Long-timescale simulation methods

Keywords: Molecular dynamics, classical action, reaction paths, path sampling, milestones

Teaser: The well-known and hard limit on timescales in biological simulations (restricted to nanoseconds) is slowly falling apart, as recent and continuous developments in theory suggest.

Summary:
The outstanding challenges in computer simulations of biological macromolecules are related to their complexity. Part of the complexity revealed in biological systems is of physical size. Enumerating atoms from a few in small signal molecules to millions of particles in biological complexes is an obvious example of biological hierarchy. Another aspect is the extremely broad range of time scales of life-science processes (many orders of magnitude), which adds another dimension of complexity. These extended timescales are observed even for a single biomolecular process and are the focus of the present review.

While the basic time step of atomically detailed simulations is about a femtosecond, it is not difficult to find molecular processes in biology that span more than ten orders of magnitude of relevant times, making the straightforward simulations of these events very difficult. Consider for example the R to T transition in hemoglobin. The complete conformational change occurs in tens of microseconds. However, there is more than one time scale in the system. Considerable activity occurs on a range of time scales before the final event (heme relaxation – picoseconds, tertiary relaxation – nanoseconds, ligand escape from the protein matrix and rebinding – hundreds of nanoseconds, etc.).

In the last few years the time scale problem was examined from a number of different directions. Rather than focusing on a selected subset of techniques, this review is meant to give a comprehensive picture of the methodologies that are out there so the interested reader can dig in more deeply (following the references). The ideas discussed include the coarser description of protein molecules (the example used is of elastic networks), smoothing protocols to accelerate dynamics, identification of reaction paths, optimization of actions, simulation of rare (but fast) events, simulations of diffusive processes by milestoning, and the use of massively distributed computing.

Introduction
Atomically detailed simulations of biological macromolecules added significantly to our understanding of biological processes. We better understand the driving forces for conformational transitions, and for energetic and thermodynamic binding. We use atomically detailed simulations to interpolate between independent experimental observations and create a single unifying picture of biological processes. Nevertheless, significant challenges in atomically detailed simulations of kinetics of biological molecules remain. It all boils down to the fundamental integration time step in molecular dynamics simulations, which is about a femtosecond. To reach a time scale of tens of

microseconds (typical of the R to T transition in hemoglobin) 10^{10} steps are required. This number of steps is beyond what can be modeled today on readily accessible computers.

The time step of about a femtosecond is necessary in order to maintain the stability of the integration. Considerable efforts were devoted to analyze the basic algorithm and to make this step larger. Many insightful observations on the numerical properties of the algorithm were obtained, and numerous improvements have been proposed [1-3]. The increases of the time step for the integration of the Newton's equations of motion were however small (a factor of several) and did not change the time scale reachable by simulations in a fundamental way.

Elastic models

The first approach extending the accessible timescales that we consider is modeling proteins by elastic networks. A possible realization of the elastic network model is described in a paper by Xu et al. [4]: A connectivity matrix (called Γ) is built for the C_α coordinates of the protein molecule. The connectivity matrix is different from zero for adjacent alpha carbons and is zero otherwise. More specifically

$$\Gamma_{ij} = \begin{Bmatrix} -1 \cdot H_{ij} & i \neq j \\ \sum_{i \neq j} H_{ij} & i = j \end{Bmatrix} \qquad H_{ij} = \begin{Bmatrix} 1 & \text{if } R_{ij} \leq r_c \\ 0 & \text{f } R_{ij} > r_c \end{Bmatrix}$$

The cutoff distance (r_c) was 7Å regardless of the existence (or non-existence) of covalent connectivity between the C_α-s. The connectivity matrix is used in harmonic analysis of protein fluctuations (every contact is translated into a spring) with a single unknown parameter, the spring force constant. The single parameter is fitted to experimental data.

This amazingly simple model for protein fluctuations in the neighborhood of their native state originated by Tirion [5]. Erman, Bahar, and Jernigan exploited this idea in considerably more depth in a series of papers starting from 1997 [6]. They demonstrate remarkable success in reproducing experimentally measured fluctuations (B factors). The recent paper about the conformational transition in hemoglobin mentioned above [4] extends the elastic description of proteins to multi-state systems, attempting to interpolate between the R and T conformations using elastic analysis at the two minima. This analysis suggests a clear path for studying large-scale and slow processes. Another interesting application of material-level description of biological macromolecules was given in [7] in which a rotating flagella used for bacteria propulsion was modeled using a "quantized elastic deformation model." The combination of coarse-grained description of the protein chain and harmonic analysis is especially intriguing. The effective energy of the coarse-grained model is more likely to be harmonic compared to atomically detailed models that experience significant local roughness. The atomic roughness is translated to multiple minima [8] and barriers that cannot be described within the harmonic model. In the same sense that linear elasticity works well for highly complex microscopic material, here an elastic model that captures the overall shape and mass density of the molecule is able to reproduce well the low frequency fluctuations.

Smoothing protocols

Another intriguing idea on how to accelerate molecular dynamics simulation is to change the potential to allow faster exploration of configuration space, and at the same time to accurately describe thermodynamic and kinetic properties. Potential smoothing to assist simulations goes way back to the diffusion transform of Scheraga [9] aiming at global optimization and to the extensive set of tools proposed by Straub [10,11] and Shalloway [12] for energy minimization.

Computations of thermodynamic properties with distorted potentials are now routinely done using the language of umbrella sampling [13] and generalized ensembles (see for instance [14-16]). In approaches for thermodynamic sampling biasing the potential is a useful paradigm to approach the most effective sampling – that of free diffusion, and then correcting for the potential distortion to obtain the exact result.

Applications of potential smoothing to kinetics are less trivial since the interpretation must rely on stronger assumptions. The basic idea is to reduce barrier heights separating alternative minima and accelerate the dynamics of barrier crossing (figure 1).

*** PLACE FIGURE 1 HERE ***

Barrier reduction is achieved (at least partially) by increasing the energy of the minimum for which more information is available. Since the theory of activated processes is well developed it is possible to estimate what the effect of the smoothing on the transition time is likely to be [17]. In an activated process the rate is modified exponentially such that the new time scale t_{new} is given by $t_{new} = t_{old} \exp\left(\frac{\Delta}{kT}\right)$. Clearly, if the change in the barrier height, Δ, is negative a significant reduction in simulation time can be obtained. An adjustment of the results back to the exact result can be done using statistical arguments based on (for example) the transition state theory. The pioneering studies in this regard of Voter (hyperdynamics [18]) and Grubmüller (conformational flooding [17]) come to mind. A recent paper that further enhances this idea was published recently by Hamelberg, Mongan, and McCammon [19]. In this paper the expensive computation of the second derivative matrix, estimating the well curvature, was avoided. In summary, the smoothing approach is very effective in exploring alternative conformations. Once the trajectory gets stuck in a particular minimum, raising the energy of the minimum enables the escape of the system from the local trap, and accelerates sampling of rare events. The modeling of the kinetic relies on the assumption that the process is a series of stochastically independent hopping events. If the process is diffusive or the separation of time scales between vibrations within a minimum and hopping between minima is not so clear other approaches may be more appropriate.

Reaction paths

Considerable renewed interest focuses on a popular approach in chemical physics and its application to biophysical processes, namely calculations of reaction paths and order parameters. An attempt is made to describe the sequence of events along a one-dimensional reaction coordinate without an explicit reference to time. The hope is that once a sequence of events in space is established a variety of procedures can use this

spatial guideline to estimate the true kinetics of the system by using (for example) statistical assumptions of how the system will propagate along this coordinate.
Of particular interest are the global path search methods that are especially appropriate for complex systems with many minima and barriers. These techniques avoid the calculations of Hessians and use information on both reactants and products to increase stability. Notable are the NEB methodology [20], its recent numerical enhancement by Chu, Trout and Brooks [21], and the elegant string method [22]. These techniques are significant improvements over the earlier boundary value approaches to compute reaction paths that were conceptualized by Pratt [23] and developed by Elber et al. [24-27]. Considerable room for improvement still exists in the global optimization of the functionals that define the paths. Nevertheless, the rigorous formulation is attractive, and is advantageous compared to the numerous heuristics that are used in the field (such as the targeted molecular dynamics), for which no guaranteed features of the path are available. Also of considerable interest is the formulation of Vanden Eijnden that takes into account the thermal volume of the trajectories to compute the reaction coordinate. While reaction path formulation taking into account the temperature of the system were proposed in the past (Berkowitz, Morgan, McCammon and Northrup [28], Elber and Shalloway [29], Huo and Straub [30]), the recently proposed solution is particularly elegant and relies on a numerically generated ensemble of paths—a more consistent approach compared to previous studies.

Computing trajectories by optimization of actions

The idea to simulate classical dynamics of biological systems with optimization of actions (instead of solving the Newton's equation of motion) has been promoted in the Elber's group for the last few years [31,32]. In this approach an initial (discrete) guess of the trajectory is used to start a minimization process of a classical action (figure 2).

** PLACE FIGURE 2 HERE **

The global approach to the calculation of the trajectory adds significantly to the numerical stability of the solution and enables the use of much larger steps. The resulting algorithm provides approximate but much longer time trajectories. Passerone, Ceccarelli, and Parrinello introduced interesting enhancements to the action optimization protocol [33].
Recent applications to complex systems include the simulation of the folding of Cytochrome *c* [34] and the folding of a helical peptide [35]. Similarly to the calculation of reaction coordinates, these trajectories provide a sequence of events (e.g. the N and the C helices fold before the 60 helix in Cytochrome *c*), while the actual time scale of the process is hard to come by due to the approximate nature of the trajectories. Further enhancements of this approach by multi-grid methods and the use of these trajectories to define milestones (see below) are likely to enable atomically detailed simulations for long time diffusive (low barrier) processes.

Simulation of rare and fast events

An important class of slow processes is events that occur very quickly but only infrequently. Sampling these trajectories is difficult (since they are rare), but once they

are initiated they are easy to compute since they last for only a short period. A brainchild of Dellago, Boluis, Csajka and Chandler [36,37], the transition path sampling gained in popularity in the last few years, and is a technique of sampling rare trajectories and estimating their statistical weight. The beauty of the technique is that a reaction coordinate is not needed, only a definition of the basin of attraction of a reactant and a product. Perhaps the secret of success of the transition pathway sampling is that the method avoids the need to simulate the considerable waiting time that the system spends in the basin of the reactant, a time that is required to build the correct phase for the hopping event. A recent intriguing application is the study of the last phase of beta hairpin folding by Bulhois [38]. Also of interest is the breakdown of an enzymatic reaction to a sequence of activated processes (studied by the transition sampling approaches) by Schlick [39].

Milestoning

If the process does not have a single dominant barrier (like in the initial phase of protein folding – chain collapse [34]) then the approaches discussed above are not appropriate. Milestoning is a recent theoretical advance that is specifically aimed at diffusive long time processes [40]. This approach uses "milestones" along a reaction coordinate to guide the simulation process. For example, a rotation of a one torsion angle (of many torsions) to a desired value can be considered a milestone in protein folding. The complete folding process is decomposed into a sequence of transitions between milestones. Instead of computing the whole trajectory in one chunk the trajectory is broken into segments and the dynamics of each segment (from one milestone to the next) is computed independently (can be done in parallel). The fragmentation implies statistical independence, which is an essential assumption of the computing protocol.

** PLACE FIGURE 3 HERE **

The validity of the assumption can be controlled by the proper choice of the milestones. Significant saving in human time is obtained due to parallelization [40] and the diffusive nature of the processes that we considered. Eric Vanden Eijnden showed us the following "saving" argument: The diffusion time is roughly proportional to L^2 where L is the length of the path. Dividing the path into N segments of length L/N each ($N-1$ milestones) means that the simulation time of all segments will be L^2/N -- a significant saving. The promise of the milestoning approach is its straightforward application to corrugate energy surfaces that are difficult to address with other approaches. Previously mentioned techniques require clear transition states and (or) separation of timescales. Bulhois and co-workers put forward a related approach that assumes a pure Markovian behavior [41].

Massively distributed computing

The use of grid computing, hundreds of thousands of computing nodes, to study protein folding was pioneered by Pande who already carried out a number of important and extremely difficult calculations [42-46]. This unique resource that provides hundreds to thousands times more computing power than what we typically find in a typical computing laboratory enables straightforward studies at domains we had not thought

possible before. Of course, a larger number of computers do not necessarily mean long time trajectories since time steps are made in sequential manner and are difficult to parallelize. The clever parallelization implemented by Pande relies on exponential kinetics. Pande is running a large number of independent trajectories. If a single time scale dominates the process (which is the prime assumption) observing a few short-time reactive trajectories is possible (the probability is low but not zero), and further interpolation to determine the rate can be made. An impressive set of small peptides have been folded to date, suggesting grid computing to be one of the most useful tools around for the study of activated processes in a straightforward (almost assumption-free) manner.

In summary, numerous approaches to long-time simulations were briefly discussed. While significant advances are expected along many of these lines, an interesting direction in the future will be to combine some of these techniques. We anticipate that the simulations of sub-millisecond phenomena will become more accessible in the next few years and simulations of microsecond processes routine.

Acknowledgement: The review was supported by an NIH grant GM59796 to the author. The author is grateful to Stacey Shirk for her editorial assistance.

Figure legends

1. A schematic drawing of a smoothing of an energy surface. The drawing presents an increase in the energy value of the relevant minimum without affecting the barrier height.
2. A schematic drawing of the action optimization approach on a two-dimensional energy surface. The two end points (R and P) are fixed and a trajectory interpolating the two end points is computed (shown by thick line) by optimizing the intermediate structures to give a stationary classical action. The action is a function of all the intermediate structures X_i, and the small arrows denote directions of minimization
3. A schematic drawing of a reaction coordinate on a two-dimensional energy surface and sequential planes perpendicular to it (representing the milestones). Also shown is a trajectory (thick line) that starts at the second slice from the right (denoted by *i*) and terminates at the third slice from the right. Trajectories initiated at each of the milestones are used to estimate the first passage time distribution between the milestones. These distributions are used in modeling the overall transition time.

Figure 1

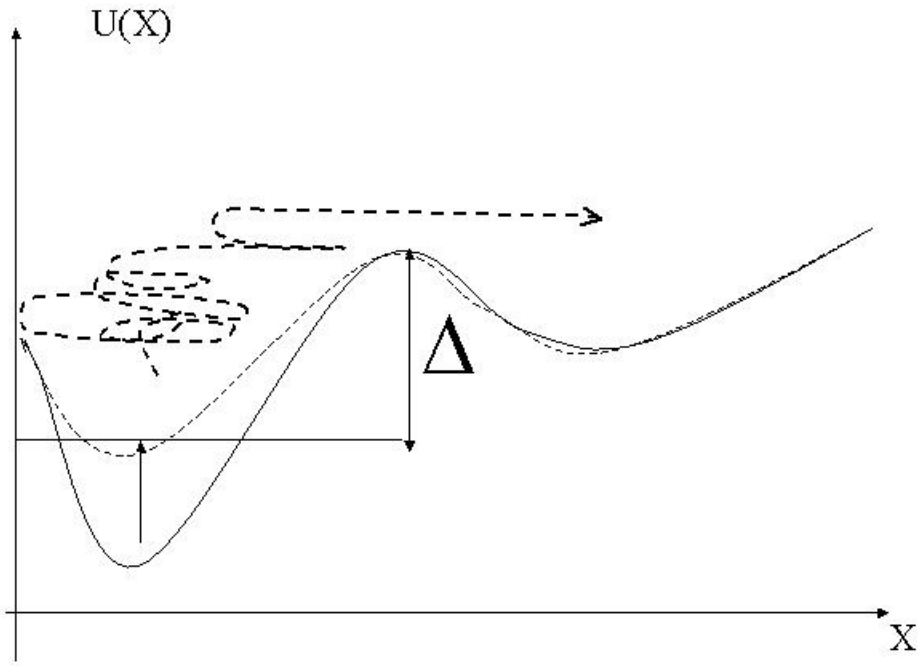


Figure 2

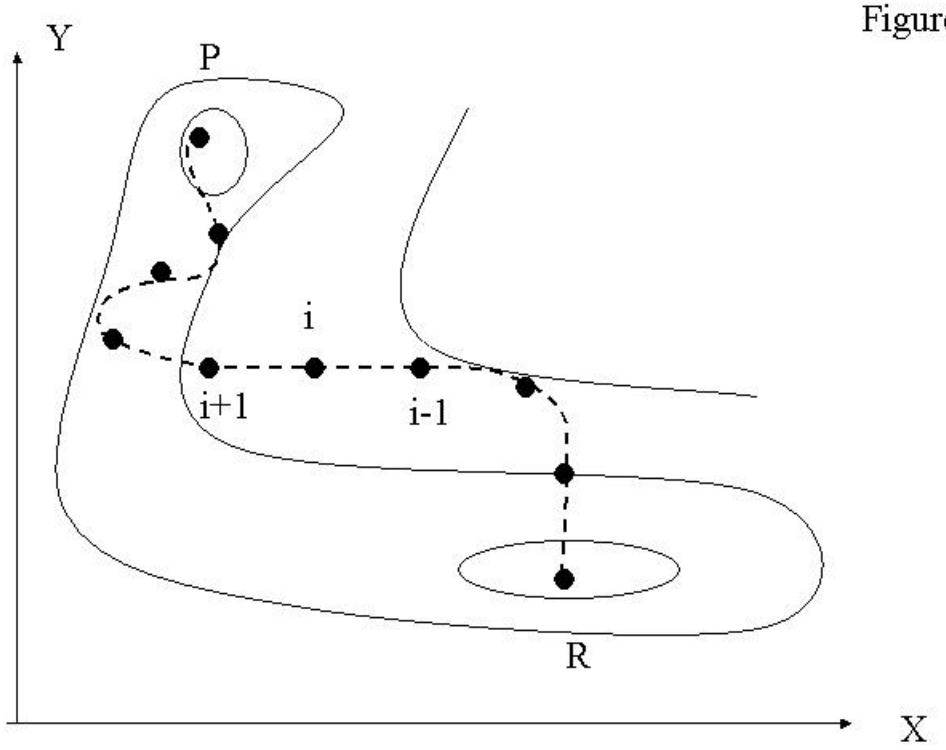
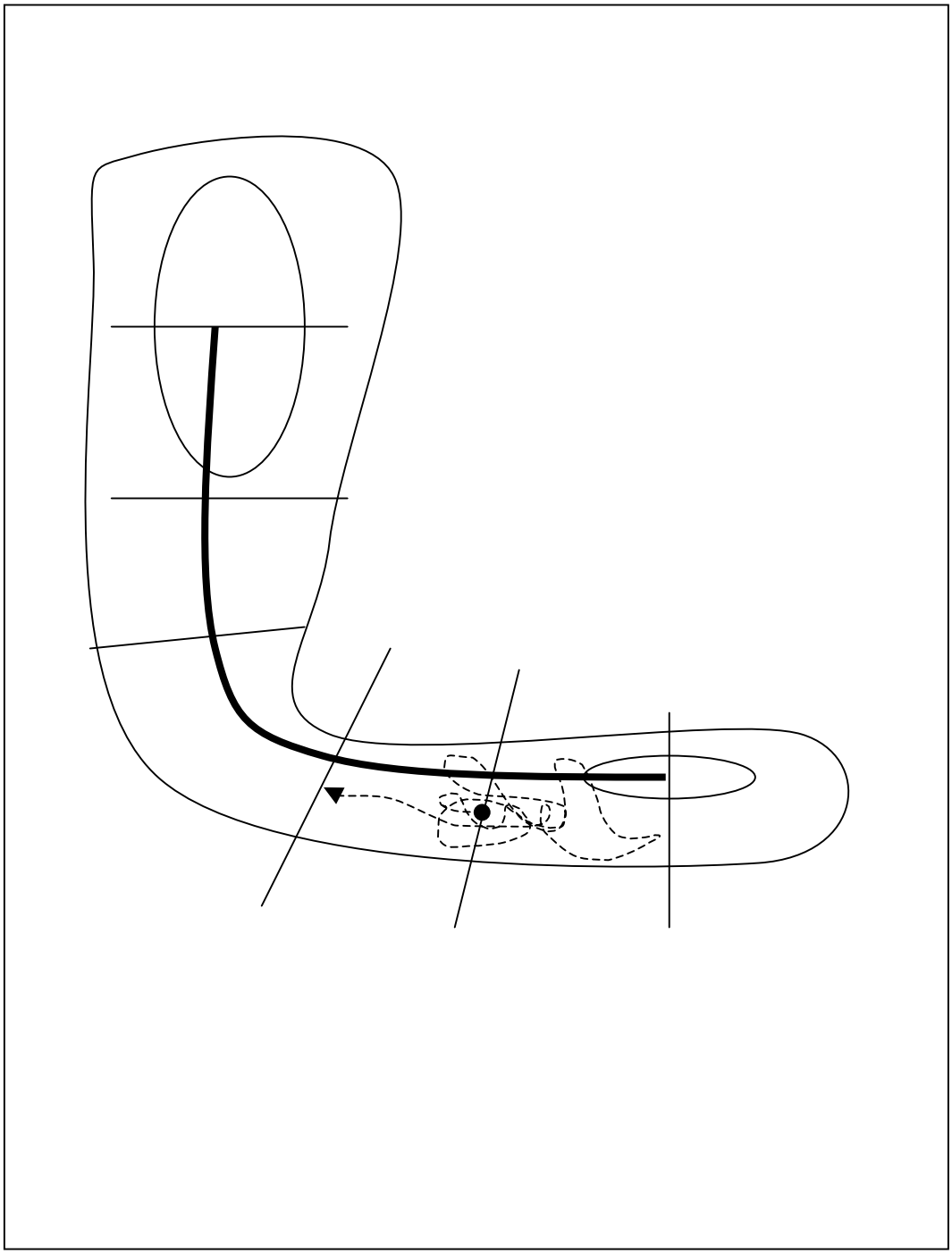


Figure 3.



References

*1. Ma Q, Izaguirre JA, Skeel RD: **Verlet-I/r-RESPA/Impulse is limited by non-linear instabilities**. *SIAM Journal on scientific computing* 2003, **24**:1951-1973.

An interesting negative paper suggesting that we are at a "theoretical limit" solving initial value problems.

2. Tuckerman M., Berne BJ, Martyna G.J.: **Reversible multiple time scale molecular dynamics**. *Journal of Chemical Physics* 1992, **97**:1990-2001.

3. Barash D, Yang LJ, Qian XL, T. S: **Inherent speedup limitations in multiple time step/Particle Mesh Ewald algorithms**. *Journal of Computational Chemistry* 2003, **24**:77-88.

*4. Xu C, Tobi D, Bahar I: **Allosteric changes in protein structure computed by a simple mechanical model: Hemoglobin T<->R2 transition**. *Journal of Molecular Biology* 2003, **333**:153-168.

A new interesting direction for elastic networks. It is attempted to interpolate between quadratic basins of elastic networks to estimate directions of transitional events. While such interpolations were used in the past with harmonic analysis of atomic models, the extension and success of drastically simplified models is intriguing.

5. Tirion MM: **Large amplitude elastic motions in proteins from a single parameter atomic analysis**. *Physical Review Letters* 1996, **77**:1905-1908.

6. Bahar I, Erman B, Haliloglu T, Jernigan RL: **Efficient characterization of collective motions and interresidue correlations in proteins by low resolution simulations**. *Biochemistry* 1997, **36**:13512-13523.

*7. Flynn. T.C., Ma, J.P.,: **Theoretical analysis of twist/bond ratio and mechanical moduli of bacterial flagellar and filament**. *Biophysical Journal* 2004, **86**:3204-3210.

An amusing application of continuum/elastic description of a system that is going beyond typical molecular scales

8. Elber R, Karplus M: **Multiple Conformational States of Proteins: A molecular dynamics analysis of myoglobin**. *Science* 1987, **235**:318-321.

9. Pillardy J, Liwo A, Groth M, Scheraga HA: **An efficient deformation-based global optimization method for off-lattice polymer chains: Self-consistent basin-to-deformed-basin mapping (SCBDBM). Application to united-residue polypeptide chains**. *Journal of Physical Chemistry B* 1999, **103**:7353-7366.

10. Andricioaei I, Straub JE: **Global optimization using bad derivatives: Derivative-free method for molecular energy minimization**. *Journal of Computational Chemistry* 1998, **19**:1445-1455.

11. Whitfield TW, Straub JE: **Gravitational smoothing as a global optimization strategy**. *Journal of Computational Chemistry* 2002, **23**:1100-1103.

12. Church BW, Shalloway D: **Top-down free-energy minimization on protein potential energy landscapes**. *Proceeding of the National Acadamy of Science USA.* 2001, **98**:6098-6103.

13. Torrie GM, Valleau JP: **Non-physical sampling distributions in Monte-Carlo free energy estimation - umbrella sampling"**. *Journal of Computational Physics* 1977, **23**:187.

14. Peng Y, Hansmann U, H.,E.,, Alves NA: **Solution effects and the order of the helix-coil transition in polyalanine**. *Journal of Chemical Physics* 2003, **118**:2374-2380.
15. Mitsutake A, Sugita Y, Okamoto Y: **Replica-exhange multicanonical and multicanonical replica exchange Monte Carlo simulations of peptides. I. Formulation and benchmarks**. *Journal of Chemical Physics* 2003, **118**:6664-6675.
16. Okamoto Y: **Protein folding simulations by simulated annealing and generalized-ensemble algorithms**. *Progress of theoretical physics supplement* 2000, **138**:301-310.
*17. Muller EM, de Meijere A, Grubmüller H: **Predicting unimolecular chemical reactions: Chemical flooding**. *Journal of chemical physics* 2002, **116**:897-905.
*18. Voter AF, Montalenti F, Germann TC: **Extending the time scale in atomistic simulation of materials**. *Annual Review of Materials Research* 2002, **32**:321-346.
The above two papers propose an "umbrella-sampling" like procedure for simulations of kinetics, in which sampling of transitional events is artificially enhanced and then corrected to the physical value using statistical theory.
19. Hamelberg D, Mongan J, McCammon JA: **Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules**. *Journal of Chemical Physics* 2004, **120**:11919-11929.
20. Jonsson H, Mills G, Jacobsen KW: **Nudge Elastic Band Method for Finding Minimum Energy Paths of Transitions**. In *Classical and Quantum Dynamics in Condensed Phase Simulations*. Edited by Berne BJ, Ciccotti G, Coker DF: World Scientific; 1998:385.
21. Chu JW, Trout BL, Brooks BR: **A super-linear minimization scheme for the nudged elastic band method**. *Journal of Chemical Physics* 2003, **119**:12708-12717.
22. Weinan E, Weiqing R, Vanden-Eijnden E: **String method for the study of rare events. *Physical Review B* 2002, **66**:52301.
An elegant and efficient implementation of a reaction path algorithm applicable for highly corrugated systems
23. Pratt LR: **A statistical-method for identifying transition-states in high dimensional problems**. *Journal of Chemical Physics* 1986, **85**.
24. Elber R, Karplus M: **A method for determining reaction paths in large molecules: Application to myoglobin**. *Chemical Physics Letters* 1987, **139**:375.
25. Ulitsky A, Elber R: **A New Technique to Calculate Steepest Descent Paths in Flexible Polyatomic Systems**. *Journal of Chemical Physics* 1990, **92**:1510-1511.
26. Czerminski R, Elber R: **Self-Avoiding Walk between 2 Fixed-Points as a Tool to Calculate Reaction Paths in Large Molecular-Systems**. *International Journal of Quantum Chemistry* 1990, **24**:167-186.
27. Olender R, Elber R: **Yet another look at the steepest descent path**. *Theochem-Journal of Molecular Structure* 1997, **398**:63-71.
28. Berkowitz M., Morgan J.D., McCammon J.A., Northrup J.H.: **Diffusion-controlled reactions: A variational formula for the optimum reaction coordinate**. *Journal of Chemical Physics* 1983, **79**:5563-5565.

29. Elber R, Shalloway D: **Temperature dependent reaction coordinates**. *Journal of Chemical Physics* 2000, **128**:118-127.
30. Huo SH, Straub JE: **The MaxFlux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body systems at finite temperature**. *Journal of Chemical Physics* 1997, **107**:5000-5006.
31. Olender R, Elber R: **Calculation of classical trajectories with a very large time step: Formalism and numerical examples**. *Journal of Chemical Physics* 1996, **105**:9299-9315.
32. Elber R, Ghosh A, Cardenas A, Stern H: **Bridging the gap between reaction pathways, long time dynamics and calculation of rates**. *Adv. Chem. Phys.* 2003, **126**:93-129.
33. Passerone D, Ceccarelli M, Parrinello M: **A concerted variational strategy for investigating rare events**. *Journal of Chemical Physics* 2003, **118**:2025-2032.
*34. Cardenas AE, Elber R: **Kinetics of cytochrome C folding: Atomically detailed simulations**. *Proteins-Structure Function and Genetics* 2003, **51**:245-257.
This study presents approximate atomically detailed molecular dynamics trajectories for a highly complex folding process with significant success in reproducing the orders of events in protein folding.
35. Cardenas A, Elber R: **Atomically detailed simulations of helix formation with the stochastic difference equation**. *Biophysical Journal* 2003, **85**:2919-2939.
36. Bolhuis PG, Chandler D, Dellago C, Geissler PL: **Transition path sampling: Throwing ropes over rough mountain passes, in the dark**. *Annual Review of Physical Chemistry* 2002, **53**:291-318.
37. Dellago C, Bolhuis PG, Csajka FS, Chandler D: **Transition path sampling and the calculation of rate constants**. *Journal of Chemical Physics* 1998, **108**:1964-1977.
38. Bolhuis PG: **Transition-path sampling of beta-hairpin folding. *Proceeding of the National Acadamy of Science USA*. 2003, **100**:12129-12134.
A most impressive application of the transition path sampling approach to the folding of a model peptide (beta hairpin)
39. Radhakrishnan R, T. S: **Orchestration of cooperative events in DNA synthesis and repair mechanism unraveled by transition path sampling of DNA polymerase beta's closing**. *Proc. Natl. Acad. Sci USA* 2004, **101**:5970-5975.
40. Faradjian A. K, R. E: **Computing time scales from reaction coordinates by milestoning**. *J. Chem. Phys.* 2004, **120**:10880-10889.
41. Moroni D, Bolhuis PG, van Erp TS: **Rate constants for diffusive processes by partial path sampling**. *J. Chem. Phys.* 2004, **120**, 4055-4065.
* The above two papers (40 & 41) suggest a novel and interesting way to deal with corrugated surfaces with many but relatively small barriers that lead to long time diffusion processes. An example for a study appropriate for the above approach is a first step in protein folding that includes chain collapse to the molten globule state.
*42. Pande VS: **Simulations of the role of water in the protein folding mechanism**. *PNAS* 2004, **101**:6456-6461.

An impressive and computationally expensive study of the role of waters in protein folding using ten of thousands of independent trajectories on the author's unique computational resource

43. Pande VS: **Meeting halfway on the bridge between protein folding theory and experiment**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**:3555-3556.
44. Pande VS, Baker I, Chapman J, Elmer SP, Khaliq S, Larson SM, Rhee YM, Shirts MR, Snow CD, Sorin EJ, et al.: **Atomistic protein folding simulations on the sub millisecond time scale using worldwide distributed computing**. *Biopolymers* 2003, **68**:91-109.
45. Shirts MR, Pitera JW, Swope WC, Pande VS: **Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force field for proteins**. *Journal of Chemical Physics* 2003, **119**:5740-5761.
46. Larson SM, England JL, Desjarlais JR, Pande VS: **Thoroughly sampling sequence space: Large-scale protein design of structural ensembles**. *Protein science* 2002, **11**:2804-2813.