Long-Timestep Biomolecular Dynamics Simulations: LN Performance on a Polymerase β / DNA System [1]

Tamar Schlick* and Linjing Yang

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

Key words: molecular dynamics, Langevin dynamics, multiple timesteps, biomolecular applications, polymerase/DNA system

* Corresponding author: schlick@nyu.edu; Fax (212) 995-4152.

Abstract

Multiple-timestep (MTS) integrators for biomolecular simulations are promising approaches for enhancing computational efficiency. To overcome resonance limitations on the outer timestep size, nonsymplectic variants can be used. A Langevin/MTS algorithm termed LN based on extrapolative force splitting is described that alleviates these resonance artifacts and allows larger timesteps. LN performance is detailed on a large biomolecular system of 45,000 atoms in terms of energetics, dynamics, and geometry. These results show good reproduction of small-timestep trajectories and substantial computational gains.

1 Introduction

The structural details of biomolecules are of great importance because of the crucial biological link between structure and function. With the advent of supercomputers, workstations clusters, and parallel computing software, molecular modeling and simulation has become a standard partners to experiment [1]. Molecular dynamics (MD) simulations, in particular, offer a powerful tool to refine experimental models and to probe systematically how molecules fold and reshape to perform the basic functions of life. Some problems that are beyond experiment can also be tackled by simulation.

Yet several practical limitations face biomolecular modelers. They include the approximate nature of the governing force fields as well as simulation protocols, the limited range of configurational sampling and relatively short trajectory times, and the enormous computational requirements. New algorithmic approaches, hierarchical spatial representations, and improved computing platforms are needed to enhance the reliability of macromolecular simulations and approach biological time frames (see [2, 3], for example, for reviews of methods).

[1]This article is based in part on a presentation at the Eilat Workshops on Multiscale Computations, April 5–11, 2000: I. NATO Advanced Research Workshop on Multiscale Computation in Chemistry and Biology (April 5-7), and II. Israel Science Foundation Workshop on Multiscale Computational Methods in Chemistry (April 9-11).

One area of intense focus has been the design on efficient, long-time integrators for biomolecular dynamics. This contribution describes multiple-timestep (MTS) integrators for biomolecular simulations, including their resonance limitations. A Langevin/MTS algorithm termed LN that alleviates these resonance artifacts and allows larger timesteps is described, with recent performance on a large biomolecular system detailed.

2 Verlet: The Gold Standard

Classical Newtonian physics forms the basis to standard MD simulations. We simulate molecular motion of a system of N atoms under the influence of a specified, position-dependent force field $F(X)$ by propagating recursively configurations (X) and velocities (V) as a function of time t.

We write the equation of motion as

$$\ddot{X}(t) = F(X(t)) \tag{1}$$

where

$$F(X(t)) = -\mathbf{M}^{-1}\,\nabla E(X(t))\,. \tag{2}$$

In these equations, $X \in \boldsymbol{R}^{3N}$ denotes the collective Cartesian vector of the system, V is the corresponding collective velocity vector, $\mathbf{M}$ is the diagonal mass matrix, and the dot superscripts denote differentiation with respect to time, t. The total force F is composed of the systematic force, which is the negative gradient of the potential energy E and, possibly, additional terms that mimic the environment (see below for Langevin additions).

One of the simplest and best family of integrators for biomolecular dynamics is the Verlet/Störmer group [4, 5]. The favorable numerical properties of this family — exceptional stability over long times and good conservation of energy — are attributed to its symplecticness.

The original Verlet algorithm [5] describes an update for trajectory positions, as follows:

$$X^{n+1} = 2X^n - X^{n-1} + \Delta t^2 F^n\,. \tag{3}$$

Here superscripts n denote the finite-difference approximations to quantities at timestep $n\Delta t$. The flexibility in defining compatible velocity formulas from the Verlet position propagation formula has led to Verlet variants. These variants rely on a velocity formula consistent with eq. (3).

Three popular variants of Verlet are the *leapfrog* scheme [6], *velocity Verlet* [7], and *position Verlet*. The force is evaluated at the endpoints in leapfrog and velocity Verlet but at the midpoint of position Verlet.

The leapfrog Verlet method is given by:

$$V^{n+1/2} = V^{n-1/2} + \Delta t\, F^n\,, \tag{4}$$

$$X^{n+1} = X^n + \Delta t\, V^{n+1/2}\,. \tag{5}$$

The velocity Verlet scheme is given as:

$$X^{n+1} = X^n + \Delta t\, V^n + \frac{\Delta t^2}{2} F^n\,, \tag{6}$$

$$V^{n+1} = V^n + \frac{\Delta t}{2}\,(F^n + F^{n+1})\,. \tag{7}$$

Both Verlet variants can be written in a symmetric fashion using the following *propagation triplet*:

$$\begin{aligned} V^{n+1/2} &= V^n + \frac{\Delta t}{2} F^n \\ X^{n+1} &= X^n + \Delta t \, V^{n+1/2} \\ V^{n+1} &= V^{n+1/2} + \frac{\Delta t}{2} F^{n+1} \, . \end{aligned} \tag{8}$$

Another variant of Verlet is known as *position Verlet*, written as the triplet

$$\begin{aligned} X^{n+1/2} &= X^n + \frac{\Delta t}{2} V^n \\ V^{n+1} &= V^n + \Delta t \, F^{n+1/2} \\ X^{n+1} &= X^{n+1/2} + \frac{\Delta t}{2} V^{n+1} \, . \end{aligned} \tag{9}$$

3 Multiple-Timestep (MTS) Methods

MTS schemes were introduced over two decades ago [8, 9] to reduce the computational cost of molecular simulations. Savings can be realized if the slow-varying forces due to distant interactions (e.g., electrostatics) are held constant over longer intervals than the more rapidly-varying short-range forces (e.g., bonded forces). Standard integration procedures can then be modified by evaluating the long-range forces less often than the short-range terms. Between updates, the slow forces can be incorporated into the discretization either as constants (via *extrapolation*) or as zero forces (i.e., considered only at the beginning and end of each macrostep).

We consider below a force splitting scheme that involves three components for illustration:

$$F_{\text{fast}}, \; F_{\text{med}}, \text{ and } F_{\text{slow}},$$

with corresponding timesteps

$$\Delta\tau, \; \Delta t_m, \text{ and } \Delta t$$

for updating the fast, medium, and slow forces. These timesteps are related via the ratios

$$k_1 = \Delta t_m / \Delta\tau \, , \qquad k_2 = \Delta t / \Delta t_m \, , \qquad r = k_1 \, k_2 = \Delta t / \Delta\tau \, . \tag{10}$$

For example, the fast forces can be taken as the bonded terms: bond length, bond angle, and dihedral angle components; the medium forces can be defined as other force terms computed within the distance range of around 6 or 7 Å; the slow forces are the remaining terms. In Ewald approaches, the electrostatic interactions are also divided between the medium and slow force classes (direct and reciprocal lattice terms, respectively). The schemes described below can be generalized to more than three classes.

3.1 Extrapolation

A position-Verlet based, extrapolative force-splitting approach can be written as the following doubly-nested loop for covering each Δt sweep.

Extrapolative MTS

$X_r^0 \equiv X + \frac{\Delta t_m}{2} V$
$F_{\text{slow}} \equiv -\mathbf{M}^{-1} \nabla E_{\text{slow}}(X_r)$
For $j = 1$ to k_2
  $X_r \equiv X_r^j \leftarrow X + \frac{\Delta t_m}{2} V$
  $F_{\text{med}} \equiv -\mathbf{M}^{-1} \nabla E_{\text{med}}(X_r)$
  $F \leftarrow F_{\text{med}} + F_{\text{slow}}$
  For $i = 1$ to k_1
    $X \leftarrow X + \frac{\Delta \tau}{2} V$
    $V \leftarrow V + \Delta \tau \, (F + F_{\text{fast}}(X))$
    $X \leftarrow X + \frac{\Delta \tau}{2} V$
  End
End

In each Δt sweep, the *slow force* is evaluated *once*, the *medium force* is evaluated k_2 *times*, and the *fast force* is evaluated $k_1 k_2$ (or r) *times*. If the slow force calculations take the majority of the CPU time, the MTS approach will result in significant computational savings. In practical implementations, the force distance classes are best treated by a smooth force-switching approach.

Simple extrapolation formulations, as above, were the first to be tried for MD [8, 10, 9, 11, 12]. However, they exhibited systematic energy drifts, a result of their nonsymplecticness.

3.2 Impulses

Work continued in the 1980s on MTS methods in a variety of contexts [13, 14], leading to the introduction in 1991 by Schulten and coworkers [11] and independently in 1992 by Berne and coworkers [15] to an MTS method which is symplectic and time-reversible. This MTS variant, based on impulses rather than extrapolation, was termed Verlet-I by the former group [11] and r-Respa [15] by the latter. The more general Respa method was derived by subjecting the Verlet method to the Trotter factorization of the Liouville operator.

A sweep over one Δt interval by an impulse-MTS Verlet approach based on the velocity Verlet method can be written as the following iteration process:

Impulse MTS

$V \leftarrow V + \frac{\Delta t}{2} F_{\text{slow}}(X)$
For $j = 0$ to $k_2 - 1$
  $V \leftarrow V + \frac{\Delta t_m}{2} F_{\text{med}}(X)$
  For $i = 0$ to $k_1 - 1$
    $V \leftarrow V + \frac{\Delta \tau}{2} F_{\text{fast}}(X)$
    $X \leftarrow X + \Delta \tau \, V$
    $V \leftarrow V + \frac{\Delta \tau}{2} F_{\text{fast}}(X)$
  End
  $V \leftarrow V + \frac{\Delta t_m}{2} F_{\text{med}}(X)$
End
$V \leftarrow V + \frac{\Delta t}{2} F_{\text{slow}}(X)$

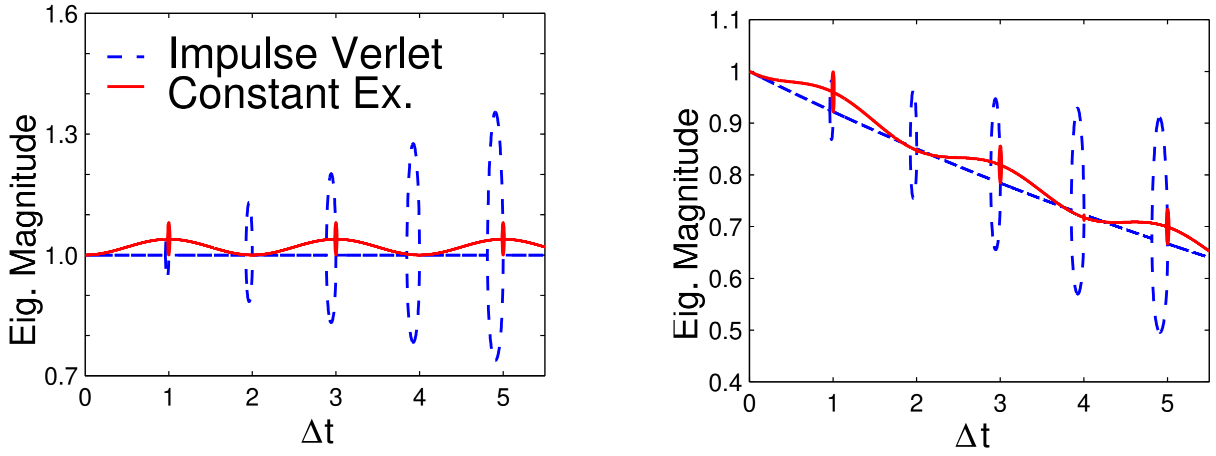

Figure 1: Propagation-matrix eigenvalue magnitudes for Newtonian (left) and Langevin (right, $\gamma = 0.162$) *impulse* and *extrapolative* force-splitting methods as a function of the outer timestep as calculated for a one-dimensional linear test system [20]. The inner timestep is $\Delta\tau = 0.001$ units.

3.3 Impulses and Resonance

The application of the slow force results in an *impulse*. The velocities are modified only outside of the inner loop (i.e., at the onset and at the end of a sweep covering Δt) by a term r times larger than the changes made to X and V in the inner loop. As the time interval between slow-force updates increases, the size of these "impulses" grows. This causes a *resonance artifact* when the impulse frequency (Δt) occurs near a natural frequency of the system.

The original work [11] expressed reservations regarding force impulses. A subsequent work [16] discussed resonance as well the systematic drift of extrapolative treatment in simple oscillator systems, but it conjectured that resonance artifacts were not likely to be so clear in nonlinear systems. Other works noted a rapid energy growth [17, 18] when the interval between slow-force updates approaches and exceeds 5 fs, half the period of the fastest oscillations in biomolecules

Simple linear analyses of resonance in [19, 20] show that impulse methods are *generally stable* except at integer multiples of half the period of the fastest motion, with the severity of the *instability worsening* with the timestep. Extrapolation methods are *generally unstable* for the Newtonian model problem, but the *instability is bounded* for increasing timesteps. This is shown in Figure 1, left [19, 20].

These resonance patterns also generalize to stochastic dynamics, where friction and random terms are added [19, 20]. Furthermore, applications to nonlinear systems, such as solvated proteins, exhibit similar overall patterns for impulse versus extrapolative force splitting treatments (Figure 1, right).

The presence of these resonances limits the outermost timestep, Δt (or the interval of slow-force update), in MTS schemes and thus the overall speedup that can be achieved in practice over a standard Verlet trajectory [15, 17, 18] for biomolecules. Larger computational gains can be achieved by the mollified impulse method of Skeel [21, 22], but this extension requires a Langevin coupling following Barth and Schlick [23] to allow larger timesteps.

Another approach is to use extrapolation in the context of a stochastic formulation, as in the LN method [24, 2, 23, 19] (see below). This combination avoids the systematic energy drift, alleviates severe resonance, and allows much longer outer timesteps [25, 26].

4 Langevin Dynamics and MTS Methods

A stochastic alternative to Newtonian dynamics, namely Langevin dynamics, has been used in a variety of biomolecular simulation contexts for *numerical* (e.g., to counteract damping, mask instabilities, or enhance sampling) and *physical* (e.g., approximate solvent representation) considerations.

The Langevin model is phenomenological [27] — adding friction and random forces to the systematic forces — but with the physical motivation to represent a simple heat bath for the macromolecule by accounting for molecular collisions. The continuous form of the simplest Langevin equation is given by:

$$\mathbf{M}\ddot{X}(t) = -\nabla E(X(t)) - \gamma \mathbf{M}\dot{X}(t) + R(t), \tag{11}$$

where γ is the collision parameter (in reciprocal units of time). The random-force vector R is a stationary Gaussian process with statistical properties given by:

$$\langle R(t) \rangle \quad = \quad 0, \qquad \langle R(t)R(t')^T \rangle \ = \ 2\gamma k_B \mathrm{T}\mathbf{M}\ \delta(t-t'), \tag{12}$$

where k_B is the Boltzmann constant, T is the target temperature, and δ is the usual Dirac symbol.

4.1 The LN Method

LN arose fortuitously [24] from the Langevin/Normal-mode method LIN [28, 29]. The combination of force splitting via extrapolation with Langevin dynamics alleviates severe resonance effects, and allows larger outer timesteps to be used with respect to impulse-based MTS schemes [23, 19, 24].

The exchange of Hamiltonian dynamics for stochastic dynamics can guarantee better numerical behavior, but the resulting detailed dynamics are not the same. Stochastic dynamics is, however, suitable for many thermodynamic and sampling questions. Small-timestep dynamics simulations can always be performed in tandem once an interesting region of conformation space is identified.

LN can be sketched as follows for a three-class force-splitting scheme.

LN Algorithm

$X_r^0 \equiv X$
$F_{\text{slow}} \equiv -\nabla E_{\text{slow}}(X_r)$
For $j = 1$ to k_2
  $X_r \equiv X_r^j \leftarrow X + \frac{\Delta t_m}{2} V$
  $F_{\text{med}} = -\nabla E_{\text{med}}(X_r)$
  $F \leftarrow F_{\text{med}} + F_{\text{slow}}$
  For $i = 1$ to k_1

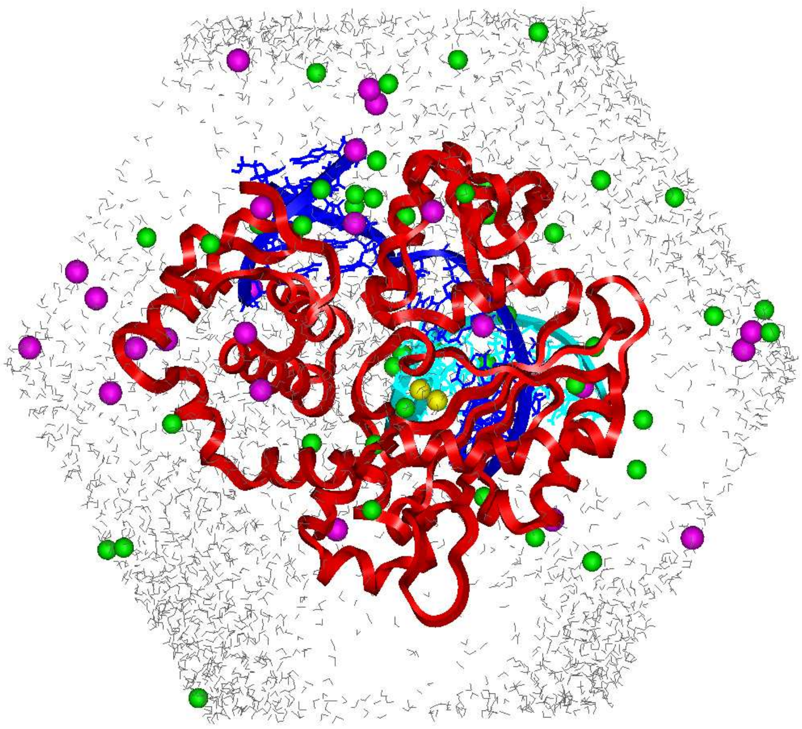

Figure 2: Solvated model of the ternary DNA/Pol β complex in a face-centered cubic domain. The protein atoms are red, the DNA atoms are blue, and the ions are magnesium (yellow), sodium (green), and chloride (purple).

$$
\begin{aligned}
X &\leftarrow X + \frac{\Delta\tau}{2} V \\
V &\leftarrow \Big(V + \mathbf{M}^{-1}\,\Delta\tau (F + F_{\text{fast}}(X) + R)\Big)/(1+\gamma\Delta\tau) \\
X &\leftarrow X + \frac{\Delta\tau}{2} V
\end{aligned}
$$

End

End

The Gaussian random force R is chosen each inner timestep by standard procedures (e.g., Box/Muller method).

4.2 LN Performance

Here we analyze performance of the LN scheme on a large polymerase/DNA system under investigation in our group [30] (Figure 2). DNA polymerase β (Pol β) is a multi-task master enzyme that also repairs single DNA bases following their damage or loss. Pol β fills single-stranded gaps in the DNA by recruiting the correct complementary nucleotide-containing unit precisely as needed [31].

The polymerase mechanism explaining fidelity of the repair process remains a mystery. One popular hypothesis today advocates an 'induced fit' mechanism between the DNA-bound polymerase and the correct substrate: the correct substrate aligns the catalytic groups of the

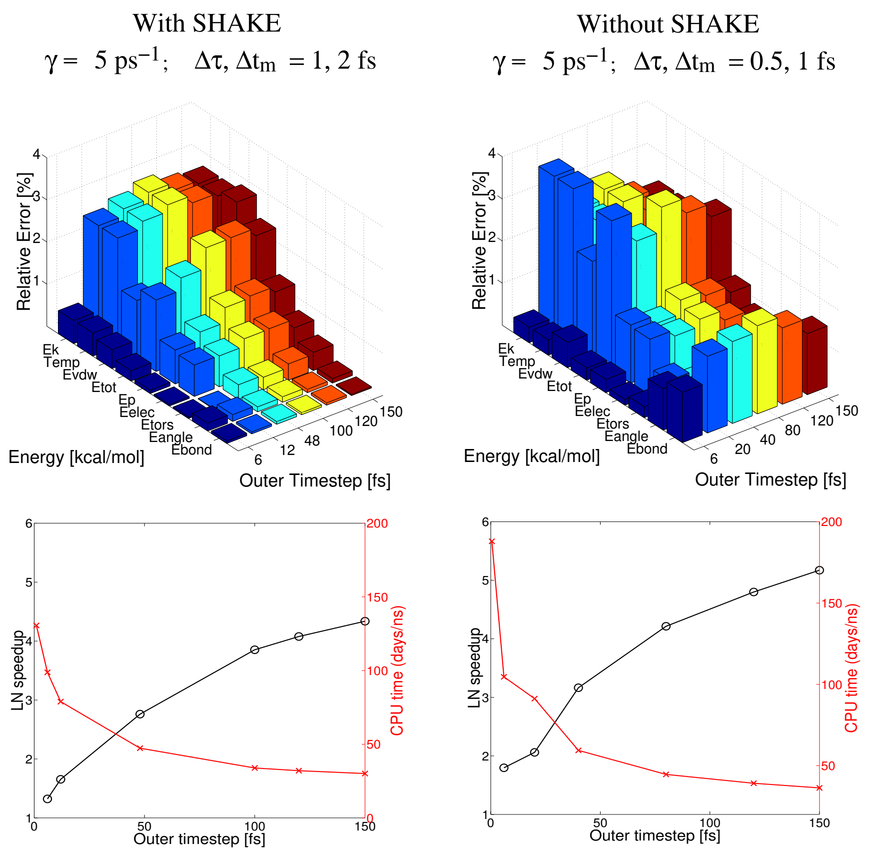

Figure 3: Manhattan plots for the closed structure of the solvated polymerase/DNA system showing the LN relative errors in different energy components as a function of the outer timestep compared to single-timestep Langevin values at $\Delta\tau$, along with corresponding speedups and CPU times. The speedup data (circles) have scale at the left, while the associated CPU times per nanosecond (x symbols) have scales at right. Data were obtained from 6 ps simulations started from the closed structure.

protein as required for proper synthesis, while the incorrect nucleotide units misalign the components so that repair activity is largely repressed.

Offering strong support to this hypothesis is recent evidence that the rate-limiting step involves a conformational change of the polymerase that induces formation of a ternary complex (DNA, enzyme, substrate) [32]. Presumably, the binding of the correct nucleotide unit triggers this conformational change, producing a 'closed', tightly-bound enzyme/product complex. After repair, the DNA/polymerase system returns to its 'open' state, releasing the DNA to allow translocation for the next cycle of polymerization [33].

The research group of Sam Wilson has solved three complexes of human Pol β that are believed to represent general intermediates of this reaction between the closed and open forms [34]: a *binary complex* with DNA containing a nucleotide gap, a *ternary complex* including the correct incoming dinucleotide, and the *product complex* containing only a nicked DNA (i.e., bases repaired but backbone still nicked). We are using MD simulations to sample the pathway between the closed and open conformations and explore how the correct nucleotide unit triggers this conformational movement while the incorrect one does not. The biological

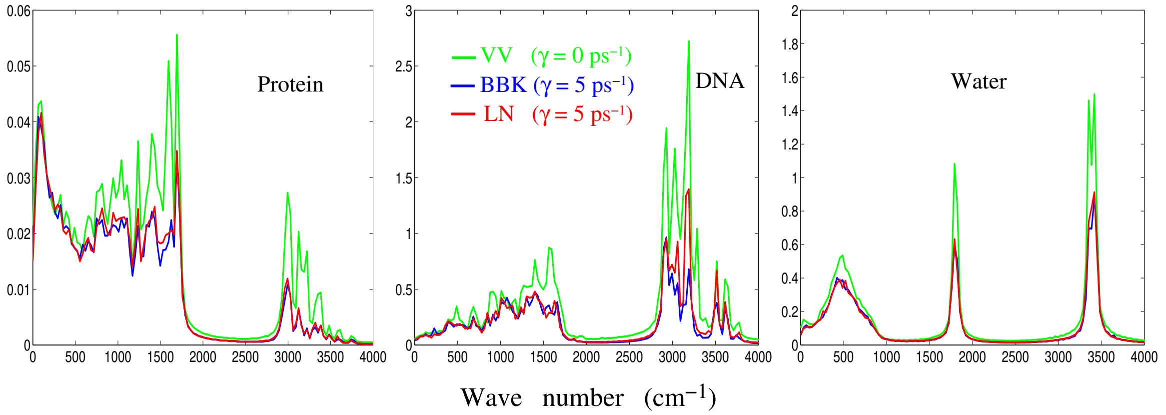

Figure 4: Spectral density functions calculated by Velocity Verlet (VV, $\gamma = 0$) versus single-timestep Langevin (BBK) and multiple-timestep Langevin (LN) at $\gamma = 5$ ps^{-1}, the latter with the SHAKE protocol with 1/2/120 fs timesteps, for a solvated polymerase/DNA system. Data were obtained over 0.9 ps simulations started from the intermediate structure.

findings will be presented separately [30].

Our system size for the closed structure is 43,751 atoms: 322 crystal water molecules, 11,096 bulk waters, 2 each of magnesium and chloride ions from the crystal, and 45 sodium and 23 chloride counterions placed for charge neutralization. Our system size for the intermediate structure is 41,973 atoms including 11,830 bulk waters, 44 sodium and 21 chloride counterions.

Our prior experience with LN has shown that good parameter choices for 3-class schemes are $\Delta\tau = 0.5$ fs, $\Delta t_m = 1$ fs, and Δt of around 120 fs. If constrained dynamics (e.g., SHAKE algorithm) for the light atom bonds is used, the inner timestep can be increased to 1 fs and the medium timestep to around 2 fs. Various biomolecular applications [23, 35, 36, 30] have shown good agreement of LN trajectories with small-timestep Langevin simulations and significant speedup factors.

Figure 3 plots the relative energy differences of LN at different outer timesteps compared to a reference, single-timestep Langevin simulation performed at the inner timestep $\Delta\tau$ for our polymerase system. Two LN protocols at $\gamma = 5$ ps^{-1} (with and without SHAKE for constraining all atom bonds involving hydrogens) are used.

We see that relative energy differences below 4% result for the larger timesteps examined in both protocols. The computational is a factor of 4–5, reducing the CPU time per nanosecond from around 150 to 35 days.

The Langevin parameter γ in the LN scheme is chosen to maintain numerical stability but at the same time minimize the perturbations to Hamiltonian dynamics as much as possible. For biomolecular simulations, we recommend $\gamma = 10$ ps^{-1} or smaller.

To assess the effect of γ of dynamic properties, spectral density functions are computed by Fourier transforming the velocity autocorrelation time series for each atom in the system to obtain a power spectrum for each atom. These spectra are averaged over the water and biomolecule atoms separately for global characterization of the motion. See [20], for example, for the detailed protocol.

The spectral densities computed for the large polymerase system in Figure 4 for $\gamma =$ 5 ps^{-1} show good agreement between the single-timestep Langevin (BBK) and LN-computed frequencies for the same γ. This emphasizes the success of MTS integrators as long as the inner timestep is small. As expected, a damping of Newtonian signals by the stochastic approach is evident, but the signals are well approximated for this γ value.

Detailed comparisons of the evolution of various geometric variables reflect this agreement between the single and multiple-timestep Langevin integrators as well [30] (Figure 5). This figure also shows the divergence of individual trajectories, as expected, but all ranges sampled are reasonable.

5 Conclusions

Developing efficient algorithms for biomolecular simulations will remain a challenging area of research. There are ample opportunities for models on various levels spatial and temporal resolution, from the detailed all-atom and small-timestep trajectories to macroscopic and stochastic approaches [37]. Multidisciplinary interactions, as stimulated by the Eilat workshops, will continue to be vital to success in such endeavors.

6 Acknowledgments

The workshops would not have been a success without the dedication of the local team — Achi Brandt, Carol Weintraub, and Raanan Michael.

Research support from the National Science Foundation, the National Institutes of Health, and the John Simon Guggenheim Foundation is gratefully acknowledged. Computations on the NCSA SGI Origin2000 computer system were made possible by support from the National Computational Science Alliance under award MCA99S021N. T. S. is an investigator of the Howard Hughes Medical Institute.

References

[1] T. Schlick, R. D. Skeel, A. T. Brünger, L. V. Kalé, J. A. Board, Jr., J. Hermans, and K. Schulten. Algorithmic challenges in computational molecular biophysics. *J. Comp. Phys.*, 151:9–48, May 1999. (Special Volume on Computational Biophysics).

[2] T. Schlick, E. Barth, and M. Mandziuk. Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation. *Ann. Rev. Biophys. Biomol. Struc.*, 26:179–220, 1997.

[3] T. Schlick. Computational molecular biophysics today: A confluence of methodological advances and complex biomolecular applications. *J. Comp. Phys.*, 151:1–8, May 1999. (Special Volume on Computational Biophysics).

[4] C. Störmer. Sur les trajectoires des corpuscules électrisés dans l'espace. *Archives des Sciences Physiques et Naturelles*, 24:5–18, 113–158, 221–247, 1907. (This reference is the first in a three-part essay. The second part appeared in the same journal in 1911 (pages 190, 277, 415, and 501), and the third part appeared in the 1912 volume of the journal, pages 51–69).

[5] L. Verlet. Computer 'experiments' on classical fluids: I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98–103, July 1967.

[6] R. W. Hockney and J. W Eastwood. *Computer Simulation Using Particles.* McGraw-Hill, New York, NY, 1981.

[7] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Applications to small water clusters. *J. Chem. Phys.*, 76:637–649, 1982.

[8] W. B. Streett, D. J. Tildesley, and G. Saville. Multiple time step methods in molecular dynamics. *Mol. Phys.*, 35:639–648, 1978.

[9] R. D. Swindoll and J. M. Haile. A multiple time-step method for molecular dynamics simulations of fluids of chain molecules. *J. Chem. Phys.*, 53:289–298, 1984.

[10] W. B. Streett, D. J. Tildesley, and G. Saville. Multiple time step methods and an improved potential function for molecular dynamics simulations of molecular liquids. In Peter Lykos, editor, *Computer Modeling of Matter*, volume 86 of *ACS Symposium Series*, pages 144–158. ACS, Washington, D. C., 1978.

[11] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Sim.*, 6:121–142, 1991.

[12] M. E. Tuckerman, B.J. Berne, and G. J. Martyna. Molecular dynamics algorithm for multiple time scales: Systems with long range forces. *J. Chem. Phys.*, 94:6811–6815, 1991.

[13] M. E. Tuckerman, B. J. Berne, and A. Rossi. Molecular dynamics algorithm for multiple time scales: Systems with disparate masses. *J. Chem. Phys.*, 94:1465–1469, 1991.

[14] M. E. Tuckerman and B. J. Berne. Molecular dynamics in systems with multiple time scales: Systems with stiff and soft degrees of freedom and with short and long range forces. *J. Comp. Chem.*, 95:8362–8364, 1992.

[15] M. E. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97:1990–2001, 1992.

[16] J. J. Biesiadecki and R. D. Skeel. Dangers of multiple-time-step methods. *J. Comp. Phys.*, 109:318–328, 1993.

[17] R. Zhou and B. J. Berne. 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.*, 103:9444–9459, 1995.

[18] M. Watanabe and M. Karplus. Simulations of macromolecules by multiple time-step methods. *J. Phys. Chem.*, 99:5680–5697, 1995.

[19] E. Barth and T. Schlick. Extrapolation versus impulse in multiple-timestepping schemes: II. linear analysis and applications to Newtonian and Langevin dynamics. *J. Chem. Phys.*, 109:1632–1642, 1998.

[20] A. Sandu and T. Schlick. Masking resonance artifacts in force splitting methods for biomolecular simulations by extrapolative langevin dynamics. *J. Comp. Phys.*, 151:74–113, May 1999. (Special Volume on Computational Biophysics).

[21] B. García-Archilla, J.M. Sanz-Serna, and R.D. Skeel. Long-time-step methods for oscillatory differential equations. *SIAM J. Sci. Comp.*, 20:930–963, 1998.

[22] J. A. Izaguirre. *Longer Time Steps for Molecular Dynamics.* PhD thesis, University of Illinois at Urbana-Champaign, 1999. Also UIUC Technical Report UIUCDCS-R-99-2107. Available via http://www.cs.uiuc.edu/research/tech-reports.html.

[23] E. Barth and T. Schlick. Overcoming stability limitations in biomolecular dynamics: I. combining force splitting via extrapolation with Langevin dynamics in LN. *J. Chem. Phys.*, 109:1617–1632, 1998.

[24] E. Barth, M. Mandziuk, and T. Schlick. A separating framework for increasing the timestep in molecular dynamics. In W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, editors, *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, volume III, chapter 4, pages 97–121. ESCOM, Leiden, The Netherlands, 1997.

[25] P. Koehl and M. Levitt. Theory and simulation: Can theory challenge experiment? *Curr. Opin. Struct. Biol.*, 9:155–156, 1999.

[26] S. Doniach and P. Eastman. Protein dynamics simulations from nanoseconds to microseconds. *Curr. Opin. Struct. Biol.*, 9:157–163, 1999.

[27] D. A. McQuarrie. *Statistical Mechanics.* Harper & Row, New York, NY, 1976. Chapters 20–21.

[28] G. Zhang and T. Schlick. LIN: A new algorithm combining implicit integration and normal mode techniques for molecular dynamics. *J. Comp. Chem.*, 14:1212–1233, 1993.

[29] G. Zhang and T. Schlick. The Langevin/implicit-Euler/Normal-Mode scheme (LIN) for molecular dynamics at large time steps. *J. Chem. Phys.*, 101:4995–5012, 1994.

[30] L. Yang, S. Broyde, and T. Schlick. An exploration by molecular dynamics of the repair mechanism of human polymerase β. In Prepration, 2000.

[31] A. Kumar, J. Abbotts, E. M. Karawya, and S. H. Wilson. Identification and properties of the catalytic domain of mammalian DNA polymerase β. *Biochemistry*, 29:7156–7159, 1990.

[32] Y. Li, S. Korolev, and G. Waksman. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of thermus aquaticus DNA polymerase I: Structural basis for nucleotide incorporation. *EMBO J.*, 17:7514–7525, 1998.

[33] S. S. Carroll and S. J. Benkovic. Mechanistic aspects of DNA polymerases: *Escherichia Coli* DNA polymerase I (klenow fragment) as a paradigm. *Chem. Rev.*, 90:1291–1307, 1990.

[34] M. R. Sawaya, R. Prasad, S. H. Wilson, J. Kraut, and H. Pelletier. Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: Evidence for induced fit mechanism. *Biochemistry*, 36:11205–11215, 1997.

[35] D. Strahs and T. Schlick. A-tract bending: Insights into experimental structures by computational models. *J. Mol. Biol.*, 0, 2000. In Press.

[36] X. Qian, D. Strahs, and T. Schlick. Simulations of 13 TATA-box variants reveal the evolutionary selection of DNA elements by TBP according to protein-complementary structure and flexibility properties. In Preparation, 2000.

[37] T. Schlick, D. Beard, J. Huang, D. Strahs, and X. Qian. Computational challenges in simulating large DNA over long times. *IEEE Comp. Sci. Eng.*, 0, 2000. (Special Issue on Computational Chemistry).

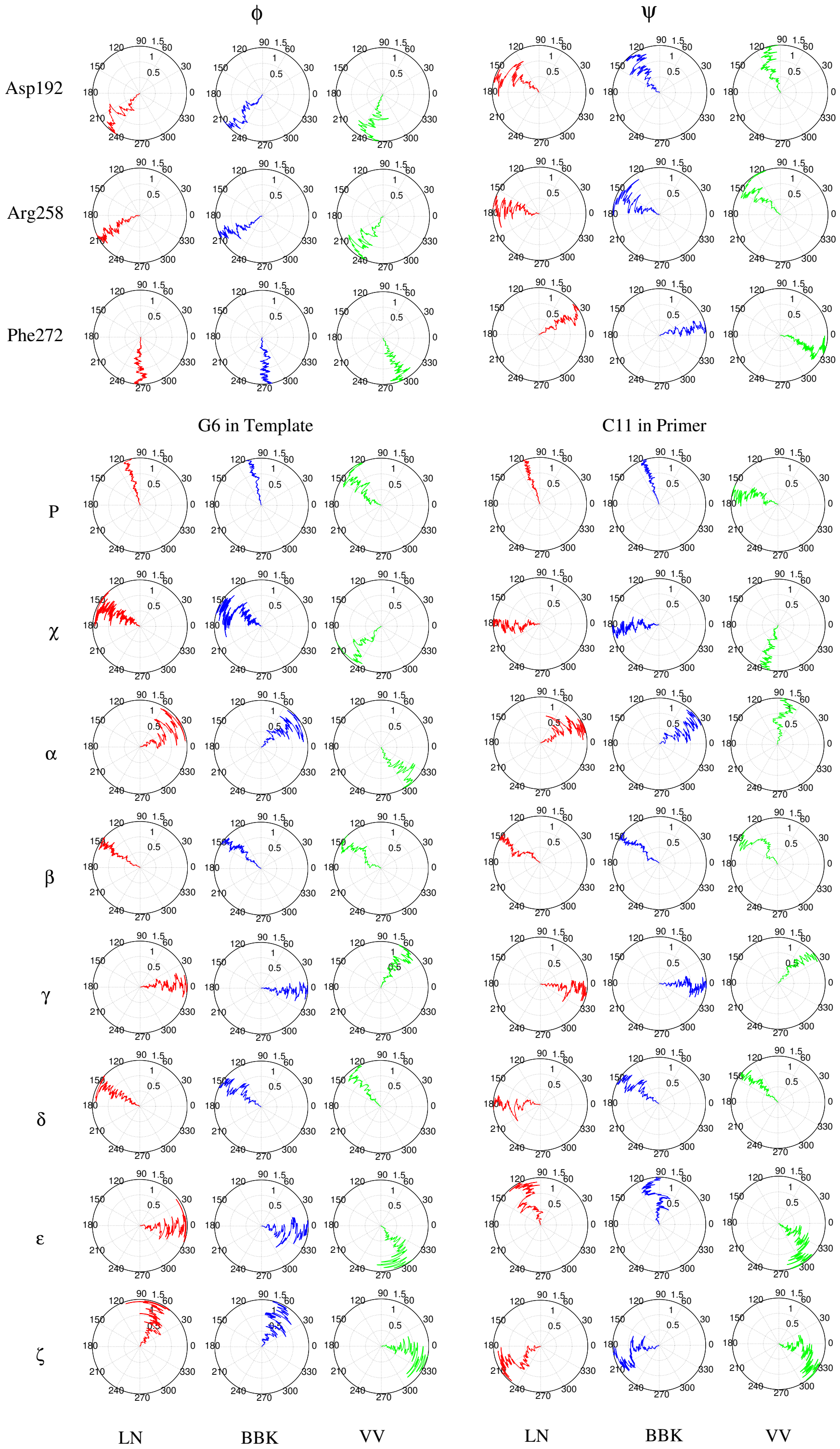

Figure 5: Comparisons of the evolution of several protein and DNA dihedral angles by the single-timestep Verlet and Langevin schemes (VV, BBK) vs. the MTS LN scheme for the solvated polymerase/DNA system. Data were obtained from 1.5 ps simulations started from the intermediate structure.