Toward an Integrated Computational Environment for Multiscale Computational Design of Nanoscale Ion Channel Semiconductors

Shreedhar Natarajan[1, 2], Sameer Varma[1, 2], Yuzhou Tang[1, 2], Scott Parker[1], Jay Mashl[1] and Eric Jakobsson[1, 2*]

[1]NCSA and [2]Department of Biophysics and Computational Biology, University of Illinois.

***Contac**t: 4021 Beckman Institute, 405 North Mathews, Urbana, IL 61801, 217 244 2896 (voice) 217 244 2909 (fax), jake@ncsa.uiuc.edu

ABSTRACT

This paper describes the design and operation of an integrated multiscale computational environment for design of nanoscale ion channel semiconductors, the Ion Channel Workbench. The present work builds on an earlier multiscale calculation from our lab [1] in which we showed that this approach could provide a close correspondence to experimental electrophysiological data on potassium channels. The current paper advances the previous work by incorporation of multiscale into a single integrated computation, in which the results of calculation at one stage automatically feed as input to calculations at other stages. It also employs more advanced electrostatics and Brownian Dynamics techniques than the previous calculations. In addition, integration of molecular dynamics and transport Monte Carlo into the Ion Channel Workbench is being actively pursued.

Keywords: electrostatics, brownian dynamics, ion channels, nanotubes, CAD

1 OVERVIEW

Ion channels are natural nanotransistors; development of biomimetic devices using these nanotransistors would greatly benefit from a computer-aided design (CAD) environment. Compared to other gene products, data on ion channels is rich in precise quantitative functional data. This motivates the development of computational models and methods that can predict the electrophysiology of channels given their 3D structure. A challenge to computational models is the range of significant time scales from femtosecond for rapid atomic motions to microsecond for experimental observable like flux. Techniques that can capture all time scales in explicit detail within the scope of a single simulation would require a prohibitive amount of computer time. Multi-scale approaches follow a divide and conquer strategy. The simulation is split into several layers in a hierarchical fashion with the time scale of phenomena defining the layering and statistical mechanics providing the connectivity between the layers.

A multi-scale strategy is well suited for device simulation purposes. One of the requirements of a CAD system is to enable rapid prototyping. This drives the need to decompose the processes within a nanotransistor CAD system so that they can be computed efficiently. However, as we scale down in both size and time, the uncertainty inherent in the models increases and it is therefore necessary to have robust connections between these processes that capture the uncertainties effectively within a probabilistic framework. Our multi-scale approach to ion channel simulation uses well-studied statistical mechanical theory to connect across scales. This makes it a good candidate for a nanotransistor CAD system. Arguably, any biomimetic semiconductor CAD system could be modeled using this paradigm - process engineering with statistical mechanical theories providing the underlying conceptual connectivity.

2 SOFTWARE ENGINEERING

We designed our *in-silico* framework to address minimally the following issues:

- Predict accurately and efficiently electrophysiology of the channel being studied.
- Extract knowledge from these simulations and formulate strategies for designing novel channels with specific IV characteristics; as is often a requirement in device design (Reverse Engineering)
- Allow efficient pluggability of new theories and methodologies
- Allow for effective dissemination through a Portal environment

2.1 Requirements and Design

We created software modules to address vertical decomposition along guidelines specified by the hierarchical approach. Horizontal decomposition within layers is necessitated by variety of programs and theories that

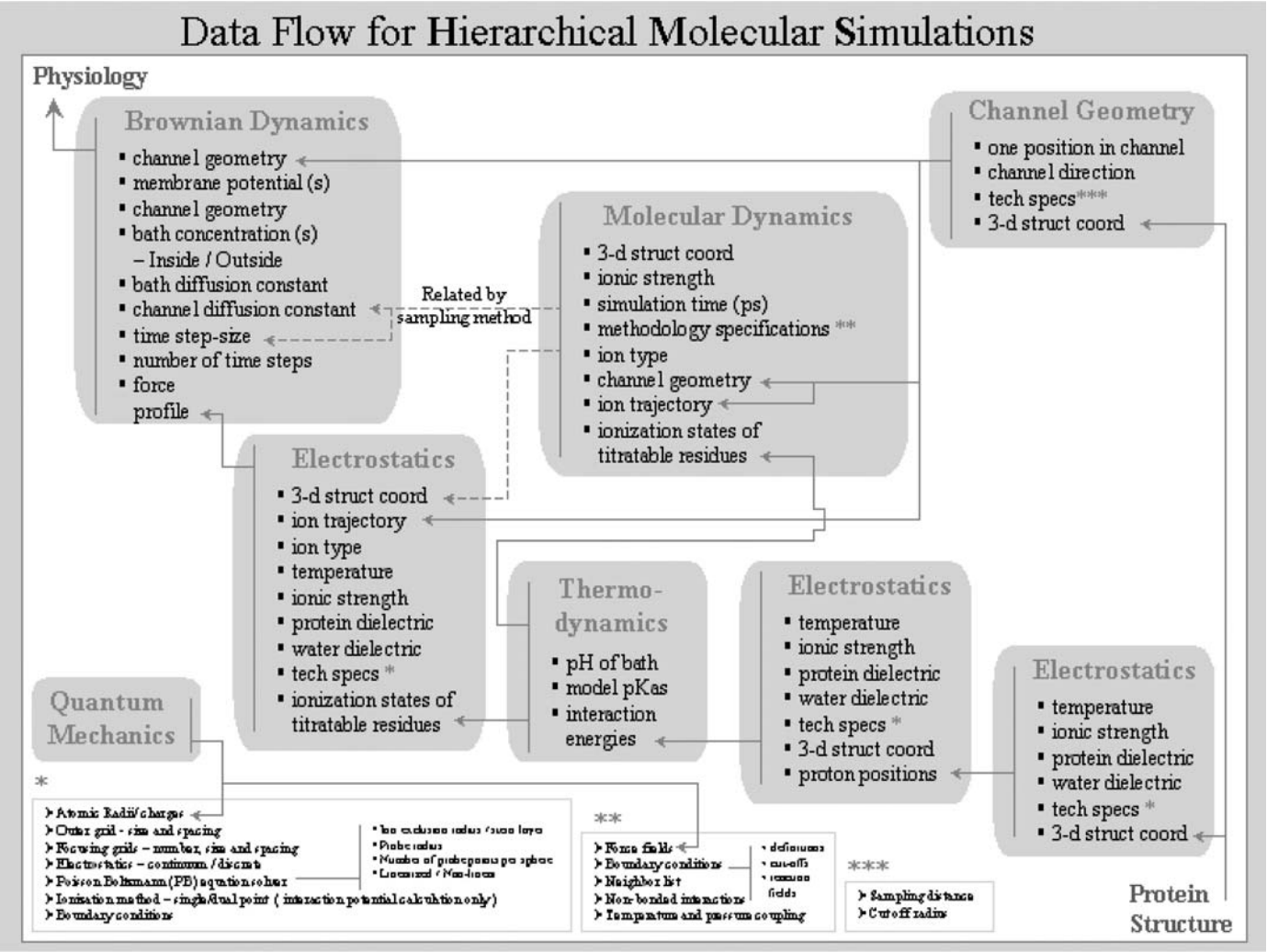

Fig 1: Data Flow Diagram for the Ion Channel Workbench.

contribute to it. E.g. Electrostatics. The workflow is shown in Fig. 1.

A module is characterized by its input and output. To ensure smooth connectivity across modules, it is necessary that the representation of information be uniform throughout the workflow. We have developed a simplistic model of module descriptors using XML [2] as the representation language. Using element overlap as the criterion enforces the connectivity between modules. These modules along with their XML metadata have been plugged and connected using a framework that we have developed locally. The framework includes a workflow interface, a workflow engine and a workflow metadata description also using XML.

Our approach to use a metadata based modular framework will have significant advantages in curating, archiving, searching, retrieving and analyzing simulation data in the future. A typical scenario in device design might be to "design a nanotube" with specific "ion flux" rates for a specific ion size and type. The first step would be to examine existing simulations and experiments for data that could suggest a naturally occurring ion channel that had the required output characteristic. It is therefore necessary to have an automated analysis technique that can search through precompiled datasets. However, simulation data can run into gigabytes in size and are typically stored in Mass Storage systems. Hence it is necessary to create metadata archives that contain the essence of the simulation results. An advantage to using semi-structured metadata is that we can formulate precise queries using structured query languages to the existing datasets. Such query languages and optimizers for queries are being developed by a large and organized community of researchers and are readily accessible. This preempts the need to write specific programs tailored to address each and every analysis. Another potential advantage is that the CAD system can incorporate multiple theories and external programs into various "plug-points" within the framework in a seamless fashion. There is an increasing drive towards interoperability [3] in the digital biology community with increasing adoption of XML based representations of their outputs. The use of semi-structured data in the creation of open source NanoDevice simulation

software can have a symbiotic and mutually beneficial relationship with the biological modeling community.

2.2 Implementation

The individual modules have been optimized to run in parallel if the problem allows for parallelization. The parallelized implementations of electrostatics and the Brownian dynamics components are described in Section 3 and Section 4. We discuss here the challenges involved in connecting the programs that reside on different computational clusters. Our implementation of the Ion Channel Portal addresses the issue of running a complete computational workflow on a computational grid seamlessly using a thin-web client. The web framework has been implemented mostly in Object-Oriented Perl. Each of the modules is wrapped with OO-Perl templates.

The modular metadata based framework discussed above lends itself to an implementation that utilizes the recently developed infrastructure of grid computing. Grid computing employs multiple, independent, distributed, heterogeneous, computational resources, coupled through a network and a common middleware standard, to perform complex computational tasks. In implementing the modular framework, the grid infrastructure allows a web server resource, which provides the web based workflow interface and the workflow engine, to be established and maintained separately from a set of remote high performance computational resources that are used to perform the computational tasks.

The workflow engine directs the execution of the ordered series of module calculations needed to characterize the electrophysiology of the channel being studied. These calculations may be performed on one or more remote computational resources, with the workflow manager assigning a module task to each resource and then receiving and managing the resulting output data. Tasks are assigned and data transferred using an asymmetric connection between resources. The connection from the web-server/workflow manager to the remote computational resource is made through the Globus grid middleware [4]. The connection from the remote compute resource back to the workflow manager is made through the web server and a custom remote http based client.

There are several advantages to an implementation based on the grid infrastructure. The coupling of distributed heterogeneous resources allows tasks to be performed on the resources best suited to their execution. Additionally, as more resources are made available on the grid, these resources can be utilized, extending the computation power of the portal. Finally, new modules can be integrated seamlessly into the framework, even those maintained on remote resources controlled by disparate members of the community.

3 ELECTROSTATICS

The long-range nature of Coulumbic forces makes electrostatics a vital player in both intra- and intermolecular interactions. Mathematically, the laws governing the distribution of potentials and electric fields are all encapsulated in the Poisson-Boltzmann equation (PBE). The energetics of conformational states, proton dissociation and diffusion; all have a strong electrostatic component. Within the context of our multiscale approach, we have decomposed electrostatic calculations into three modules.

Firstly, electrostatics is coupled with empirical equations (derived from quantum mechanics and probability density functions [5] to estimate and optimize coordinates of hydrogen atoms that are not resolved or ambiguous in protein structures [6].

Secondly, several functional groups in proteins (and some in synthetic nanotubes) have a tendency for proton dissociation. Proton dissociation alters the net charge of a functional group and thus plays a significant role in defining ion flux rates/ratios. Proton dissociation occurs with a probability (pKa) that is strongly influenced by environmental conditions. This probability is calculated from thermodynamic equations that require free energy functions calculable from PBE theory. We have recently formulated a systematic strategy that eliminates most parameter-dependent uncertainty in these calculations [7].

Finally, electrostatics has also been used to calculate the deterministic part of the force experienced by an ion in a Brownian dynamics simulation. We have also recently made several refinements to this methodology [8].

All the three modules in our hierarchy have been parallelized to efficiently run on multiple processors. Instead of parallelizing the PBE solver itself, each of these modules has been parallelized by decomposing the entire problem domain into a set of smaller domains. For instance, if one needs to determine electrostatic forces for a given set of trajectory points inside a nanotube (i.e. force profile), one can distribute these points across different processors. Each processor then calculates the forces for its own set of points. Such domain decomposition, as can be seen, is ingrained in the nature of the problem. We have used the same strategy to parallelize all the three modules [7, 8]. This context based parallelization strategy has two main advantages. Firstly, it does not depend upon the availability of an ingrained parallelization in the PBE solver. This allows the freedom to plug in any kind of a sequential PBE solver. And secondly, this strategy demands for a minimum possible inter-processor communication, which makes this parallelization ideal for grid-computing environments.

4 BROWNIAN DYNAMICS

Brownian motion is the random movement of solute molecules in dilute solutions that result from repeated collisions with solvent molecules. The basic principle behind Brownian dynamics (BD) is the same as molecular dynamics, but it introduces some new approximations that allow us to perform simulations at a microsecond time scale. We introduce implicit solvent description and ignore internal molecular motions that allow us to use a larger time step.

Our earlier implementation of BD simulation implemented this concept [9], along with the added assumption that diffusion of ion is along the channel axis; based on the observation that the pore region of ion channel is very narrow so that ions could only move in single-file style. These simplifications made it possible for the model system to be simulated for as long as 1ms within only 1 hr of CPU time; thus allowing us to retrieve statistically meaning results, such as current-voltage curves that are observable in experiments. We have made several improvements to the original algorithm [8]. It is now possible to do the simulation for multiple species simultaneously. Ion-ion pairwise interactions are pre-calculated by electrostatic calculations and converted into a look-up table so that this interaction could be more accurately represented during the simulation. The calculations have also been parallelized by temporal decomposition of the total simulation interval among multiple processors.

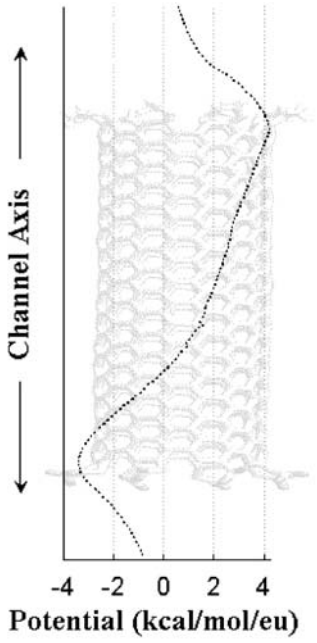

Fig 2: The Potential as seen by a K^+ ion as it moves along the permeation pathway of the Nanotube. The nanotube is shown as a watermark.

5 RESULTS AND DISCUSSION

Using the potential profile of nanotube showed in Fig 2 as input, the 1-D BD simulation generated the current-voltage (I-V) curve of K^+ that is shown in Fig 3. From this figure, it can be seen that there is a preference of influx than efflux of K^+ current because of the asymmetry of the potential profile. For this calculation, a total of 11 simulations were done with simulation time of 180 microseconds for each one. It cost 4 hour CPU time on NCSA IBM p690 machine.

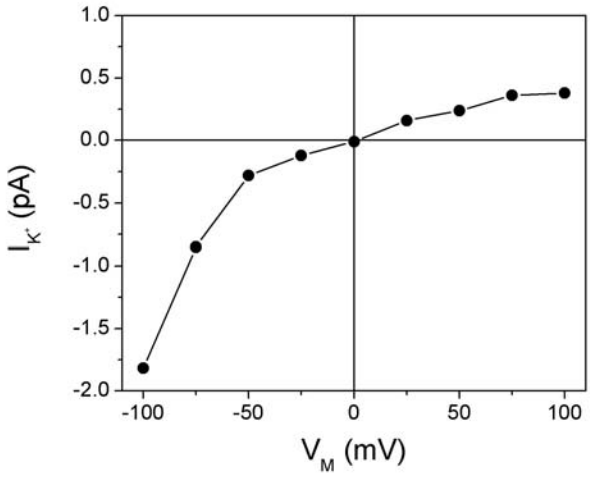

Fig 3: The I-V curve of a nanotube in a pseudo bilayer. $[K^+]$ on both sides is 150mM

References

[1] Mashl, R.J., Y. Tang, J. Schnitzer, and E. Jakobsson. Biophysical Journal. **81**. 2473-2483. 2001.
[2] URL: http://www.w3.org/XML/.
[3] URL: http://www.sbml.org.
[4] URL: http://www.globus.org.
[5] Hooft, R.W., C. Sander, and G. Vriend. Proteins. **26**(4). 363-76. 1996.
[6] Malmberg, N.J., S. Varma, E. Jakobsson, and J.J. Falke. 2004.
[7] Varma, S. and E. Jakobsson. Biophysical Journal. 2004.
[8] Tang, Y., S. Varma, and E. Jakobsson. 2003.
[9] Bek, S. and E. Jakobsson. Biophysical Journal. **66**. 1028-1038. 1994.