Novel methods for molecular dynamics simulations

Ron Elber

In the past year, significant progress was made in the development of molecular dynamics methods for the liquid phase and for biological macromolecules. Specifically, faster algorithms to pursue molecular dynamics simulations were introduced and advances were made in the design of new optimization algorithms guided by molecular dynamics protocols. A technique to calculate the quantum spectra of protein vibrations was introduced.

Address
Department of Physical Chemistry and Biological Chemistry, The Fritz Haber Research Center and the Wolfson Center for Applied Structural Biology, The Hebrew University, Jerusalem 91904, Israel; e-mail: ron@batata.fh.huji.ac.il

Current Opinion in Structural Biology 1996, 6:232–235

Abbreviations

LIN	Langevin, implicit Euler and normal mode scheme for molecular dynamics integration
MD	molecular dynamics
RESPA	reference system propagator algorithm

Introduction

Molecular dynamics (MD) simulations aid the understanding of biomolecular properties on several fronts. Firstly, MD simulations provide a qualitative 'feel' for the range of motions and fluctuations in biomolecular structures [1,2]. Secondly, the simulations are employed in the calculation of thermodynamic properties such as conformational free energy or structural fluctuations [3]. Thirdly, MD simulations are also useful in computing short-time dynamics of biochemical processes [4,5].

Throughout this review I will refer to MD simulations as computations that describe the biomolecule and the solvent in detail. Considerable progress was made in the design and application of implicit approximations of solvation and of reduced models of biomolecules (Jernigan and Bahar, this issue, pp 195–209). It is obvious that these models are enormously cheaper (computationally) than a full atomic description of the system. At present, reduced models are our prime hope in exploring long-time processes and large-scale conformational transitions such as protein folding. Nevertheless, reduced models have their limitations. For example, when an enzymatic reaction is investigated, it is difficult to study the reaction progress without a picture of all the relevant atoms at the active site [6]. This information is difficult to obtain from a model with a spatial resolution of a single amino acid. Furthermore, reduced models of biomolecules embedded in solvents other than water (such as the lipid environment) are difficult to generate, due to the lack of calibration data. Atomically detailed models of arbitrary solvents are more straightforward to construct.

Yet another useful feature of atomic models of MD protocols is their relatively small variation from research group to research group. Nowadays, atomic force fields use essentially the same functional form and the variations in the parameter sets are relatively small. In contrast, the variations in reduced models of biomolecular interactions are wide and wild, both in the functional form and in the value of the parameters. This is not surprising, considering the relatively 'young' age of the reduced models and their parameterization. The current number of creators of reduced force fields is also significant. Assuming, however, that only one correct force field exists, there is a need to sort out the different suggestions. One possible approach is to compare simulations with simulations or with experimental data on small molecular systems. One field in which a useful comparison can be made is that of solvation energies, for which a wealth of experimental and atomically detailed computational data is available. Comparison to solvation energies is an approach already taken by other groups [7].

After the promotion of the atomically detailed simulations, it is about time to clearly state their limitations. The simulations usually suffer from severe constraints. The constraints are on the timescales that can be explored and on the conformational space that can be sampled during a single trajectory. In this review, I am focusing on research that aims to reduce the limitations of MD. The length of MD simulations are limited at present to a few nanoseconds. This timescale is not sufficient to explore large-scale conformational changes, which is the reason for the restriction on the searches of conformational space in addition to the restriction on the timescale.

Searches for plausible reaction coordinates

One way of getting around the problem is to change the parameters of the simulation, for example, the temperature. A higher temperature speeds the atomic motions and makes it possible to examine significant structural changes at short times. This approach was employed primarily in unfolding simulations [8], in which a folded configuration was 'destroyed' and the sequence of events of the unfolding process was examined. This approach makes it possible to suggest qualitative mechanisms on the dynamics and the structure of unfolding intermediates. Nevertheless, an essential question is how relevant the results at the elevated temperatures are to ordinary (room temperature) conditions.

Another approach in which protein folding can be investigated is the application of the reaction coordinate

method. In this scheme, a reaction coordinate is assumed, and an effective energy profile is computed assuming local equilibrium. A biasing force is employed to drive the system along the reaction coordinate and at each point, the free energy of the system (the potential of mean force) is computed. The free-energy profile provides a basis for a kinetic theory of the process dynamics [9•]. The calculations of the potential of mean force are expensive and difficult and raise the question of proper sampling of conformational space. The big question is how the computed folding mechanism is influenced by the way in which it was induced on the computer. Two possible ways of unfolding proteins were mentioned above: the high-temperature runs (which are the most popular [8]) and the ad hoc set-up of a reaction coordinate [9•]. Other options are the addition of denaturants [10], and pressure increase [11••]. It is important to compare the different pathways, because interpolation to different conditions (close to room temperature and to one atmosphere) is made. Such interpolation is more likely to be correct if the general features of the folding pathways are similar for the different computational perturbations. This, unfortunately, is not the case.

New molecular dynamics integration

An interesting paper [11••] addressed the above question. It examined different approaches to induce unfolding in lysozyme. The perturbations were temperature, pressure and an assumed reaction coordinate for the unfolding. It was shown in this study that the unfolding process was qualitatively modified by the method in which it was induced. Great care must therefore be exercised in the interpretation of unfolding computer experiments.

Raising the temperature serves two purposes. It accelerates the motions of the atoms and also enhances the sampling of different conformations. The problem is that the conformations are sampled with a weight different to what we expect in room-temperature simulations. The trajectory tends to visit higher energy structures. A way of 'speeding up' the atomic motions while maintaining the correct sampling is to increase the time step. This is a problem in numerical analysis more than in biopolymers but it is clearly of great significance. Progress in introducing better algorithms was, however, slow. Obstacles to maintaining the stability and the accuracy of the trajectories made it difficult to come up with a new algorithm with a much larger time step. Though slow, progress has been made, and one of the most promising directions is the RESPA (reference system propagator algorithm) [12,13••]. In RESPA, the equations of motion are integrated using multiple timescales and partitioning to fast and slow ranging forces. The partitioning to quickly and slowly varying interactions is primarily based on chemical and physical intuition (it would be nice if this partition was done automatically too). Perhaps the most impressive achievement of the RESPA is the extensive work invested in the optimization of a variety of features. A recent publication [13••] demonstrated this approach and further advanced the RESPA to include fast multipole expansion of the long-range forces. The success of the RESPA is also evident in the re-confirmation of its favorable properties by another group that discusses its application [14].

An automatic approach which does not rely on an *a priori* identification of fast and slow coordinates is the LIN (Langevin, implicit Euler and normal mode scheme for molecular dynamics integration) technique [15]. This method enables the use of a very large time step. However, the calculations associated with a single step require the manipulation of the second derivative matrix. The single step is therefore significantly more expensive compared to the more common approaches that employ only the forces (first derivatives of the potential energy). At present, LIN does not speed things up that much. Nevertheless, it is a promising lead, since not all possibilities of speeding up the individual step were exhausted (for example, exploring the fact that the matrix is highly sparse) and further improvement in the efficiency of the protocol is expected.

Another approach that makes it possible to increase the time step somewhat is the SHAKE [16] or the related RATTLE [17] algorithm. It is a common practice today to push the time step in regular MD simulations to 1 or even 2 fs, and to employ the above algorithms to constrain the fastest degrees of freedom, the bond lengths, to their equilibrium position. SHAKE and RATTLE are iterative procedures in which the results of an unconstrained step are corrected in an iterative way to satisfy the constraints. A detailed and careful numerical analysis of the SHAKE and the RATTLE algorithms, along with a number of useful tricks on how to enhance the rate of convergence of these algorithms, was published recently [18•]. This paper is useful reading for those who employ constraints in MD applications.

It is worth stressing that dynamics with constraints are different from dynamics without them. For example, if accurate time correlation functions are of interest, it is necessary to integrate the equations of motion explicitly. This is another useful feature of the RESPA in addition to the increase in the time step. RESPA makes it possible to obtain more accurate trajectories because the fast degrees of freedom are integrated separately.

MD trajectories are the solutions of first order differential equations of motion. It appears to be difficult to speed up the computations by more than one order of magnitude by advancing the numerical algorithms only. A promising approach gives up on the precision maintained in the above two algorithms, but maintains the stability of the solution for an arbitrary time step [19]. This is the prime advantage of the new algorithm. The method is based on the use of functionals to compute stochastic trajectories (R Olender, R Elber, unpublished data). In

this approach, an optimized functional is computed, as in the quantum path integral approach [20]. In the quantum path integral, the optimized trajectory corresponds to the classical path. Using the functional of Onsager and Machlup [20], one obtains an optimized trajectory which corresponds to the most probable stochastic trajectory between two configurations. This technique is therefore useful in the investigation of processes in which the reactant and the product structures are known, and it is conceptually different from trajectory calculations that rely on initial conditions only. Clearly, the considerable experience acquired in the past in the computations of path integrals also benefits this new approach for trajectory calculations.

Enhanced sampling

Going back to the issue of timescales and sampling in classical mechanics models of proteins, we note that until now, only the time step was discussed. Searching for faster ways of computing MD trajectories (and getting more steps at the same computational cost) is one solution to the sampling problem. An alternative is to get more than one structure in a single time step. Unfortunately there is no such thing as a free lunch in the MD field and such an increase in the sampling is achieved at the cost of the use of an approximation. The approximations involved limit the range of applicability of these algorithms. The algorithms that enhance sampling in the above way are primarily used as optimization tools in which a more effective exploration of alternate configurations is achieved.

In one approximate technique, the probability density of the particle is assumed to be a Gaussian function [21•]. Point particles (ordinary trajectories) are completely defined by their x,y and z coordinates. In the new scheme, an additional parameter, the Gaussian width, is needed to describe the system and it therefore adds flexibility to the representation of the physical model. An exact (ordinary) trajectory is obtained in the limit in which the Gaussian width is infinitely narrow. However, the approximate trajectory that is obtained with a finite Gaussian width for the probability density has some advantages. For example, the effective energy surface that is felt by the center of the Gaussian is smoother than the original potential felt by the point particles. Therefore, an annealing scheme and searches for the global energy minimum are much more effective with the finite Gaussian width as compared to regular annealing with a 'Gaussian' of zero width (a regular MD trajectory). This protocol is an extension of the diffusion equation approach [22,23], in which direct minimization is employed on Gaussian transformed potentials.

Yet another smoothing technique of the energy surface for more rapid optimization is the LES (locally enhanced sampling) method [24]. In LES, the probability density is approximated by a sum of mean-field trajectories. A possible set up is of several copies of a peptide that are embedded in a single box of water. The different copies of the peptide do not see each other and they feel only the full force of the single copy of the water box. On the other hand, the water molecules feel the average (mean) force of the peptide copies. The averaging smoothes the energy surface and the multiple copies of the peptide make it possible to obtain more statistics for alternative conformations. The smoothing in LES is less effective than in the Gaussian transforms, since the density (represented here by the sum of trajectories) is fluctuating rapidly. However, LES has other advantages. In a single time step, many different conformations of the peptide can be analyzed by consideration of the structures of each of the individual copies. The LES was recently employed in the determination of the structure of small peptides in explicit water solution [25], a difficult calculation without the help of LES. The role of hydrophobic interactions as a drive to structure in small peptides was demonstrated.

Vibrational analysis

MD of biological systems is primarily aimed at reproducing structures and the energetics of solvated macromolecules. The force fields and the computational methods available are insufficient for the accurate comparison of computations to spectroscopic data. Spectroscopy was and is used extensively to refine force fields of small molecules and to follow their dynamics. It is therefore desirable to extend these techniques, together with the appropriate interpretation tools, to large biomolecules. I restrict the discussion below to spectroscopy that requires a quantum mechanical description of the system. There is a wide range of experimental techniques that can provide this type of information. Unfortunately, little computational methodology is available to interpret the results quantitatively.

The only well established technique capable of the quantum mechanical description of protein dynamics and spectroscopy is the harmonic approximation in conjunction with the normal-mode analysis. In the normal-mode method, an approximation of small displacements is employed. The molecular system is assumed to deviate only slightly from the equilibrium position, resulting in forces linear in the displacement. An analytical solution (within the harmonic approximation) for the time-dependent properties can be obtained. The normal-mode approach is further advanced, either computationally [26], or as an analysis tool [27]. In analysis, the normal-mode framework suggests a scheme to interpret results of full MD trajectories. Unfortunately, the fundamental assumption of the normal-mode analysis (small displacements in the neighborhood of a single minimum) is questionable in proteins, and it is important to go beyond the harmonic approximation if a quantitative description of individual modes (and spectral lines) is desired. Such a computational technique for quantum vibrational analysis of proteins,

which is going beyond the normal mode analysis [28] was recently introduced. The effect of anharmonic coupling in the interaction potential was computed using a mean-field approach. The method — the self-consistent field approach — was demonstrated to accurately describe vibrations of small molecules [29]. It was applied to a protein for the first time, showing the significant effect of anharmonicity (deviation from linearity in the force) on the frequencies and on the wavefunctions.

Conclusions

In the last year, we have seen a continuation of ongoing refinement of MD algorithms. Gradual but significant improvements in numerical algorithms are now apparent, making it possible to speed up computations by a factor of about 10. In addition to the development of algorithms for exact solutions of the equations of motion, approximate methodologies for more rapid exploration of conformational space were further developed and tested. In a new line of research, anharmonic quantum mechanical calculations of wavefunctions relevant to spectroscopy were pursued for the first time using the self-consistent field method.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. McCammon JA, Harvey SC: *Dynamics of proteins and Nucleic Acids.* Cambridge: Cambridge University Press; 1987.
2. Brookes CL III, Karplus M, Pettitt BM: *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics.* New York: John Wiley & Sons; 1988.
3. Kirshenbaum K, Daggett V: **Sequence effects on the conformational properties of amyloid β (1–28) peptide: testing a proposed mechanism for the α→β transition.** *Biochemistry* 1995, **34**:7640–7647.
4. Li H, Elber R, Straub JE: **Molecular dynamics simulations of NO recombination to myoglobin mutants.** *J Biol Chem* 1993, **266**:17908–17916.
5. Schaad O, Zhou HX, Szabo A, Eaton WA, Henry ER: **Simulation of the kinetics of ligand binding to a protein by molecular dynamics: geminate rebinding of nitric oxide to myoglobin.** *Proc Natl Acad Sci USA* 1993, **90**:9547–9551.
6. Warshel A, Hwang JK, Aqvist J: **Computer simulations of enzymatic reactions: examination of linear free-energy relationships and quantum mechanical corrections in the initial proton-transfer step in carbonic anhydrase.** *Faraday Discuss* 1992, **93**:225–238.
7. Sitkoff D, Sharp KA, Honig B: **Correlating solvation free energies and surface tensions of hydrocarbon solutes.** *Biophys Chem* 1994, **51**:397–403.
8. Hirst JD, Brooks CL III: **Molecular dynamics simulations of isolated helices of myoglobin.** *Biochemistry* 1995, **34**:7614–7621.
9. • Boczko EM, Brooks CL III: **First principles calculation of the folding free energy of a three helix bundle protein.** *Science* 1995, **269**:393–396.

An application of an assumed reaction coordinate (the radius of gyration) to the construction of the potential of mean force of unfolding of protein A.

10. Schiffer CA, Doetsch V, Wüthrich K, Van Gunsteren WF: **Exploring the role of the solvent in the denaturating of a protein: a molecular dynamics study of DNA binding domain of the 434 repressor.** *Biochemistry* 1995, **34**:15057–15067.
11. •• Huenenberger PH, Mark AE, Van Gunsteren WF: **Computational approaches to study protein unfolding: hen egg white lysozyme as a case study.** *Proteins* 1995, **21**:196–213.

This paper describes a detailed and careful comparison of four different computational tricks to induce the unfolding of a protein. An important conclusion of this study is that the mechanisms and sequence of events depend on the way in which the folding was induced. For example, the unfolding depends on the location of the zero of the radial force that is used to 'expand' the protein.

12. Tuckerman ME, Berne BJ, Martyna GJ: **Reversible multiple time scale molecular dynamics.** *J Chem Phys* 1992, **97**:1990–2001.
13. •• Zhou R, Berne BJ: **A new molecular dynamics method combining the reference system propagator algorithm with a fast multipole method for simulating proteins and other complex systems.** *J Chem Phys* 1995, **104**:9444–9459.

A highly optimized algorithm to pursue MD simulations is described. In addition to the introduction of multiple time steps, the long-range forces are computed with multipole expansion leading to a saving in computer time of up to a factor of 20.

14. Watanabe M, Karplus M: **Simulations of macromolecules by multiple time-step methods.** *J Phys Chem* 1995, **99**:5680–5697.
15. Derreumaux P, Schlick T: **Long time step dynamics of peptides by the dynamics driver approach.** *Proteins* 1995, **21**:282–302.
16. Ryckaert JP, Ciccotti G, Berendsen HJC: **Numerical integration of Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.** *J Comput Phys* 1977, **23**:327–341.
17. Andersen HC: **Rattle: a velocity version of the SHAKE algorithm for molecular dynamics calculations.** *J Comput Phys* 1983, **52**:24–34.
18. • Barth E, Kuczera K, Leimkuhler B, Skeel RD: **Algorithms for constrained molecular dynamics.** *J Comput Chem* 1995, **16**:1192–1209.

An analysis of many computational approaches to pursue MD with constraints. A discussion on convergence properties and a trick to speed up the convergence of SHAKE (over relaxation) are provided.

19. Simmerling C, Elber R, Zhang J: **MOIL-View – a program for visualization of structure and dynamics of biomolecules and STO – a program for computing stochastic paths.** In *Modeling of Biomolecular Structures and Mechanisms.* Edited by Pullman A, Pullman B, Jortenr J. Amsterdam: Kluwer; 1995:241–265.
20. Feynman RP, Hibbs AR: **Quantum mechanics and path integrals.** New York: McGraw-Hill; 1965.
21. • Straub JE, Ma J, Amara P: **Simulated annealing using coarse grained classical dynamics: Smoluchowski dynamics in the Gaussian density approximation.** *J Chem Phys* 1995, **103**:1574–1581.

An annealing algorithm, in conjunction with smoothing the energy surface, is proposed and demonstrated to be a very useful alternative to other optimization protocols.

22. Piela L, Kostrowicki J, Scheraga HA: **The multiple-minima problem in the conformational analysis of molecules. Deformation of the potential energy hypersurface by the diffusion equation method.** *J Phys Chem* 1989, **93**:3339–3346.
23. Shalloway D: **Application of the renormalization group to deterministic global minimization of molecular conformation energy functions.** *J Global Optimization* 1992, **2**:281–311.
24. Roitberg A, Elber R: **Modeling side chains in peptides and proteins: application of the locally enhanced sampling (LES) and the simulated annealing methods to find minimum energy conformations.** *J Chem Phys* 1991, **95**:9277–9287.
25. Simmerling C, Elber R: **Computer determination of peptide conformations in water: different roads to structure.** *Proc Natl Acad Sci USA* 1995, **92**:3190–3193.
26. Brooks BR, Janezic D, Karplus M: **Harmonic analysis of large systems. I. Methodology.** *J Comput Chem* 1995, **16**:1522–1542.
27. Hayward S, Kitao A, Go N: **Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis.** *Proteins* 1995, **23**:177–186.
28. Roitberg A, Gerber RB, Elber R, Ratner MA: **Anharmonic wavefunctions of proteins: quantum self-consistent field calculations of BPTI.** *Science* 1995, **268**:1319–1322.
29. Gerber RB, Ratner MA: **Self consistent field method for vibrational excitation in polyatomic systems.** *Adv Chem Phys* 1988, **70**:97–132.