will provide LLNL with a capability to examine large conformational changes in proteins. In combination with LLNL large supercomputer platforms, LLNL will be uniquely positioned to examine larger biological systems than others. We will integrate this new methodology into our computational biology proposal to NIH, DHS, DOD, and BES.

References

- Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for macromolecular energy, minimisation, and dynamics calculations. *J. Comp. Chem.* 4:187-217.
- de Koning, M., W. Cai, B. Sadigh, T. Opperstrup, M.H. Kalos, and V.V. Bulatov. 2005. Adaptive importance sampling Monte Carlo simulation of rare transition events. *J Chem Phys* 122(7):074103.
- Doyle, D.A., J.M. Cabral, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Cahit, and R. MacKinnon. 1998. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* 280:69-77.
- Kuo, A., Gulbis JM, Antcliff JF, Rahman T, Lowe ED, Zimmer J, Cuthbertson J, Ashcroft FM, Ezaki T, and D. DA. 2003. Crystal structure of the potassium channel KirBac1.1 in the closed state. *Science* 300(5627):1922-1926.
- Long, S.B., E.B. Campbell, and R. MacKinnon. 2005. Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* 309(5736):897-903.
- Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781-1802.
- Tirion, M.M. 1996. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters* 77(9):1905-1908.
- Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J. Berendsen. 2005. GROMACS: fast, flexible, and free. *J Comput Chem* 26(16):1701-1718.
- Zhou, Y., J.H. Morais-Cabral, A. Kaufman, and R. MacKinnon. 2001. Chemistry of ion coordination revealed by a K⁺-channel-Fab complex at 2.0 Å resolution. *Nature* 414:43-48.