Adaptive dimensionality reduction of stochastic differential equations for protein dynamics

Jesús A. Izaguirre*, Christopher R. Sweet*
*University of Notre Dame, Notre Dame, IN, USA

***Abstract*— The dynamics of proteins can be described as the superposition of motions at a continuum of time scales. In the special case of a protein immersed in an implicit solvent, a stochastic differential equation (SDE) can model the dynamics of the solute protein. Traditional model reduction techniques fail because *a priori* characterization of the slow variables in these SDEs is nearly impossible. We present an approach that instead, does a local dimensionality reduction of the SDE in a neighborhood of phase space, which is adaptively performed when the reduced model is no longer valid. The local slow variables, which we call approximate normal modes (ANM), are found using the diagonalization of a coarse-grained Hessian (CGH) from the potential energy function. We call this procedure coarse-grained normal mode analysis, or CNMA. Diagonalization of the CGH can be achieved in $O(N \log N)$ time and $O(N)$ memory rather than $O(N^3)$ time and $O(N^2)$ memory of ordinary diagonalization. CNMA is able to capture the low frequency motions of the protein. An SDE on the ANM is found by using a *saddle-point approximation* of the mean fast-frequency force experienced by the slow variables, and an implicit solvent model that considers the protein as a Brownian particle. This mean force can be computed at a cost no greater than a fine-grained force evaluation. Discretization of the resulting SDE achieves very long time steps compared to the discretization of the fine-grained SDE. A metric is used to monitor the validity of the ANM as slow variables and prompt re-diagonalization of the CGH or adaptation of the time step used. I will present numerical results on small peptide and protein system that show that this coarse-graining scheme allows up to three orders of magnitude speedup due to increase in the SDE discretization time step, and that the scheme is able to preserve kinetics when compared to the fine-grained SDE.**

I. Introduction

Proteins are polymers of mostly naturally occurring aminoacids. Proteins are molecular machines, and as any machine, they must move in order to function. Understanding protein motion or dynamics is critical to solving problems as diverse as protein folding, functional conformational changes, and to computationally predict the effectiveness of drugs that target proteins. Simulating protein dynamics remains very challenging. The most straightforward approach, molecular simulation of Newton's equations of motion using standard atomistic models, quickly runs into a significant sampling problem for all but the most elementary of systems. While small proteins fold or have biologically relevant conformational changes on the microseconds to second timescale, detailed atomistic simulations are currently limited to the nanosecond regime, with a few "heroic" simulations breaking the microsecond timescale.

The fundamental challenge to overcome is the presence of multiple time scales: typical bond vibrations are on the order of femtoseconds (10^{-15} sec) while proteins fold on a time-scale of microsecond to millisecond. This is a 10^{12} difference in time scales! We tackle the problem of developing timescale-coarse-grained models of proteins to understand their thermodynamics (e.g. statistical properties) and kinetics (e.g. rates). Among other multiscale problems, coarse graining dynamical protein models in a rigorous and computationally tractable way is particularly challenging. Two specific difficulties can be identified.

The *identification of the slowest variables* in the system (e.g. associated with the slowest time scales and transition rates) is to a large extent an unresolved problem. Even when people agree on a specific definition, the actual computation can be intractable. This is the case for example if one attempts to calculate the committor function, the probability at a given point that the protein folds rather than unfolds, by brute force. This is typically done by starting many trajectories from a given point and directly computing how many fold the protein. Additionally, computing dynamics of the slow variables is non-trivial because they are intricately coupled to the fast variables. In other words, there is in general *no timescale separation*. In protein modeling, there is a dense spectrum of frequencies due to the highly coupled nature of the force field. Unfortunately, most multiscale methods start by assuming that it is in fact possible to extract variables whose time scale is significantly slower than the rest. Some of the unresolved variables will have time scale (faster but) comparable to time scales of some of the resolved variables. No sharp cutoff can be found. Therefore special techniques need to be developed to go beyond the time scale separation approximation.

II. Coarse-grained Normal Mode Langevin dynamics (CNML)

Rather than attempting to identify slow variables that are valid globally, our approach is based on adaptively identifying slow variables valid locally. Once these slow variables have been identified, we derive a SDE where the effect of fast variables is described through average and fluctuating forces. We discretize this SDE, which allows much longer time steps than the original fine-grained equations of motion. Whenever these slow variables are no longer valid, defined by a metric explained below, we identify a new set of slow variables, or alternatively adapt the time step for the solution of the SDE. Earlier,

we presented an approach using similar slow variables and SDE, but that attempted no adaptive dimensionality reduction in [7]. Our current approach is more robust and scalable.

Choice of slow variables. Our slow variables are approximate low-frequency normal modes (ANM) derived from a coarse-grained Hessian (CGH). Normal modes are the eigenvectors of the Hessian matrix of the potential energy U at an equilibrium or minimum point x_0 with proper mass normalization. More formally assume a system of N atoms with $3N$ Cartesian positions and diagonal mass matrix $\mathbf{M}$. To perform the normal mode analysis (NMA) for these systems we can formally expand the potential energy about an equilibrium point, which we assume to be a local minimum. The Hessian $\mathbf{H}$ is a factor in the first non-constant, non-zero term of this expansion and a harmonic approximation to the original system can be found by truncating the expansion at this point. To rewrite the harmonic approximation as a set of decoupled oscillators it is insufficient to diagonalize the Hessian as the resulting oscillators would be coupled through the projected mass matrix. Instead we mass re-weight the system using and then diagonalize the resulting mass re-weighted Hessian, $\mathbf{M}^{-\frac{1}{2}}\mathbf{H}\mathbf{M}^{-\frac{1}{2}}\mathcal{Q} = \mathcal{Q}\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is the diagonal matrix of ordered eigenvalues and $\mathcal{Q}$ the matrix of column eigenvectors $\mathbf{e}_1, \cdots, \mathbf{e}_{3N}$. The frequency of a mode is equal to $\sqrt{|\lambda|}$ where λ is the eigenvalue. If we choose a cut-off frequency $\sqrt{\lambda_i}$ to partition the normal modes such that $\mathcal{Q} = [\mathbf{Q}, \bar{\mathbf{Q}}], \mathbf{Q} = [\mathbf{e}_1, \cdots, \mathbf{e}_i]$ and $\bar{\mathbf{Q}} = [\mathbf{e}_{i+1}, \cdots, \mathbf{e}_{3N}]$ are rectangular matrices whose columns span a slow subspace C and fast subspace $C^{\perp}$ respectively. In the following discussion we will assume the dimensions of $\mathbf{Q}$ to be $3N \times m$. In the linear case the time-step is bounded by the asymptotic stability of the method [2] at a frequency equal to the $\sqrt{|\lambda_i|}$, rather than the highest frequency in the system. Our results show this is a good heuristic to choose the time-step.

Dynamics of the slow variables. Once the slow variables have been identified, equations for the rate of change of these variables need to be formulated. Let us denote by q a set of resolved variables with momenta p. We assume that the number of coordinates q is very small compared to $3N$ where N is the number of atoms in the system. Typically N can be as large as hundreds of thousands while the number of resolved variables is 10 to 100. We wish to find a way to calculate $\mathrm{d}p/\mathrm{d}t$ in terms of q and p only. The following exact equation for $\mathrm{d}p/\mathrm{d}t$ can be derived:

$$\frac{\mathrm{d}q(t)}{\mathrm{d}t} = p,$$

$$\frac{p(t)}{t} = -\overbrace{\nabla_q A}^{\textbf{drift}} - \overbrace{\int_0^t C_r(s) \cdot p(t-s)\,\mathrm{d}s}^{\textbf{friction}} + \overbrace{r(t)}^{\textbf{noise}}, \quad (1)$$

$$C_r(s) = \left\langle r(\tau+s)\, r(\tau)^T \right\rangle, \forall \tau \quad (2)$$

(fluctuation dissipation theorem)

These equations are in reduced units and we neglected the dependence of the memory kernel C_r on q and p. The brackets $\langle\rangle$ define the thermodynamic average in the canonical ensemble. Equation (1) can be derived using the Mori-Zwanzig projection [8]. The potential $A(q)$ is the Potential of Mean Force (PMF, or Helmholtz free energy) for variable q. The integral in (1) represents a friction. In this model the friction includes memory so this equation is often called the Generalized Langevin Equation (GLE). The last term $r(t)$ is a fluctuating force with zero mean: $\left\langle r(t)|q_0, p_0\right\rangle = 0$. This is a conditional average over Cartesian coordinates x and momenta p_x keeping $q = q_0$ and $p = p_0$ fixed.

This equation can be rigorously derived from statistical mechanics and is therefore an attractive starting point to build coarse grained models. However, it is also, in principle, very expensive to calculate. The most common approximation is to assume a separation of time scales; then $C_r(s)$ is simply equal to the auto-correlation of $\mathrm{d}p(t)/\mathrm{d}t$ which can be readily computed. As was pointed out earlier this assumption does not hold in general. We next discuss how our choice of slow variables and saddle point approximation of the drift term make this approximation feasible and results in a computationally tractable coarse-grained dynamical model.

Choosing low frequency modes as resolved variables was motivated by the physical insight that low frequency modes contain the physically relevant motions close to the minimum [3], [4]. For small numers of modes we observe that the coupling between C and $C^{\perp}$ is small, though not zero. The drift term of (1) can be simplified using a *saddle-point approximation*. The mean force is approximated by the instantaneous projection of the force onto the slow subspace, subject to the constraint that the conformation is a minimum in the fast subspace. This is the most probable value of the mean force when the spaces are decoupled. At that point the friction term can be approximated as the autocorrelation of the slow force, and the noise can be evaluated as white noise. Thus, we numerically enforce a quasi-adiabatic decoupling between C and $C^{\perp}$.

The simplification of the drift term proceeds as follows. The choice of frequency partition separates the positions around the equilibrium point x_0 into $\hat{x}$ in C and $\bar{x}$ in $C^{\perp}$ such that $x = \hat{x} + \bar{x} + x_0$. These are defined as

$$\hat{x} = \mathbf{P}_{\mathrm{x}}(x - x_0),\ \bar{x} = \mathbf{P}_{\mathrm{x}}^{\perp}(x - x_0),$$

where the projection matrices take the positions from Cartesian to mode coordinates and back to Cartesian space, and are given by

$$\mathbf{P}_{\mathrm{x}} = \mathbf{M}^{-1/2}\mathbf{Q}\mathbf{Q}^{\mathbf{T}}\mathbf{M}^{1/2},$$
$$\mathbf{P}_{\mathrm{x}}^{\perp} = \mathbf{M}^{-1/2}\left(\mathbf{I} - \mathbf{Q}\mathbf{Q}^{\mathbf{T}}\right)\mathbf{M}^{1/2}.$$

The mean force for the drift term for a particular value of can be written as

$$\text{Average } f(\hat{x}) = -\frac{1}{Z}\int \exp(-\beta U(x))\boldsymbol{\delta}$$
$$(\boldsymbol{P}_{\mathrm{x}}(x - x_0) - \hat{x})\boldsymbol{P}_{\mathrm{f}}\nabla U(x)\mathrm{d}x. \quad (3)$$

We have introduced the usual canonical ensemble partition function and the force projection matrix:

$$\boldsymbol{P}_{\mathbf{f}} = \mathbf{M}^{1/2}\mathbf{Q}\mathbf{Q}^{\mathbf{T}}\mathbf{M}^{-1/2}.$$

The average $f(\hat{x})$ is dominated by the slow force term, $-\boldsymbol{P}_{\mathrm{f}}\nabla U(x)$, corresponding to the smallest $U(x)$ due to the Boltzmann weight. This $U(x_{\min})$ is the minimum potential energy that satisfies the constraint $\boldsymbol{P}_x(x_{\min} - x_0) = \hat{x}$. We can rewrite it as $U(x_{\min}) = U(x_0 + \hat{x} + \bar{x}_{\min})$. Since $x_0 + \hat{x}$ is fixed, this is equivalent to minimizing the projection of the positions onto the fast subspace. Hence $\bar{x}_{\min} = \operatorname{argmin} U(x_0 + \hat{x} + \bar{x}_{\min})$ with $x_0, \hat{x}$ fixed. This implies that the mean force can be approximated by the instantaneous slow force:

$$\text{Average } f(\hat{x}) \approx -\boldsymbol{P}_{\mathrm{f}}\nabla U(x_{\min}). \tag{4}$$

A second important approximation is that the protein is modeled using implicit solvent (ISM). ISMs have been shown to be sufficiently accurate for a number of applications, including protein folding studies, and they are attractive because they greatly reduce the cost of simulating a protein. Thus, to model the coarse-grained dynamics of an implicitly solvated protein, (1) is simplified into a Langevin equation:

$$\begin{aligned} \mathrm{d}\mathbf{x} = \mathbf{v}\mathrm{d}t,\ \mathbf{M}\mathrm{d}\mathbf{v} = \mathbf{f}\mathrm{d}t - \boldsymbol{\Gamma}\mathbf{M}\mathbf{v}\mathrm{d}t \\ + (2k_{\mathrm{B}}T\boldsymbol{\Gamma})^{1/2}\,\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}(t), \end{aligned} \tag{5}$$

where $\mathbf{f} = -\boldsymbol{P}_{\mathrm{f}}\nabla U(x_{\min})$ is derived in (4), t is time, $\mathbf{W}(t)$ is a collection of Wiener processes, k_{B} is the Boltzmann constant, T is the system temperature, $\mathbf{v}$ are the velocities and $\boldsymbol{\Gamma}$ is the diagonalizable damping matrix. The system diffusion tensor $\mathbf{D}$ gives rise to $\boldsymbol{\Gamma} = k_{\mathrm{B}}T\mathbf{D}^{-1}\mathbf{M}^{-1}$. $\mathbf{D}$ is chosen to model the dynamics of an implicit solvent and the coarse-graining of the dynamics.

Discretization of the dynamics. We discretize (5) using the Langevin Impulse (LI) integrator [6]. We call this discretization the Normal Mode Langevin (NML) propagator. Schematically, half a step of NML performs the following steps:

Half slow kick: advance velocities using half a long timestep $\Delta t/2$ using the projection of internal and random force unto slow subspace C.

Slow Fluctuation: advance positions using the projection of internal and random force unto slow subspace C.

Fast Fluctuation: minimize positions on fast subspace $C^{\perp}$ using steepest descent.

Coarse grained diagonalization. To adaptively find the slow variables we need a cheap procedure to extract them. We introduce a *coarse-grained normal mode analysis that is scalable* (CNMA). CNMA uses a dimensionality reduction strategy that allows computation of low frequency modes in $O(N \log N)$ time, and with $O(N)$ memory, rather than $O(N^3)$ time and $O(N^2)$ memory of brute-force diagonalization. The coarse-graining strategy to computing the frequency partitioning is based on three ideas. The first is to find a reduced set of normalized vectors $\mathbf{E}$ that spans the low frequency space of interest, C. The second is to recursively extract a minimal set of vectors $\mathbf{Q}$ from $\mathbf{E}$ and the coarse grained Hessian $\mathbf{H}$. The third is form $\mathbf{H}$ in linear cost, $O(N)$.

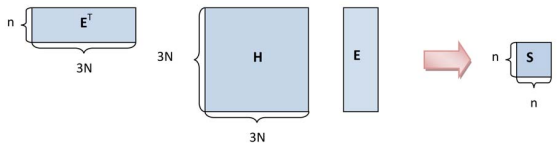


Fig. 1. Illustration of the dimensionality reduction strategy for the diagonalization. If the vectors in $\mathbf{E}$ span the low frequency space of interest in $\mathbf{H}$, then the diagonalization of $\mathbf{S}$ can produce a low frequency basis set.

Assume that we have found $\mathbf{E}$. Figure 1 illustrates the dimensionality reduction strategy. $\mathbf{H}$ is the Hessian at a given simulation step. The dimensions of $\mathbf{E}$ are $3N \times n$, where $n \ll N$. The quadratic product $\mathbf{E}^{\mathrm{T}}\mathbf{H}\mathbf{E}$ produces a matrix $\mathbf{S}$ of reduced dimensions $n \times n$. Below we show that from the diagonalization of $\mathbf{S}$ we can obtain $\mathbf{E}$. In particular, we (cheaply) diagonalize the symmetric matrix $\mathbf{S}$ to find orthonormal matrix $\tilde{\mathbf{Q}}$ s.t.

$$\mathbf{S}\tilde{\mathbf{Q}} = \tilde{\mathbf{Q}}\boldsymbol{\Omega},$$

for diagonal matrix $\boldsymbol{\Omega}$. We can then write

$$\mathbf{Q}^{\mathrm{T}}\mathbf{H}\mathbf{Q} = \boldsymbol{\Omega},$$

for $\mathbf{Q} = \mathbf{E}\tilde{\mathbf{Q}}$. Our subspace of dynamical interest, C, is then defined as the span of the first m columns of $\mathbf{Q}$. Recall that m is the number of reduced collective motions, typically in the range of 1 - 100.

We can evaluate how well the span of $\mathbf{E}$ represents C using the following result (we skip the proof for space limitation): Let the ith ordered diagonal of $\boldsymbol{\Omega}$ be $\sigma_i = \boldsymbol{\Omega}_{ii}$. Then the highest frequency mode in C, $f_{\max}$, satisfies

$$f_{\max} \leq \sqrt{|\sigma_m|}.$$

Then the Rayleigh quotient σ_m can be used to establish the maximum time step that can be taken in subspace C for stability. It follows that if λ_m is close to the mth ordered eigenvalue of $\mathbf{H}$, then the first m vectors of $\mathbf{Q}$ are a good representation of the low frequency space of interest.

We form $\mathbf{E}$, by starting from a 'local' block Hessian in which each block $\tilde{\boldsymbol{H}}_{ij}$ (composed of 1 or more residues) is zero if $i \neq j$. The remaining blocks on the diagonal are assumed to be independent of all other blocks. This block Hessian is then diagonalized, which is equivalent to performing independent diagonalization for each block. Let us determine each block Hessian eigenvectors and eigenvalues, $\mathcal{Q}_i$ and $\mathbf{D}_i$, as follows:

$$\tilde{\boldsymbol{H}}_{ii}\mathcal{Q}_i = \mathcal{Q}_i\mathbf{D}_i.$$

Our hypothesis is that interactions among residues responsible for the low frequency space of interest will be included, either by projection or directly, in the first

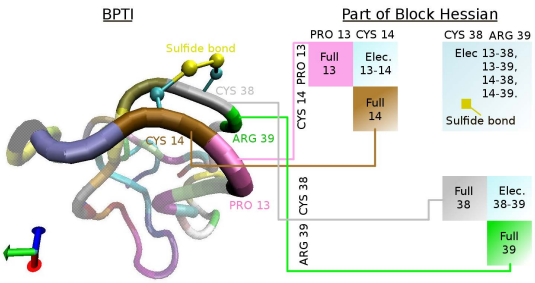


Fig. 2. Segment of a BPTI molecule and its associated block Hessian entries. Here, for illustration, a block is defined by one residue. Each residue corresponds to a Hessian block containing all of the forces within the residue, denoted 'Full'. Adjacent residues have a corresponding electrostatic block denoted 'Elec.', e.g. Elec. 13-14. Physically local residues within the cutoff distance have a corresponding electrostatic block, e.g. Elec. 13-38. Bonds connecting non-adjacent residues, such as the disulfide bond shown, correspond to small 3x3 blocks in the Hessian.

few eigenvectors of $\mathcal{Q}_i$, and need to be included in $\mathbf{E}$. The source of these vectors is as follows:

1) External low frequency motions due to nonbonded interactions can be projected onto the first 6 eigenvectors of $\mathcal{Q}_i$, corresponding to conserved d.o.f. per block. In other words, external forces manifest themselves in rotations or translations of each residue-block.
2) External low frequency motions due to bonded interactions can be projected onto the dihedral space, and will consist of up to 4 vectors of $\mathcal{Q}_i$, due to the peptide bond dihedrals of up to 2 connecting blocks.
3) Internal low frequency motions, for instance due to side-chain dihedral motions, will also be in the dihedral space and thus will be in $\mathcal{Q}_i$.

We expect that the eigenvectors identified above will correspond to the first k ordered eigenvalues. The number k will vary between blocks and will be determined by selecting a cutoff frequency from the block eigenvalues. Figure 2 illustrates the block structure of $\mathbf{H}$ for protein BPTI with cutoff for the electrostatics. This is very similar to a protein contact map. Contiguous residues give a tri-diagonal block structure. Non-contiguous residues that are nearby form off-diagonal blocks due to nonbonded forces. Special structural features like disulfide bonds create 3×3 small blocks. The block structure of $\mathbf{E}$ follows from its composition from eigenvectors of the block Hessians $\tilde{\boldsymbol{H}}_{\boldsymbol{ii}}$. Thus, the cost of the matrix-matrix multiplication will be $O(N)$.

The multilevel application of this dimensionality reduction leads to a scheme with $O(N \log N)$ cost. We first diagonalize each residue. The cost for this stage is $O(N)$ as the average number of residue atoms is fixed and the number of residues is proportional to N. We then need to consider the diagonalization cost of the 'block projected' matrices. If we took a large system and recursively assigned block size 'factor' b, each linear block dimension is b times the previous, then we get a diagonalizations with $b^{a+1} = N$, so total cost is $O(N \log N)$. This leaves the projection of the actual Hessian, but we can assign $b \propto \sqrt[3]{N \log N}$ to yield the correct scaling.

III. Numerical results

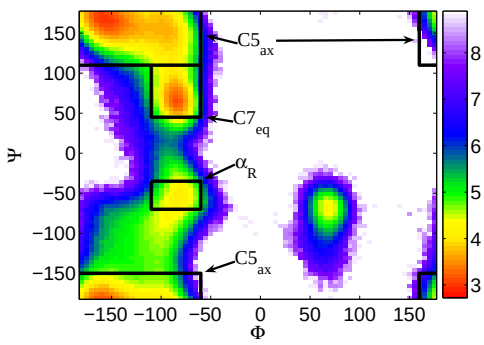
Adaptive NML recovers long time dynamics. We applied NML to study the isomerization kinetics of blocked alanine dipeptide (ACE ALA NME). With a small molecule like alanine dipeptide it is possible to sample for a sufficient length of time to measure the rates of transition between two states: in this case we measure the isomerization rate between the $C7$ equatorial and α_R conformations. The rate from states A and B, denoted k_{AB}, can be calculated using the approximation proposed by Best and Hummer [1] from the probability of transition, P_{TP}, and the average transition time $\langle t_{TP} \rangle$:

$$2c_A k_{AB} = P_{TP}/\langle t_{TP} \rangle,$$

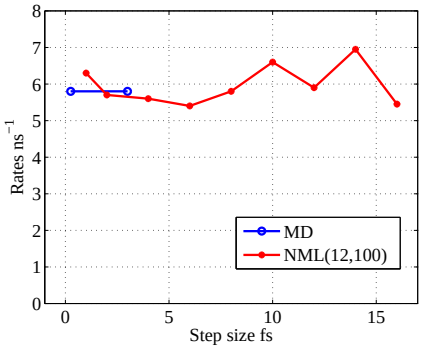
where c_A is the equilibrium mole fraction of conformation A. Figure 3(a) shows the free energy as a Ramachadran plot for Alanine Dipeptide using the sigmoidal screened Coulomb potential of [5]. Conformation A is C7 equatorial and C5 axial combined, and conformation B is α_R. Figure 3(b) shows that NML is capable of correctly computing the rate with only 12 modes with a rediagonalization frequency of 100 fs as the time step increases up to 16 fs. As a reference, the rate computed for the fine-grained SDE using molecular dynamics (MD) with time steps up to 3 fs are shown. MD cannot go beyond this time step due to the fast frequencies present in the system. Let NML(m,*freq*) be NML where m is the number of slow modes propagated, and *freq* refers to the re-diagonalization frequency in femtoseconds. Figure 3(c) shows the isomerization rate for AD running NML(m,0) (no re-diagonalization), NML(m,100) and NML(m,1000). It can be observed that whereas the rate quickly goes down for NML(m,0), the rate is correctly computed for NML(m,100) for even 7 modes (only 1 real mode excluding the 6 conserved modes). NML(m,1000) is somewhere in between the two results.

Coarse-grained normal mode analysis is scalable. Five models were used for the comparison of the 'brute force' diagonalization and the coarse grained CNMA method: PIN1 WW domain (PDB 1I6C), BPTI (PDB 4PTI), Calmodulin (PDB 1CLL), Tyrosine kinase (PDB 1QCF), and F1-ATPase (PDB 2HLD). The results can be seen in Figures 4(a) and 4(b), which match the scaling analysis of $O(N \log N)$ time and $O(N)$ memory.

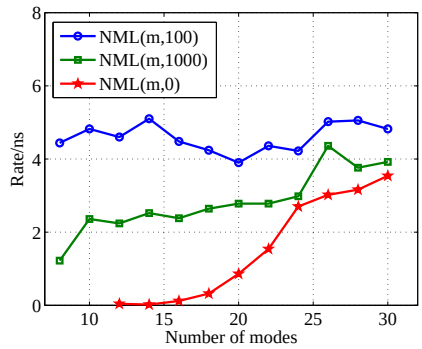
NML with re-diagonalization using CNMA can greatly accelerate dynamics calculations. We are currently applying NML with re-diagonalization using CNMA to study the folding of the WW domain and other proteins. Figure 5 illustrates analytical predictions of the accelerations in sampling the dynamics that we expect when using our approach on progressively larger protein systems. Thousand fold acceleration should be possible for systems with a few thousand atoms.



(a) Free Energy.

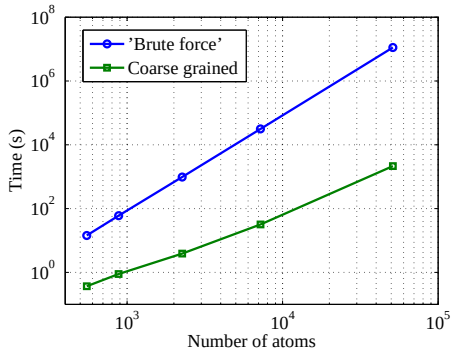


(b) Rate vs. time step

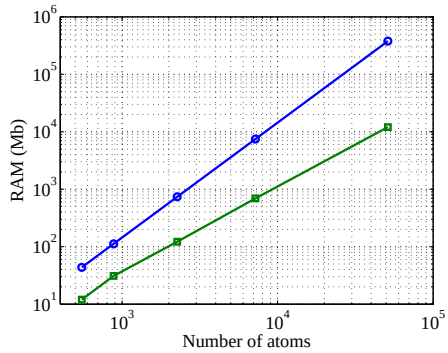


(c) Rate vs. number of modes

Fig. 3. (a) Ramachadran plot for the free energy (in kcal mol^{-1}) of alanine dipeptide using a sigmoidal screened Coulomb potential. (b) Isomerization rate of alanine dipeptide as a function of the time step using 12 modes and re-diagonalization every 100 fs. (c) Rate as a function of varying re-diagonalization frequencies and number of modes.



(a) Diagonalization time.



(b) Diagonalization RAM.

Fig. 4. (a) Comparison of the scaling with time for 'brute force' and coarse grained CNMA methods. (b) Comparison of the scaling with RAM usage for 'brute force' and coarse grained CNMA methods.

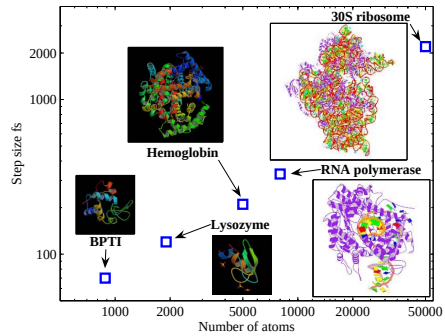


Fig. 5. Prediction of time steps possible using 100 modes and re-diagonalization using CNMA in NML.

Acknowledgment

JAI acknowledges partial funding from NSF grants DBI-0450067 and CCF-0622940. This work is part of a collaboration with Vijay S. Pande, Paula Petrone, and Eric Darve from Stanford University.

References

[1] R. B. Best and G. Hummer. Chemical Theory and Computation Special Feature: Reaction coordinates and rates from transition paths. *PNAS*, 102(19):6732–6737, 2005.

[2] B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics*. Cambridge University Press, 2004.

[3] C. S. M. Levitt and P. Stern. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.*, 181:423–447, 1985.

[4] P. Petrone and V. S. Pande. Can Conformational Change Be Described by Only a Few Normal Modes? *Biophys. J.*, 90(5):1583–1593, 2006.

[5] M.-Y. Shen and K. F. Freed. A simple method for faster nonbonded force evaluations. *J. Comp. Chem.*, 26(7):691–698, 2005.

[6] R. D. Skeel and J. A. Izaguirre. An impulse integrator for Langevin dynamics. *Mol. Phys.*, 100(24):3885–3891, 2002.

[7] C. R. Sweet, P. Petrone, V. S. Pande, and J. A. Izaguirre. Normal mode partitioning of Langevin dynamics for biomolecules. *J. Chem. Phys.*, 128(11):1–14, 2008.

[8] R. Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford, 2001.