

New and Notable

Engineering Teams Up with Computer-Simulation and Visualization Tools to Probe Biomolecular Mechanisms

Tamar Schlick

Department of Chemistry and Courant Institute of Mathematical Sciences, New York University and the Howard Hughes Medical Institute, New York, New York 10012

With the growing number of global enterprises in biology that require computer analysis of data on a large scale—what Sydney Brenner half-humorously calls “e-biology” (Brenner, 2002)—it has been remarked on occasion that some thought and creativity may be declining. Yet I’d like to argue that although some aspects of this sentiment are true in cases, modern technology also presents unprecedented opportunities for exploring biological systems. One such innovative example is described in this issue of the *Biophysical Journal* by Grayson et al. (Grayson et al., 2003). These researchers present a creative approach called “Interactive Molecular Dynamics” (IMD) for probing mechanisms of biological reactions, a technique only made possible by the marriage of state-of-the-art scientific visualization, computer simulation (theory and implementation), and engineering tools. To demonstrate its potential, IMD is applied to develop and examine substrate-selective mechanisms for conducting sugars through the glycerol-conducting channel protein (GlpF) in *Escherichia coli* and of binding/unbinding glycerol in the glycerol kinase (GK) protein, also from *E. coli* (see Figs. 1–3). Combined

with noninteractive (i.e., standard) dynamics simulations, results from IMD can be further explored to complement experimental structural studies. Most exciting, IMD is generally applicable to other molecular systems, as long as the user has all the requisite hardware and software.

Dynamics simulations of biological systems are routinely used to sample the thermally accessible conformational states of macromolecules through temporal and spatial trajectories that follow classical physics (Karplus and McCammon, 2002). In theory, simulating Newtonian mechanics can capture the desired kinetics events. In practice, modeling accuracy and computing complexity limit the time range that can be followed. Certainly, code parallelization on multiple-processor machines or distributed computing shaves off computing clock time, as demonstrated by the longest continuous simulation to date—1 μ s, for a villin headpiece, achieved in four months of dedicated CPU time on 256 processors of a Cray T3D/E (Duan and Kollman, 1998), or aggregate dynamics using the Folding@home community-wide megacluster for several hundred microseconds (Snow et al., 2002)—but typical simulations lengths are on the 10 ns range (already considered “long” (Daggett, 2000)). Thus, it is no surprise that, as one of the largest group of supercomputer consumers, macromolecular modelers are always scrambling for more computer time, faster codes, possible model approximations, and new algorithms to increase the timestep size, enhance the sampling, and improve the accuracy of the models and results.

The molecular dynamics integration timestep cannot be increased much beyond the typical 1 fs value because, with decreased intervals of molecular observations (force update frequency), the resolution of the fast processes is sacrificed and hence the overall is altered. This results from the nonnegligible effect of the fast processes on the global molecular motion due to intricate

coupling of vibrational modes: the fast small-amplitude motions can trigger a concerted series of events that produce large-scale global rearrangements. This subtle coupling, which is lacking or much weaker in other physical systems like planetary motion, limits the traditional mathematical machinery available for numerical integration of ordinary differential equations (e.g., force-splitting or multiple-timestep methods have limited success, as summarized in Schlick (2001)), and compels algorithm developers to seek inventive, tailored approaches.

As our appreciation of the need to sample biomolecular configuration space more globally heightened, numerous approaches to increase the sampling have been developed (e.g., see summary in Schlick (2002), Chapter 13). For example, multiple trajectories (rather than one long run) can be effective for improving statistics, combinations of Monte Carlo and molecular dynamics techniques can survey conformational space efficiently (e.g., Hansmann and Okamoto (1999), or alternative coordinate frameworks can be designed to extract collective motions for biomolecules that are essential to function (e.g., Kitao and Go (1999)).

Yet the fun begins when we allow altering the model—through biasing forces or guiding restraints—to capture events that would be otherwise unaccessible. For example, diffusion-limited processes associated with high-energy or entropy barrier can be “accelerated” through biasing forces in Brownian dynamics simulations, with rate calculations adjusted by associating lower weights with movements along high biases and vice versa (Zou et al., 2000). Biomolecular systems can be “steered” (Isralewitz et al., 2001) or “guided” (Wu and Wang, 1999)—subjected to time-dependent external forces along certain degrees of freedom or along a local free-energy gradient—to probe molecular details of certain experiments, or to study folding/unfolding events (Wu et al., 2002).

Submitted March 5, 2003, and accepted for publication March 7, 2003.

Address reprint requests to Tamar Schlick, Dept. of Chemistry and Courant Institute of Mathematical Sciences, New York University and the Howard Hughes Medical Institute, 251 Mercer St., New York, NY 10012. E-mail: schlick@nyu.edu.

© 2003 by the Biophysical Society

0006-3495/03/07/1/04 \$2.00

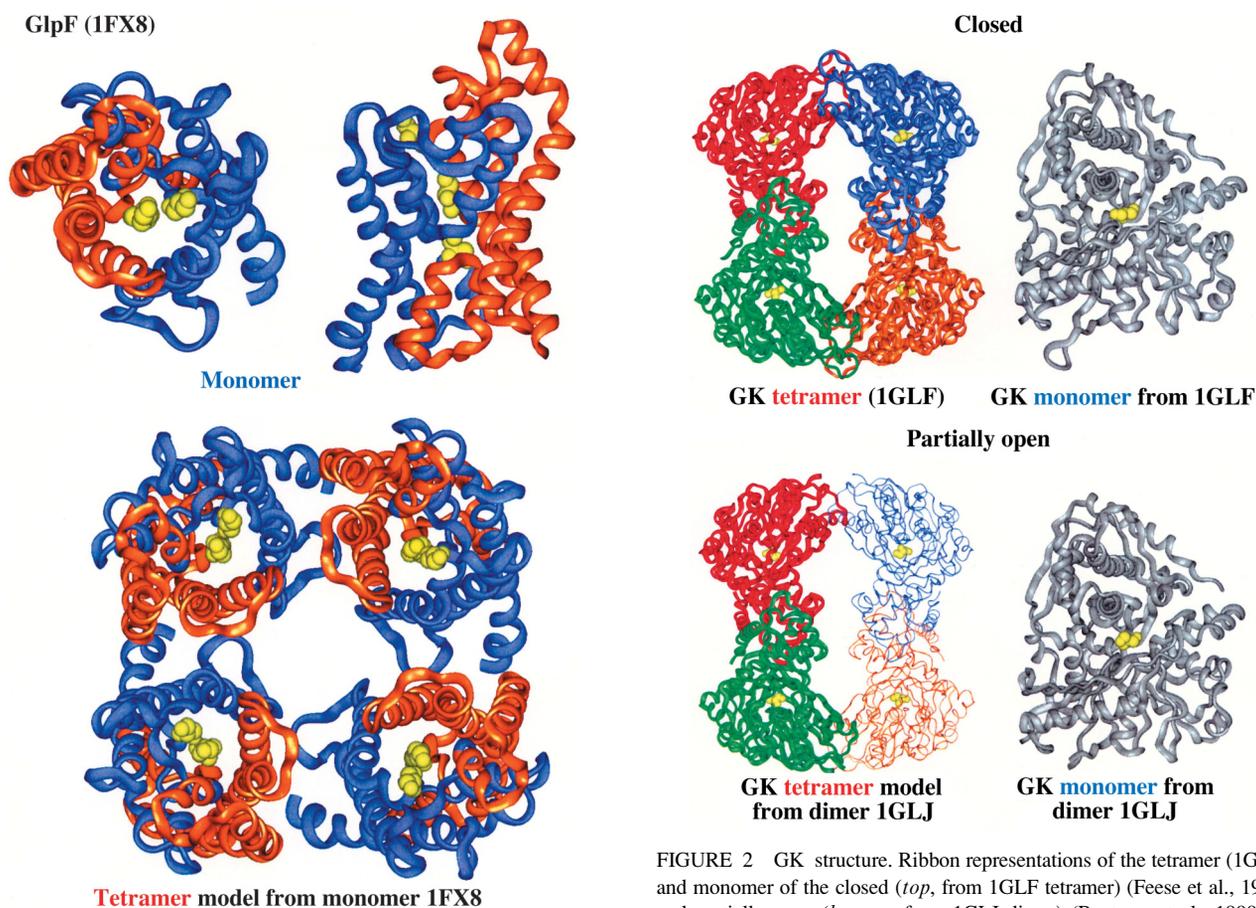


FIGURE 1 GlpF structure. Ribbon representations of the GlpF monomer (*top*, PDB entry 1FX8) and tetramer (*bottom*) forms, with the latter constructed from the monomer, with glycerol shown in a space-filling yellow model (Fu et al., 2000).

FIGURE 2 GK structure. Ribbon representations of the tetramer (1GLF) and monomer of the closed (*top*, from 1GLF tetramer) (Feese et al., 1998) and partially open (*bottom*, from 1GLJ dimer) (Bystrom et al., 1999) of GK, with glycerol shown in a space-filling yellow model. The tetramer of the partially open form is built from the dimer 1GLJ.

If generating pathways between known structures is the goal (e.g., *closed* and *open* forms of a polymerase enzyme or unfolded and folded states of a protein), the dynamic pathway may be “targeted” by use of restraints, which monitor the distance to the reference structure (Paci and Karplus, 2000). Despite the fictitious trajectory that results from such “targeted MD”, insights into disallowable configurational states can be obtained, as well as conclusions regarding common pathways. Umbrella sampling, a traditional approach for enhancing conformational sampling, can also be very effective when combined with such biased trajectories (e.g., Bernèche and Roux (2001)). Two noteworthy sophisticated approaches for sampling long-time processes and obtaining reaction pro-

files (transition state regions and associated free energies) are the stochastic path approach (Siva and Elber, 2003) and transition pathway sampling (Bolhuis et al., 2002); exciting applications to biomolecules have been, and are now, in progress with these methods.

All the approaches above are “non-interactive”. That is, parameters and instructions are prescribed at the onset of the simulation, and the researcher awaits trajectory completion before embarking on the challenging phase of data analysis. Now imagine that we can instead sit in front of a device that would allow us to apply a restoring force of any given direction and magnitude to any atoms/location in the biomolecular system and continuously adjust the force as the molecular response is observed. This idea, explored in the past in

the context of molecular mechanics, is now developed for molecular dynamics by Grayson et al. (2003). Namely, a computer-driven haptic device controlled by a researcher is connected to the sophisticated dynamic visualization (VMD) and simulation (NAMD) package developed in the Schulten group; the interface allows simulations to be run and displayed at the same time. The researcher selects atoms (e.g., sugar molecule) using a three-dimensional pointer and applies a force (of several-hundred picoNewtons in size) with the haptic device by pressing a button, which links the applied spring force to the selected object. The user of the device, in turn, feels the pulling force on the atoms by a resistance response. Better accuracy requires lower force magnitudes; otherwise, the internal

forces might be small compared to the applied force and an energy transfer can alter the global behavior along the reaction coordinate. The applied force can steer the sugar substrate, as done in the work, down the channel while indicating how much force—which evolves with the motion—is needed to reach the target. Thus, a strong resistance signals high-energy barriers and a smaller value indicates that favorable molecular interactions may have taken over (and thus a smaller external spring is needed).

In this way, data are collected over an IMD session of about one hour in length covering ~100–150 ps of simulated time; of course, the system under study must be simplified as needed (e.g., protein monomer, only a few water molecules) to make the running simulations fast enough to be displayed and adjusted interactively. (A local cluster of 32 Athlon CPUs running Linux and the Pittsburgh Supercomputing Center Terascale computer system were used in Grayson et al., (2003).)

The IMD runs are performed with variations (e.g., initial substrate orientation in channel, protein open or closed forms, ribitol or arabitol sugar substrates) to suggest structural and dynamic aspects of the biological reaction, such as molecular interactions between substrate and protein and/or solvent water molecules. But these runs are just the beginning. Hypotheses generated by IMD are further explored through several noninteractive dynamics simulations of 1–3 ns in length, which require a few days of CPU time on these advanced software/hardware platforms.

The two systems studied in the work with IMD are motivated by crystal structures along with interesting mechanistic insights of two proteins associated with the metabolism of glycerol in *E. coli*. The membrane channel protein GlpF conducts glycerol with high selectivity while excluding ions, water, and other charged solutes. Its structure suggested a narrow selectivity filter where key protein/substrate/water interactions are stabilized (Fu et al., 2000)

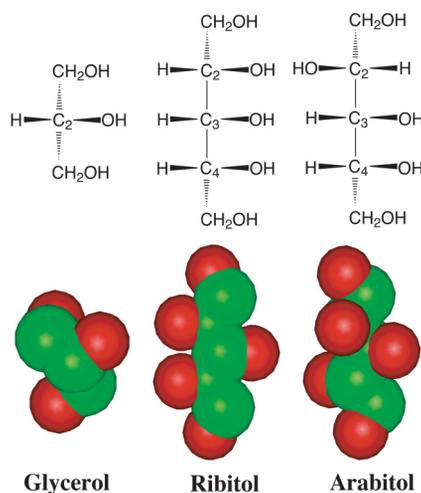


FIGURE 3 Sugar structures. Chemical formulae and space-filling models for the sugars glycerol, ribitol, and arabitol.

(see Fig. 1). Grayson et al., (2003) apply IMD to study the conduction of larger sugars than glycerol, ribitol, and arabitol (see Fig. 3), where unfavorable steric interactions or hydrogen-bonding incompatibilities can be amplified; the former pentaol has a higher rate of conduction than the latter, but the mechanism explaining this is unclear. By pulling each sugar into the GlpF monomer channel using two different orientations (carbon 1 or carbon 5 first), critical interactions as well as a conformational rearrangement of the substrate are revealed near and far from the selectivity filter. The differences in hydrogen-bonding sites and alignments of the sugars with the electrostatic field of the channel walls explain the observed disparity in conduction between ribitol and arabitol.

The second application of IMD involves the enzyme GK, which catalyzes the phosphorylation of glycerol after it is conducted into the cytoplasm (Feese et al., 1998; Bystrom et al., 1999). Structural studies have suggested that GK undergoes conformational transitions between *open* and *closed* forms during its catalysis (see Fig. 2). IMD is applied to open and closed forms of GK to simulate the unbinding of glycerol to explore the mechanism and possible effect of the two different protein forms. The two protein forms (related by a rigid-body subdomain motion) are found to

be key in defining the interactions inside the binding pocket, that is, the specificity and strength of glycerol/GK interactions. A tighter binding in the closed form makes glycerol extraction more difficult, whereas the sugar is more flexible in the open form, allowing nearby water molecules to reorient within the binding pocket.

A recurring theme in both proposed mechanisms is the crucial dynamic interplay of the following three factors involving the substrate, solvent, and protein that explains the induced fit and selectivity of the sugar conduction/unbinding events: conformational rearrangements of the linear sugar *substrate* (isomerization fluctuations that change the dipole orientation and hence the substrate's interactions with water and protein residues), *water* movement that reorient solute/solvent interactions, and conformational rearrangements of the *protein*.

Though all analyses are of qualitative nature and rely on the assumption that the external field does not disturb characteristics of the free energy profile of the system, combined with standard dynamics simulations, IMD is clearly a promising tool for generating rapidly mechanistic hypotheses for many biomolecular reactions. IMD may be particularly effective when researchers have amassed a body of structural and dynamic data to stimulate specific hypotheses. When tested with IMD, the

findings can be used to suggest new experiments (e.g., protein mutations, variations in the electrostatic environment) as well as additional theoretical studies to confirm the mechanisms. Undoubtedly, researchers fortunate enough to enjoy the experience of an IMD simulation in progress on their favorite biomolecular system must be excited to replace the standard painstaking process of trajectory generation and analysis by a one-hour idea-provoking session.

I thank Linjing Yang for preparation of the figures and Ravi Radhakrishnan for reading a draft of this article.

REFERENCES

- Bernèche, S., and B. Roux. 2001. Energetics of ion conduction through the K⁺ channel. *Nature*. 414:73–77.
- Bolhuis, P. G., D. Chandler, C. Dellago, and P. L. Geissler. 2002. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* 53:291–318.
- Brenner, S. E. 2002. The worm's turn. *Curr. Biol.* 12:R713.
- Bystrom, C. E., D. W. Pettigrew, B. P. Branchaud, P. O'Brien, and S. J. Remington. 1999. Crystal structures of *Escherichia coli* glycerol kinase variant S58 → W in complex with nonhydrolyzable ATP analogues reveal a putative active conformation of the enzyme as a result of domain motion. *Biochemistry*. 38:3508–3518.
- Daggett, V. 2000. Long timescale simulations. *Curr. Opin. Struct. Biol.* 10:160–164.
- Duan, Y., and P. A. Kollman. 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*. 282:740–744.
- Feese, M. D., H. R. Faber, C. E. Bystrom, D. W. Pettigrew, and S. J. Remington. 1998. Glycerol kinase from *Escherichia coli* and an Ala65 → Thr mutant: The crystal structures reveal conformational changes with implications for allosteric regulation. *Structure*. 6:1407–1418.
- Fu, D., A. Libson, L. J. W. Miercke, C. Weitzman, P. Nollert, J. Krucinski, and R. M. Stroud. 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*. 290:481–486.
- Grayson, P., E. Tajkhorshid, and K. Schulten. 2003. Mechanisms of selectivity in channels and enzymes studied with interactive molecular dynamics. *Biophys. J.* 85:36–48.
- Hansmann, U. H. E., and Y. Okamoto. 1999. New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.* 9:177–183.
- Isralewitz, B., J. Baudry, J. Gullingsrud, D. Kosztin, and K. Schulten. 2001. Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model.* 19:13–25.
- Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9:646–652.
- Kitao, A., and N. Go. 1999. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* 9:164–169.
- Paci, E., and M. Karplus. 2000. Unfolding proteins by external forces and temperature: The importance of topology and energetics. *Proc. Natl. Acad. Sci. USA.* 97:6521–6526.
- Schlick, T. 2001. Time-trimming tricks for dynamic simulations: Splitting force updates to reduce computational work. *Structure*. 9:R45–R53.
- Schlick, T. 2002. *Molecular Modeling: An Interdisciplinary Guide*. Springer-Verlag, New York.
- Siva, K., and R. Elber. 2003. Ion permeation through the gramicidin channel: Atomically detailed modeling by the Stochastic Difference Equation. *Proteins*. 50:63–80.
- Snow, C. D., H. Nguyen, V. S. Pande, and M. Gruebele. 2002. Absolute comparison of simulated and experimental protein folding dynamics. *Nature*. 420:102–106.
- Wu, X., and S. Wang. 1999. Enhancing systematic motion in molecular dynamics simulation. *J. Chem. Phys.* 110:9401–9410.
- Wu, X., S. Wang, and B. R. Brooks. 2002. Direct observation of the folding and unfolding of β -hairpin in explicit water through computer simulation. *J. Am. Chem. Soc.* 124:5282–5283.
- Zou, G., R. D. Skeel, and S. Subramanian. 2000. Biased Brownian dynamics for rate constant calculation. *Biophys. J.* 79:638–645.