THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

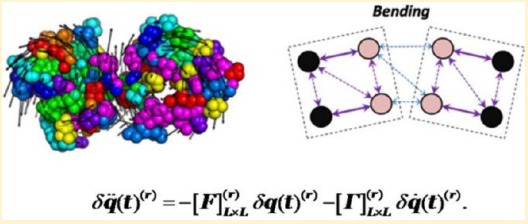
# Response of Rotation–Translation Blocked Proteins Using Langevin Dynamics on a Locally Harmonic Landscape

Anthony C. Manson and Rob D. Coalson*

Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States

Supporting Information

**ABSTRACT:** Langevin dynamics is used to compute the time evolution of the nonequilibrium motion of the atomic coordinates of a protein in response to ligand dissociation. The protein potential energy surface (PES) is approximated by a harmonic basin about the minimum of the unliganded state. Upon ligand dissociation, the protein undergoes relaxation from the bound to the unbound state. A coarse graining scheme based on rotation translation blocks (RTB) is applied to the relaxation of the two domain iron transport protein, ferric binding protein. This scheme provides a natural and efficient way to freeze out the small amplitude, high frequency motions within each rigid fragment, thereby allowing for the number of dynamical degrees of freedom to be reduced. The results obtained from all flexible atom (constraint free) dynamics are compared to those obtained using RTB-Langevin dynamics. To assess the impact of the assumed rigid fragment clustering on the temporal relaxation dynamics of the protein molecule, three distinct rigid block decompositions were generated and their responses compared. Each of the decompositions was a variant of the one-block-per-residue grouping, with their force and friction matrices being derived from their fully flexible counterpart. Monitoring the time evolution of the distance separating a selected pair of amino acids, the response curves of the blocked decompositions were similar in shape to each other and to the control system in which all atomic degrees of freedom are fully independent. The similar shape of the blocked responses showed that the variations in grouping had only a minor impact on the kinematics. Compared with the all atom responses, however, the blocked responses were faster as a result of the instantaneous transmission of force throughout each rigid block. This occurred because rigid blocking does not permit any intrablock deformation that could store or divert energy. It was found, however, that this accelerated response could be successfully corrected by scaling each eigenvalue in the appropriate propagation matrix by the least-squares fitted slope of the blocked vs nonblocked eigenvalue spectra. The RTB responses for each test system were dominated by small eigenvalue overdamped Langevin modes. The large eigenvalue members of each response dissipated within the first 5 ps, after which the long time response was dominated by a modest set of low energy, overdamped normal modes, that were characterized by highly cooperative, functionally relevant displacements. The response assuming that the system is in the overdamped limit was compared to the full phase space Langevin dynamics results. The responses after the first 5 ps were nearly identical, confirming that the inertial components were significant only in the initial stages of the relaxation. Since the propagator matrix in the overdamped formulation is real-symmetric and does not require the inertial component in the propagator, the computation time and memory footprint was reduced by 1 order of magnitude.




## 1. INTRODUCTION

The biological functions of proteins are controlled by their cooperative motions, which typically involve large structural regions.[1] Protein conformational transformations induced by externally applied perturbations, such as ligand binding, are of great biological importance.[2] Introducing a ligand molecule into a binding pocket induces forces on the protein, causing it to deform in ways that have direct functional implications.[3] Specific types of ligand binding include (i) binding of agonist molecules, (ii) binding of molecules used for transport and storage, and (iii) binding of drug molecules sufficient to induce conformational changes that can modulate biological function.[4] The general description of these complex processes has been impaired by the high intrinsic dimensionality of these systems.

Spatial coarse graining schemes have been proposed to reduce the dimensionality of large biomolecules and speed up dynamics simulations.[5] One method is to assign the rigid regions of a biomolecule as the coarse grain elements, called rotation–translation blocks (*RTBs*).[6] In order to identify rigid or semirigid blocks, one must identify the most important and strongest restorative forces and formulate them as constraints.[7] Examples include protein secondary structure, aromatic rings, covalent bond lengths and angles, and the locked dihedral angle associated with the peptide bond.[8] *RTBs* have collective properties such as mass and moment of inertia that can be incorporated into the equations of motion.[9] These reformulated equations can then be used to determine useful dynamical

**Received:** June 19, 2012
**Revised:** August 26, 2012
**Published:** August 27, 2012

   dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142–12158

properties of the system such as (1) vibrational normal modes[6,9] and (2) nonequilibrium response trajectories.[10]

Normal modes are useful for analyzing the structural and dynamical features of macromolecules and have been used for qualitative estimates of many macromolecular properties such as the magnitude of atomic fluctuations,[11] vibrational entropy,[12] and light scattering spectra.[13] They are obtained from diagonalization of the Hessian matrix[13] (second derivative of the potential) and describe global deformations around a local minimum of the molecular potential energy surface.[14] A basis set of modes is sufficient to describe the motion of the system within the locally harmonic region, typically with a modest subset accounting for most of the variation. Expressed in terms of *RTB* structures, normal modes have been found to significantly reduce the degrees of freedom of a dynamical system.[6]

Temporal coarse graining using Langevin models has been used to formulate the response of dissipative systems by adding friction and counterbalancing random forces to the atoms in a molecule.[15,16] Such a formulation replaces microscopic degrees of freedom with ones that are typically collective (macroscopic) and change slowly in comparison. Combined with the expression of motion using *RTB* normal modes, this can significantly reduce the spatial-temporal dimensionality and render complex problems tractable.

The competence of biological systems depends on their response to stimuli. Dynamic linear response theory (LRT) can relate the nonequilibrium motion of the protein atoms that occurs after the ligand molecule dissociates to equilibrium fluctuations within the potential energy surface (PES) of the unliganded protein.[17] LRT in conjunction with the Langevin formulation has been proposed to approximate this response within a locally harmonic subregion of the energy landscape.[17] This response, when integrated with the Fokker–Planck formulation, can be used to calculate the time evolution (drift and diffusion) of the probability density function of the phase space components.[18]

In a previous paper, we presented a Langevin equation based all flexible atom study of nonequilibrium relaxation dynamics of a moderate sized protein, the ferric binding protein (FBP),[17] instigated by dissociation of an $Fe^{3+}$ ligand from its binding pocket. The present work extends this to include the impact of rigid block groupings. Here the effect of the grouping constraints on the system response will be evaluated and compared to the all atom case. The character of the motion, effect of friction, and kinematic errors in the approximation will be discussed. Special attention will be paid to the robustness of the response to variations in decomposition.

**Current Work.** Following the analysis presented in ref 17, we begin with a simplistic but qualitatively defensible multidimensional harmonic oscillator model of the FBP protein motion in the apo-state. With such a model, we can systematically and accurately compare the all-atom flexible motion associated with the relaxation of the protein from the ligand-bound state to equilibrium in the apo-state vs the corresponding motion calculated from a rigid fragment model of the same system, employing the *RTB*-Langevin dynamics formulation of ref 10. There are obvious advantages in computational efficiency as well as insights into the mechanism of conformational change associated with describing the system dynamics as a collection of internally rigid fragments, particularly large fragments.

We used three schemes for decomposing the protein into internally rigid fragments. Thus, part of the point of our study is to understand the strengths/weaknesses of different internal rigidity decompositions. We found that all three decomposition protocols give similar results for the relaxation dynamics of pairs of "tagged" amino acids (i.e., monitoring the selected inter-residue distance as a function of time). Moreover, the motion thus obtained is qualitatively similar to the motion obtained with the corresponding motion obtained when all atoms were fully flexible (not subject to rigidity constraints)—the motion in the internal rigid fragment (blocked) models was found to be systematically faster than that of the all-flexible atom analogue.

To understand the origin of this behavior, we performed ordinary *RTB* vibrational normal mode analysis (developed by Sanejouand et al.[11]) on FBP and compared this to the all-atom normal mode analysis (NMA). We found that the Hessian eigenvalues for the *RTB* version of the analysis were correspondingly ∼4 times larger than for the all-atom flexible analogue. This finding suggested that there was a systematic bias intrinsic to the blocked formulation 6 that induced a stiffer potential energy surface, resulting in forces being transmitted more efficiently across the protein vs in the same protein with all atoms fully flexible. This is not surprising, because if the presumed rigid fragments are actually deformable (via deformation of internal "soft" modes) these fragments would absorb or at least temporarily divert energy flowing across the entire protein structure. If, on the one hand, the blocked regions were actually rigid relative to the nonblocked regions, this effect would not be expected to yield results significantly different from the all flexible atom analogue. However, if the stiffer interactions, such as covalent bonds, traverse block boundaries (a likely occurrence in compact chain polymers),blocking can introduce artifacts into the protein dynamics.

Regardless of these circumstances, we found that, upon empirically scaling down the rigid relaxation dynamics by the slope of the correlation of the blocked vs nonblocked eigenvalue spectra, the dynamics came into very good agreement with the all flexible atom model for the decompositions analyzed. This result can be understood by considering how the stiffness of a block can vary with the block size in a situation where stiff interactions such as covalent and hydrogen bonds permeate the protein. A relatively simple analysis, presented in the Discussion section, shows how blocking a protein along the covalent backbone can induce a higher stiffness relative to the nonblocked case. We show here that a set of $n$ comparable springs (covalent bonds) configured in series (like the protein backbone) has a lower stiffness than just one spring. The softer, series configuration is analogous to the all-atom case and the stiffer, single spring case analogous to the blocked instance (where the remaining $n - 1$ springs in the block are held to a fixed distance by the blocking constraint).

Further analyzing the Langevin modes, we found that only a few modes (<5) were dominant. This finding has implications for identifying critical functionally important modes of motion of large proteins, thereby simplifying the computation of their time evolution.

**Outline of Paper.** In section 2, we review relevant aspects of the theory of Langevin dynamics of coupled harmonic oscillators, as well as the basic principles of rotation translation block (*RTB*) decomposition of a macromolecule into internally rigid fragments each of whose motion can be represented by six degrees of freedom (three overall translations and three overall

rotations). In section 3, we present essential logistical details of our model of ferric binding protein (FBP), an ca. 300 residue protein which will be the focus of our computational studies. In section 4, we follow the time course of the distance between two "tagged" amino acid residues (AARs) on the surface of the FBP molecule, as might be probed, for example, by single molecule FRET.[19] In section 5, we attempt to explain the origin of the temporal scaling relation between all flexible atom ("non-blocked") vs *RTB* ("blocked") versions of the protein system. Because it is easier to analyze, we turn to the overdamped (OD) version of the model. Following the presentation of our basic results in sections 4 and 5, further discussion is provided in section 6 and our conclusions are presented in section 7.

## 2. BASIC THEORY

**RTB Projection Matrix.** The projection matrix implements the dimensional reduction of the relevant macromolecule. As its name implies, it attempts to project out less relevant dimensions from the full phase space. In this study, rigid clustering will be used as the criteria for dimensional reduction.

The projection matrix is a sparse, generally nonsquare, orthonormal matrix composed of *matrix* sub-blocks that map the six *RTB* displacements of each rigid block to the Cartesian coordinates of its member atoms (Figure 1). Full notational details and logical derivations are provided in the Supporting Information.

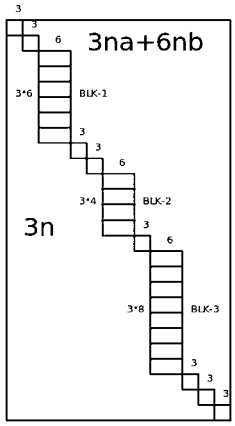



**Figure 1.** Example of projection matrix mapping 11 general *RTB* blocks to 26 atoms. The blocks are a mix of three (nb = 3) *nontrivial* blocks (each having 6, 4, and 8 atoms, respectively) and eight (na = 8) *trivial* (point mass) blocks.

The projection matrix, based on *mass weighted displacements*, is derived from the corresponding Cartesian displacements through

$$\delta \boldsymbol{q}^{(a)} = [\boldsymbol{M}^{1/2}]^{(a)}_{N\times N}\delta \boldsymbol{x}^{(a)}$$
$$\delta \boldsymbol{x}^{(a)} = \boldsymbol{x}^{(a)} - \boldsymbol{x}_0^{(a)} \quad (1)$$

where $N = 3 \times |\{\boldsymbol{atoms}\}|$ and $\boldsymbol{M}$ is the diagonal matrix of atomic masses. The superscript on $\delta \boldsymbol{x}^{(a)}$ and $\delta \boldsymbol{x}^{(r)}$ refers to the subspace in which the vector occurs (e.g., *all atom (AA)* or rotation-translation-block *(RTB)* subspace here; the *RTB* subspace is described below). The notation $[\boldsymbol{H}]_{N\times M}$ indicates a matrix having $\boldsymbol{N}$ rows and $\boldsymbol{M}$ columns. This notation will help to identify the changes in dimensionality in transitioning between the various spaces (all atom, *RTB* coarse-grained, full phase space...).

The original ordering of atoms may not be most suitable for the blocked formulation. It is convenient to group the atoms of a block together because this will simplify the generation of the projection matrix. To achieve this reordering, one can apply a *permutation* matrix.[20] This is a unit matrix in which the *rows* have been permuted. A reordering (a) to (b) has the property that

$$[\boldsymbol{PM}^{(a)\to(b)}]^{-1}_{N\times N} = [\boldsymbol{PM}^{(a)\to(b)}]^{T}_{N\times N} \quad (2)$$

Formally, the *blocking projection matrix*, $\boldsymbol{B}$, maps the mass-weighted *RTB* space to the mass-weighted *AA* space.

$$\delta \boldsymbol{q}^{(a)} = [\boldsymbol{B}]_{N\times L}\delta \boldsymbol{q}^{(r)}$$

inverse:

$$\delta \boldsymbol{q}^{(r)} = [\boldsymbol{B}]^{T}_{L\times N}\delta \boldsymbol{q}^{(a)} \quad (3)$$

where $\delta \boldsymbol{q}^{(sp)} = \boldsymbol{q}^{(sp)} - \boldsymbol{q}_0^{(sp)}$ is the displacement in space $(\boldsymbol{sp})$ and $\boldsymbol{L} = |\{\boldsymbol{dof}_{rtb}\}|$ is the number of degrees of freedom in the reduced dimensional *(RTB)* space.

**Langevin Formulation.** A major contributor to the high dimensionality of solute−solvent systems is the large number of solvent molecules. Hence, one is compelled to find a simplified description of the solvent. The Langevin formulation substitutes the details of the solute−solvent interactions with a frictional force and a random force component that is due to thermal fluctuations.

The *Langevin equation*, in the regime where the thermal fluctuations are zero (i.e., at $T = 0$), is given by

$$\delta \ddot{\boldsymbol{q}}^{(a)}(t) = -\boldsymbol{\Gamma}^{(a)}\delta \dot{\boldsymbol{q}}^{(a)}(t) - \boldsymbol{F}^{(a)}\delta \boldsymbol{q}^{(a)}(t) \quad (4)$$

where $\delta \boldsymbol{q}^{(a)}(t)$ is the all atom, mass weighted displacement vector, $\boldsymbol{\Gamma}^{(a)}$ the mass weighted atomic level friction matrix, and $\boldsymbol{F}^{(a)}$ the mass weighted atomic level force matrix. Note that the force and friction matrices will likely be computed initially using the original (.pdb) ordering. If so, the elements of these matrices will need reordering before being input into the present formulation:

$$\boldsymbol{F}^{(a)} = [\boldsymbol{PM}^{(\boldsymbol{pdb}\to\boldsymbol{blk})}]_{N\times N}[\boldsymbol{F}^{(\boldsymbol{a},\boldsymbol{pdb})}]_{N\times N} \quad (5)$$

$$\boldsymbol{\Gamma}^{(a)} = [\boldsymbol{PM}^{(\boldsymbol{pdb}\to\boldsymbol{blk})}]_{N\times N}[\boldsymbol{\Gamma}^{(\boldsymbol{a},\boldsymbol{pdb})}]_{N\times N} \quad (6)$$

The hydrodynamic radii of the atoms in the protein were estimated from the surface area exposure of each atom to the solvent. The *NACCESS*[21] molecular surface program was used with a probe radius of 1.4 Å to determine the solvent accessible surface area (SASA) for each atom. For a fully exposed atom, a hydrodynamic radius of 0.77 Å was employed. After the degree of surface exposure was determined, the hydrodynamic radius of each atom was scaled to the fully exposed atomic radii using the square root of the fraction of surface accessible area calculated from the *NACCESS* program. For hydrogen atoms, the hydrodynamic radius was set to a maximum value of 0.2 Å if the resulting effective radius was larger than this value. The solvent viscosity was set to $\eta = 1.0$ cP (roughly the viscosity of water at room temperature) in the tagged pair relaxation studies.

The mass weighted *AA* (diagonal) *friction matrix* is defined as


dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142−12158

$$[\boldsymbol{\Gamma}^{(a)}]_{N\times N} = \left[\frac{\gamma_{i,i}}{m_i}\right]_{N\times N} \tag{7}$$

where $\gamma_{i,i}$ is the friction constant associated with atom $i$. Each $\gamma_{i,i}$ is given by the Stokes formula:[22]

$$\gamma_{i,i} = 6\pi\eta a_i, \quad 1 \le i \le N \tag{8}$$

where $\eta$ is the solvent viscosity and $a_i$ is the hydrodynamic radius of atom $i$.

The *RTB* version of the full phase space $T = 0$ Langevin equation is given by[10]

$$\delta\ddot{\boldsymbol{q}}(t)^{(r)} = -[\boldsymbol{F}]^{(r)}_{L\times L}\delta\boldsymbol{q}^{(r)} - [\boldsymbol{\Gamma}]^{(r)}_{L\times L}\delta\dot{\boldsymbol{q}}(t)^{(r)} \tag{9}$$

where the *RTB* projected friction and force matrices are computed from their all atom counterparts by

$$[\boldsymbol{F}]^{(r)}_{L\times L} = [\boldsymbol{B}]^{T}_{L\times N}[\boldsymbol{F}]^{(a)}_{N\times N}[\boldsymbol{B}]_{N\times L}$$
$$[\boldsymbol{\Gamma}]^{(r)}_{L\times L} = [\boldsymbol{B}]^{\mathbf{T}}_{L\times N}[\Gamma]^{(a)}_{N\times N}[\boldsymbol{B}]_{N\times L} \tag{10}$$

**System Propagation Matrix.** The solution of system (9) yields a propagator matrix which prescribes the time evolution of the phase space components from a set of initial conditions. This *propagator matrix* (in *RTB*) space is[10]

$$\boldsymbol{A}^{(r)}_{2L\times 2L} = \begin{bmatrix} 0_{L\times L} & \boldsymbol{I}_{L\times L} \\ -\boldsymbol{\Gamma}^{(r)}_{L\times L} & -\boldsymbol{F}^{(r)}_{L\times L} \end{bmatrix} \tag{11}$$

which, when diagonalized, yields a system having (generally) complex eigenvalues and eigenvectors.

$$\boldsymbol{A}^{(r)}_{2L\times 2L} = [\boldsymbol{P}]_{2L\times 2L}[\boldsymbol{D}]_{2L\times 2L}[\boldsymbol{P}]^{-1}_{2L\times 2L}$$
$$\text{with} \quad [\boldsymbol{D}]_{2L\times 2L} = \begin{bmatrix} \lambda_{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{2L,2L} \end{bmatrix} \tag{12}$$

Here, $[\boldsymbol{D}]_{2L\times 2L}$ is the eigenvalue matrix associated with $\boldsymbol{A}^{(r)}_{2L\times 2L}$ and the columns of $[\boldsymbol{P}]_{2L\times 2L}$ are the corresponding eigenvectors. Once the solution of the propagator matrix is obtained, one can generate the full dynamical response experienced by an arbitrary initial condition (set of atomic displacements and velocities) at any time.

The phase space time evolution of the lab frame Cartesian coordinates is (see the Supporting Information for full details)

$$\boldsymbol{X}^{(a)}_{N}(t)^{\binom{\text{cartesian,}}{\text{lab-frame}}} = [\boldsymbol{M}]^{-1/2}_{N\times N}[\boldsymbol{PM}]^{-1}_{N\times N}[\boldsymbol{B}]_{N\times L}[\boldsymbol{P}]_{2L\times 2L} [\exp([\boldsymbol{D}]_{2L\times 2L}t)]_{2L\times 2L}[\boldsymbol{P}]^{-1}_{2L\times 2L}\boldsymbol{X}^{(r)}_{2L}(0) \tag{13}$$

where

$$\boldsymbol{X}^{(r)}_{2L}(0) = \begin{bmatrix} \delta\boldsymbol{q}^{(r)}_{L} \\ 0 \end{bmatrix}$$

is the phase space initial state having zero initial momentum, and

$$\delta\boldsymbol{q}(0)^{(r)}_{i\in(1\ldots 2L)} = [\boldsymbol{B}]^{T}_{L\times N}[\boldsymbol{PM}]_{N\times N}[\boldsymbol{M}^{1/2}]^{(a)}_{N\times N}\delta\boldsymbol{x}(0)^{(a)}_{j\in(1\ldots N)} \tag{14}$$

is the initial all atom displacement projected to the *RTB* space.

We end this subsection by noting that the protein dynamics we are computing typically occurs at room temperature (ca. $T$ = 300 K), not at or near $T$ = 0. However, within the multidimensional harmonic oscillator model adopted here, expectation values of the atomic displacement coordinates and their corresponding momenta are independent of the system temperature, and, moreover, are determined by the frictively damped harmonic oscillator equations of motion (with no counterbalancing random force terms) noted in eq 4. This follows from the fact that the time-evolving phase space probability distribution function for such a system is Gaussian at all times, and, furthermore, according to this distribution function the expectation values of the coordinates are given by the solution of eq 4. Finite temperature effects manifest themselves in the dispersion ("spread") of the phase space distribution. The equations of motion that prescribe the time evolution of the spread matrix *do* depend on temperature.[10,18] Thus, the average over this Gaussian phase space distribution of any function which is nonlinear in the displacement coordinates and/or their conjugate momenta *will* in general be temperature dependent.

In the present work, our ultimate goal is to compute the distance between two selected atoms, namely, the C$\alpha$ atoms of two tagged residues of FBP. Strictly speaking, this distance is a nonlinear function of the atomic displacement coordinates. However, since the time-variation in all distances of interest (cf. Figures 5−7 below) is small relative to the corresponding reference values (the values they attain in the apo-equilibrium conformation of FBP), a first-order Taylor-series expansion in the displacement variables represents the distance changes of interest to a satisfactory level of approximation. Thus, if $\overline{d(\delta\boldsymbol{q}^{(a)})}$ represents the average of a tagged amino-acid pair distance $d$ over the phase space distribution of the harmonic oscillator system at some time $t$, then we can safely invoke the approximation $\overline{d(\delta\boldsymbol{q}^{(a)})} \cong d(\overline{\delta\boldsymbol{q}^{(a)}})$. However, $\overline{\delta\boldsymbol{q}^{(a)}}$ is precisely the (temperature independent) quantity determined by eq 4. (The same argument obviously applies to the RTB analogue, eq 9.)

**Overdamped Response Formulation.** To further explore the character of the response, it is helpful to consider the formulation in the *overdamped limit* where the acceleration (inertial component) of the Langevin equation is ignored. One advantage of this formulation is that the velocity components of the phase space vector are not needed; i.e., the formulation involves only the configuration space of atomic displacements. This provides a substantial savings in the computation time and memory footprint. Neglecting the acceleration term on the lhs of eq 4, the general *AA* Langevin equation becomes

$$[\boldsymbol{\Gamma}]^{(a)}_{N\times N}\delta\dot{\boldsymbol{q}}^{(a)}(t) = -[\boldsymbol{F}]^{(a)}_{N\times N}\delta\boldsymbol{q}^{(a)}(t) \tag{15}$$

where $\boldsymbol{\Gamma}^{(a)}_{N\times N}$ is the all atom friction matrix, $\boldsymbol{F}^{(a)}_{N\times N}$ is the all atom force matrix, and $\boldsymbol{q}^{(a)}$ is the mass weighted displacement. (The *RTB* instance of the overdamped limit, starting from eq 9, is similar in structure and is derived in the Appendix.)

For notational convenience, we will drop the (a) superscript in the remainder of this subsection. The preceding equation has the solution

$$\delta\boldsymbol{q}(t) = \exp(-\boldsymbol{A}t)\delta\boldsymbol{q}_0 \tag{16}$$

where

$$\boldsymbol{A} = \Gamma^{-1}\boldsymbol{F} \tag{17}$$


dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142−12158

It is useful to perform the transformation

$$\delta \boldsymbol{q}' = \boldsymbol{\Gamma}^{1/2}\delta \boldsymbol{q} \tag{18}$$

Doing so, one obtains

$$\delta \dot{\boldsymbol{q}}'(t) = -\boldsymbol{A}'\delta \boldsymbol{q}' \tag{19}$$

with the symmetric propagator matrix

$$\boldsymbol{A}' = \boldsymbol{\Gamma}^{-1/2}\boldsymbol{F}\boldsymbol{\Gamma}^{-1/2} \tag{20}$$

and the explicit solution

$$\delta \boldsymbol{q}'(t) = \exp(-\boldsymbol{A}'t)\delta \boldsymbol{q}_0' \tag{21}$$

The time evolution of the mass-weighted displacement vector $\delta \boldsymbol{q}$ can then be obtained via the inverse transformation $\delta \boldsymbol{q} = \boldsymbol{\Gamma}^{-1/2}\delta \boldsymbol{q}'$. Clearly, the eigenvalue spectrum associated with $\boldsymbol{A}'$ (or $\boldsymbol{A}$, since these two matrices have the same eigenvalues) should bear some similarity to that of $\boldsymbol{F}$, i.e., to the nonfrictive instance. The eigenvalues, however, are scaled down as a result of the friction on the atoms of the system. As the friction increases, the eigenvalues get smaller and the response along each mode becomes slower.

The friction matrix $\boldsymbol{\Gamma}_{i,j}$ is noninvertible if some atoms in the system have zero solvent accessible surface area (see the Supporting Information). This scenario will occur for larger compact proteins in which the interior is not accessible to the solvent. To accommodate this situation, one can assign a small but finite friction to the zero valued elements. If a value of $10^{-5}$ is assigned to these elements, the matrix can be inverted and the propagator matrix computed. Indeed, in the calculations on FBP presented below, varying the range of values from $10^{-2}$ to $10^{-10}$ had minimal impact on the resultant relaxation dynamics.

**RTB Kinematic Error.** For a successful application of *RTB* blocking to dynamical relaxation processes, the *RTB* projection should duplicate the required all atom initial displacement to a good approximation. Intrinsically, blocking imposes constraints on motion by fixing the interatomic distances within each block. If a block is formed in regions where the interatomic distances undergo small changes, the blocked system will accurately duplicate the all atom motion. However, because proteins are inherently flexible, blocking may introduce error. To assess this error, a metric is needed that can estimate these kinematic defects for the entire protein or selected subregions of it.

**Sources of Kinematic Error.** As a block is perturbed, each atom in the block can undergo rotation about the centroid of the block. In order to preserve the linear form of the block projection matrix, a shear motion (see below) is used in lieu of a length preserving transformation. This approximation can introduce distortion that can increase with block rotation (Figure 2). The following equation relates this tangential, shear-type movement of the atoms within a block to the rotations about the principal rotation axes of each block. These rotations, $\delta\omega_i$, represent three of the six *dof's* for each block and are the lab-frame Cartesian components of the rotational displacement:

$$\underbrace{\begin{bmatrix}\delta t_x\\ \delta t_y\\ \delta t_z\end{bmatrix}_n}_{\substack{\text{cartesian}\\ \text{displacement}\\ \text{of atom n}}} = [\boldsymbol{R}]^{n,m}_{3\times 3}\underbrace{\begin{bmatrix}\delta\omega_x\\ \delta\omega_y\\ \delta\omega_z\end{bmatrix}_m}_{\substack{\text{rotational}\\ \text{displacment}\\ \text{of block m}}} \tag{22}$$

where each column of $[\boldsymbol{R}]^{n,m}_{3\times 3}$ (derived in the Appendix) is a tangent vector that is normal to both (1) the relative position vector (of atom $n$ wrt to the centroid of block $m$) and (2) the corresponding axis of rotation. One can see (Figure 2) that as the angular displacement increases the tangential displacement increases concomitantly and for a sufficiently large $\delta\omega_i$ deviates noticeably from the true arc of rotation.

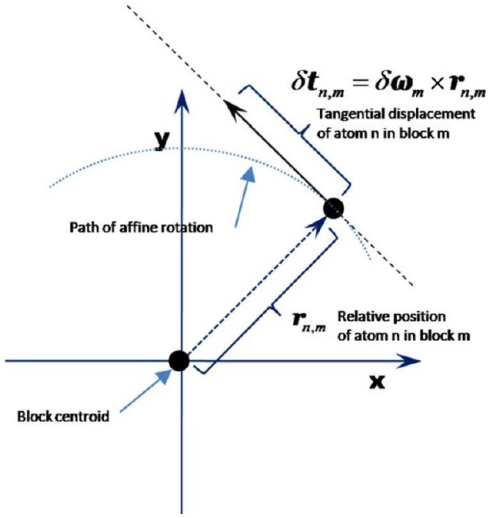



**Figure 2.** Tangential (shear rotation) vector $\delta t_{n,m}$ of atom $n$ of block $m$ in response to infinitesimal rotation about the $z$ axis. The length preserving rotation follows the dotted arc, while the approximated rotation follows a tangential path that is normal to the position vector and the axis of rotation.

Another source of error stems from artificially constraining the interatom distances in regions where the system is in actuality more flexible. Some examples of this occur when the atoms in a block undergo axial compression or if they undergo shifts in registry (contact topology).

In order to capture these errors, we define the kinematic *residual displacement*, which directly represents how well the all atom displacement is duplicated:

$$\delta \boldsymbol{q}^{(res)} = \left[\underbrace{\boldsymbol{I}_{N\times N}}_{\text{all atom}} - \underbrace{[\boldsymbol{B}]_{N\times L}[\boldsymbol{B}]^T_{L\times N}}_{\text{block constrained}}\right]\delta \boldsymbol{q}^{(a)} \tag{23}$$

It can be calculated on any subset of atoms within the molecule, such as the atoms of a single block. The first term on the rhs of eq 23 is the actual, all-atom displacement and the second term is the recovered or *back-projected (bp)* displacement under the blocking constraint. This recovered displacement is calculated by first projecting the all atom displacement into the *RTB* space and then projecting back out to the all-atom space:

$$\delta \boldsymbol{q}^{(bp)} = \left[\underbrace{[\boldsymbol{B}]_{N\times L}}_{rtb\rightarrow aa}\underbrace{[\boldsymbol{B}]^T_{L\times N}}_{aa\rightarrow rtb}\right]\delta \boldsymbol{q}^{(a)} \tag{24}$$

The back projected *RTB* vector can be decomposed into components parallel and perpendicular to the *AA* displacement. If the component that is perpendicular to the actual displacement $\delta \boldsymbol{q}^{(a)}$ is minimized, then the *RTB* approximation to the required all atom motion will be more accurate. It is expected that, as the decomposition becomes more fine-grained, this error will diminish. It is also expected that, if smaller elements are added where the curvature of the applied displacement is higher, the error will also be reduced. In our analysis of FBP (see below), the size of the elements used to approximate the applied displacement is relatively large in comparison to the maximum displacement of any of the atoms. Specifically, since the average block size is on the order of an


dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142−12158

amino-acid residue and the maximum displacement of any atom is smaller than this extent, the kinematic deviation of the blocked system from the all-atom control is expected to be small.

The decomposition of the back projected displacement can be formulated as follows:

$$\delta \boldsymbol{q}^{(bp)} = (\delta \boldsymbol{q}^{(bp)})^{\parallel} + (\delta \boldsymbol{q}^{(bp)})^{\perp} \tag{25}$$

By defining

$$\delta \boldsymbol{uq}_{\parallel}^{(a)} = \frac{\delta \boldsymbol{q}^{(a)}}{|\delta \boldsymbol{q}^{(a)}|} \tag{26}$$

we identify

$$(\delta \boldsymbol{q}^{(bp)})^{\parallel} = (\delta \boldsymbol{q}^{(bp)} \cdot \delta \boldsymbol{uq}_{\parallel}^{(a)}) \delta \boldsymbol{uq}_{\parallel}^{(a)} \tag{27}$$

which determines

$$(\delta \boldsymbol{q}^{(bp)})^{\perp} = \delta \boldsymbol{q}^{(bp)} - (\delta \boldsymbol{q}^{(bp)})^{\parallel} \tag{28}$$

Now, defining the unit vector

$$\delta \boldsymbol{uq}_{\perp}^{(bp)} = \frac{(\delta \boldsymbol{q}^{(bp)})^{\perp}}{|(\delta \boldsymbol{q}^{(bp)})^{\perp}|} \tag{29}$$

we arrive at the equivalent expression:

$$\delta \boldsymbol{q}^{(bp)} = \alpha^{\parallel} \delta \boldsymbol{uq}_{\parallel}^{(a)} + \alpha^{\perp} \delta \boldsymbol{uq}_{\perp}^{(bp)}$$

with

$$\alpha^{\parallel} = \delta \boldsymbol{q}^{(bp)} \cdot \delta \boldsymbol{uq}_{\parallel}^{(a)}$$

$$\alpha^{\perp} = |\delta \boldsymbol{q}^{(bp)} - \alpha^{\parallel} \delta \boldsymbol{uq}_{\parallel}^{(a)}| \tag{30}$$

The value $\alpha^{\perp}$ provides a measure of the kinematic error due to the projection transformation and can be computed for the entire molecule or some subset of the molecule such as a block. This metric will be calculated at $t = 0$ (ps) where the error is expected to be the largest. By correlating this metric on a block-by-block basis with select block parameters, such as (1) block size or (2) rotational displacement of each block, it can be used to check for any systematic bias in which a given formulation (choice of blocks) can influence the results. In this context, systematic bias can refer to kinematic and dynamic perturbations relative to the control case (all atoms are fully flexible) that are due to the blocking operation itself.

**Spectral Projection of Normal Modes.** The spectral projection shows the composition of the atomic displacement vector within a given mode eigenvector decomposition spectrum. The spectrum of projections is determined by the blocking state (all flexible atom vs blocked) and the dynamic mode that applies. The three possible dynamic modes considered here are: (1) ordinary, (2) full-phase-space Langevin, and (3) overdamped Langevin.

To compute the *RTB* spectral projection coefficients $\boldsymbol{\beta}$ from a given all-atom Cartesian displacement $\delta \boldsymbol{x}^{(a)}$, one can use the following steps:

$$\begin{aligned}
&\text{step1: } \delta \boldsymbol{q}^{(a)} = [\boldsymbol{M}]_{N\times N}\, \delta \boldsymbol{x}^{(a)} \\
&\text{step2: } \delta \boldsymbol{q}^{(r)} = [\boldsymbol{B}]_{L\times N}^{T}\, \delta \boldsymbol{q}^{(a)} \\
&\text{step3: } \delta \boldsymbol{uq}^{(r)} = \frac{\delta \boldsymbol{q}^{(r)}}{|\delta \boldsymbol{q}^{(r)}|} \\
&\text{step4: } \boldsymbol{\beta} = \underbrace{[\boldsymbol{nm}^{(r)}]_{L\times L}^{T}}_{\text{basis for dynamic mode}}\, \delta \boldsymbol{uq}^{(r)}
\end{aligned} \tag{31}$$

where $\delta \boldsymbol{uq}^{(r)} = \delta \boldsymbol{q}^{(r)}/|\delta \boldsymbol{q}^{(r)}|$ is the unit normalized *RTB* displacement and $\boldsymbol{nm}^{(r)}$ is the unit normalized *RTB* eigenvector basis (i.e., normal modes) for the given dynamic mode, stored as column vectors.

## 3. MOLECULAR SYSTEM

**Non-Equilibrium Response Calculation.** The response to ligand dissociation was simulated by applying the atomic displacements needed to describe the ligand bound (closed state) protein configuration, and then following its ensuing relaxation. The energy minimized open structure of FBP was used to calculate the harmonic PES and the minimized closed state (along with the open state) used to calculate the initial perturbation. These components, combined within the context of the *RTB* Langevin framework, were used to generate a full phase space Langevin trajectory in response to ligand dissociation.

**Potential Energy Surface Calculation.** The minimized open state of FBP was prepared from the (unbound) X-ray structure (pdb: 1DV9) by conjugate gradient minimization to a gradient rms value of $10^{-6}$ kcal/(Å·mol) in the gas phase using *AMBER-10*.[23] This minimized structure was then used to calculate the all-atom Hessian matrix which prescribes the harmonic energy basin in the neighborhood of the minimum.

**Initial Perturbation.** The initial displacement of *FBP* was calculated from the difference between the liganded (pdb: 1MRP) and the unliganded or apo pdb: 1DV9) states. In order to minimize any excessive strain energy in the closed (liganded or holo) state, the closed state was minimized to $10^{-3}$ kcal/(Å·mol) using *AMBER-10* and subsequently aligned using the *align* module within the *Pymol* molecular modeling system.[24] The initial nonequilibrium perturbation was computed as $\delta \boldsymbol{x}^{initial} = \boldsymbol{x}^{holo} - \boldsymbol{x}^{apo}$. The registry (contact topology) of the closed state was found to be nearly identical to the open state, with the registry of the internal regions being identical and only small deviations occurring on the surface. This all-atom displacement was then mass weighted and projected into the *RTB* space to serve as the initial condition in the time propagation stage of the calculation.

The system matrix linking all the phase space components was processed and propagated in time for >2 ns. Conformations (given in terms of Cartesian displacements) were produced every 1/20 ps and used to calculate the progress of tagged-residue distances. The propagation equations were processed using the linear algebra modules within *LAPACK/BLAS*.

**Zero Energy Displacement.** Since the zero energy modes are not explicitly removed from the initial displacement, the system will relax, in part, along the six zero eigenvalue Langevin normal modes. The asymptotic state of the relaxation, $\boldsymbol{X}_{t\to\infty}^{(r)}$, corresponds to this zero energy translation and rotation. To remove this motion, the asymptotic state is first estimated by calculating the displacement at a large time value (e.g., 3 times the estimated duration of the relaxation). The removal of this motion is applied in the *RTB* space in order to minimize any

distortion due to the *approximated* block rotation (Figure 2) before the lab frame Cartesian coordinates are calculated using the block projection operator. This is not necessary in the all flexible atom instance because the mechanical elements (point masses) do not undergo rotation. To obtain the lab frame coordinates of the relaxation trajectory, the asymptotic state is subtracted from each intermediate state in *RTB* space. This results in relaxation along the *nontrivial RTB* normal modes corresponding to internal deformation of the system. These *RTB* states are then multiplied by the projection matrix to get the mass weighted all-atom displacement vector, which when inverse-mass-weighted provides the Cartesian, lab-frame displacement data.

**Characterizing the Response.** The motion following ligand dissociation is predominantly a two-domain hinge motion (Figures 3 and 4). In order to provide a baseline of

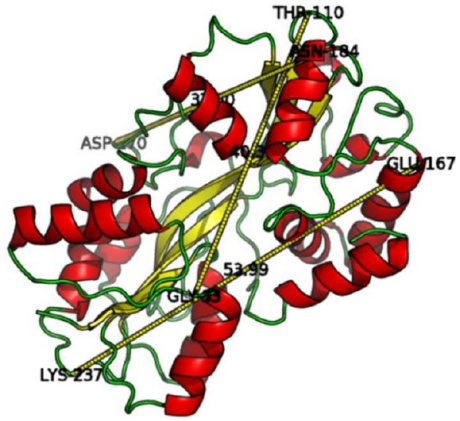



**Figure 3.** Ferric binding protein (pdb: 1DV9) shown in the open (unliganded) conformation. The tagged residue pairs used to report distance relaxation are labeled. This *unbound* state was used to calculate the harmonic potential energy surface (*PES*).

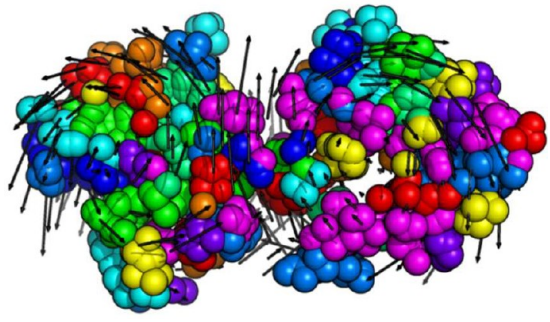


**Figure 4.** Cross section of (*FBP:AC*) decomposition cutting along the main hinge axis, showing coherence of clusters with cleft opening normal mode. The black arrows are the displacements along the "cleft opening" normal mode.

comparison, the same three $\alpha$-carbon pair distances were examined to characterize the responses that were used in a previous all flexible atom (no RTB blocking) study.[17] These pairs were chosen at various angles to the dominant displacement of the hinge motion. The pair E167−K237 is parallel to the cleft opening displacement and has one atom of the pair in each of the two domains. The pair G33−T110 is at a 45° angle to this motion and also spans two domains. Finally, the pair N184−D270 is nearly perpendicular to the motion and is contained in only one domain (Figure 3). The first two pairs, which span the two domains, will undergo a sizable (∼5 Å) change, whereas the last pair, spanning only one domain, is expected to have a more modest (∼1 Å) change.

**Block Decompositions Used.** A rigid decomposition is composed of blocks ranging in size from one atom on up. Blocks of size one (i.e., point mass) have only three degrees of freedom *DOFs*, while composite blocks having two or more atoms will be described using six DOFs. The first three describe the Cartesian translation of the block, and the remaining describe the block rotations. In general, a range of block sizes is needed to provide a good decomposition. For instance, in regions where the atomic movements are more fluid, smaller block sizes may be needed to describe the motion more accurately.

Three block decomposition methods were tested. The first was a straightforward residue level (one block per residue) grouping (*RES*) 6. Here, the decomposition was formed by partitioning the protein chain at the peptide bond. The second decomposition was made using the method developed by Shudler and Niv.[25] This method, called BlockMaster *(BM)*, partitions a protein into semirigid blocks and flexible regions based on residue−residue correlations calculated from approximate vibrational normal modes associated with an elastic network model.[25] The third was an atom level coherence (*AC*) grouping based on correlated motions observed during a low temperature MD simulation, as described in detail below. The notation (protein:decomposition) refers to a specific *RTB* decomposition; e.g., (*FBP:AC*) is the *AC* decomposition of protein FBP.

**AC Decomposition.** The *AC* decomposition, illustrated in Figure 4, was achieved using a *"generate and test"*[26] methodology. Starting with a one block per residue decomposition, it refines the decomposition by using information such as secondary structure, polar contacts, and displacement correlation simulation data. Candidate blocks are formed by marching along the backbone of the protein and refining clusters. Using the secondary structure and polar contacts visible within a molecular modeler,[24] one can perform a first pass refinement of the grouping. As candidate blocks are refined, they are removed from view in order to ease the inspection of the remainder of the protein. These blocks are then tested for coherence by running a gas phase MD simulation from the closed to the open state at 50 K for 10 ps using a Langevin thermostat. This trajectory data is then used to determine the *displacement correlation* matrix of the atoms within each block. The displacement correlation matrix is calculated from the 10 ps trajectory by computing the inner product of the atom displacements (relative to the average displacement for the time interval) obtained from each atomic trajectory with the other atoms found within its neighborhood (7.0 Å cutoff). The elements of the displacement correlation matrix for the correlation of atom $i$ with atom $j$ over the $n$ time steps of the sampling interval are given by

$$[\boldsymbol{DC}_{i,j}]_{N\times N} = [\langle \delta\vec{\boldsymbol{x}}^{i}\cdot\delta\vec{\boldsymbol{x}}^{j}\rangle_{norm}]_{N\times N}$$

$$\text{where}\quad \langle \delta\vec{\boldsymbol{x}}^{i}\cdot\delta\vec{\boldsymbol{x}}^{j}\rangle_{norm} = \frac{1}{\boldsymbol{n}}\sum_{k\in\{1\ldots n\}}\frac{\delta\vec{\boldsymbol{x}}_{k}^{i}\cdot\delta\vec{\boldsymbol{x}}_{k}^{j}}{|\delta\vec{\boldsymbol{x}}_{k}^{i}||\delta\vec{\boldsymbol{x}}_{k}^{j}|} \tag{32}$$

where $\delta\vec{\boldsymbol{x}}_{k}^{i} = (\vec{\boldsymbol{x}}_{k}^{i} - \langle\vec{\boldsymbol{x}}^{i}\rangle)$ is the displacement from the time averaged position of atom $i$, with $\vec{\boldsymbol{x}}_{k}^{i}$ being the Cartesian coordinates of atom $i$ at time index $k$, i.e., $t_k = k\Delta_t$ with $\boldsymbol{k} \in (1, \ldots, \boldsymbol{n})$, $\Delta_t = (10\text{ ps})/\boldsymbol{n}$. Additionally, as just noted, we invoke the


dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142−12158

cutoff $|\vec{x}_k^i - \vec{x}_k^j| < 7.0$ Å. If the two atoms are further apart than this, we set $\langle \delta\vec{x}^i \cdot \delta\vec{x}^j \rangle_{norm} = 0$.

If the displacement correlation of an atom with the other atoms of its assigned block is high (e.g., $[\boldsymbol{DC}_{i,j}]_{N\times N} > 0.85$), the block is unchanged; otherwise, the atom is reassigned to either a different block or to a new block. The process is repeated until all the unassigned atoms are placed into new or existing blocks. The iteration stops when each atom is found to correlate sufficiently well with the other atoms within its assigned block. Since this approach looks for low energy correlated motions of atoms as a means of identifying rigidity, it is similar to the Blockmaster approach used in ref 25.

## 4. DISTANCE RELAXATIONS AND MOLECULAR MOTIONS

**Full Phase-Space Responses.** The intrinsic relaxation of the system can be described by tracking selected interatomic distances. The responses of tagged inter-residue distances were calculated for four decompositions of FBP (one unblocked and three blocked) using the full-phase-space Langevin propagator. Specifically, the relaxation time traces of the three $\alpha$-carbon pairs (*N167−K237, G33−T110, N184−D270*) were constructed for the blocked configurations (*FBP:RES*), (*FBP:BM*), and (*FBP:AC*) and the unblocked counterpart (*FBP:AA*). The shapes of the relaxation curves of the blocked configurations are similar to each other and to the nonblocked instance (Figures 5−7). The blocked responses for the AC and

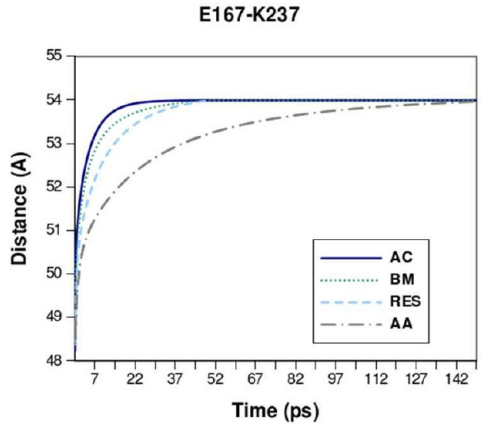



**Figure 5.** Relaxation of $\alpha$-carbon pair [*E167, K237*].

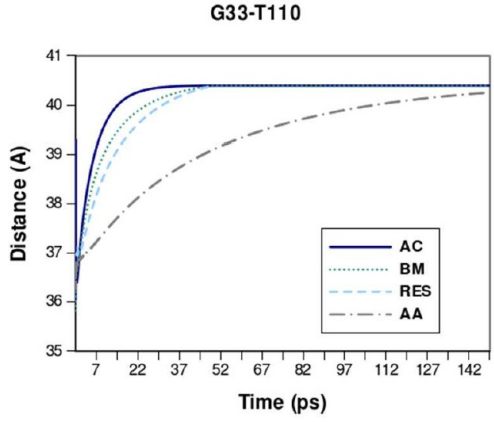



**Figure 6.** Relaxation of $\alpha$-carbon pair [*G33, T110*].

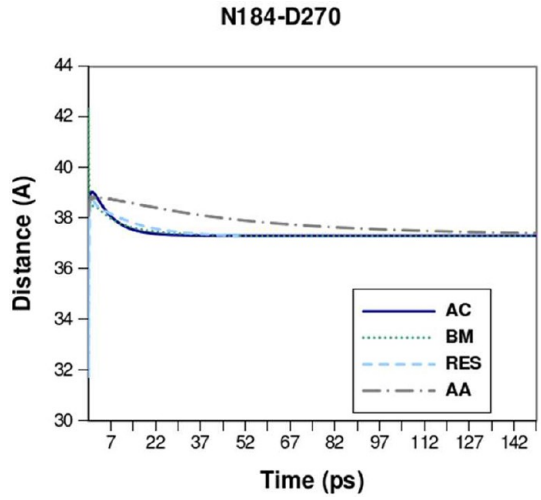



**Figure 7.** Relaxation of $\alpha$-carbon pair [*N184, D270*].

BM decompositions (which contained larger blocks than the RES decomposition) were on average ∼4× faster than the nonblocked counterpart, whereas the RES decomposition was ∼3× faster than the nonblocked counterpart. Qualitatively speaking, these results are "intuitively obvious" because of the better transmission of force across the macromolecule, i.e., by effectively replacing the spongy soft modes within the rigid groupings with stiffer modes. In other words, since the rigid groupings have their interatomic distances fixed, they will not absorb energy and so will perfectly transmit the work through the block. Furthermore, the time dependence of inter-residue distances is consistent with two previous all atom computational studies of all flexible atom Langevin normal modes in proteins,[16,22] which reported time scales of collective motions in the 10 ps to 1 ns range, and also agrees well with our previous calculation of the same quantities.[17]

**Time Scale Correction.** *RTB* blocking generally accelerates the system response compared to the corresponding all-atom instance (Figures 5−7). This effect can be understood in part by observing that the correlation of the eigenvalues of the *RTB* vs *AA* instances of *FBP* is not one-to-one (cf. Figure 15, described below). This nonunit slope is the result of the change in effective resilience (transmission of energy) of the system due to the formation of blocked atomic clusters (see the Discussion for details). The correlation of the blocked and unblocked eigenvalue spectra in this study was found to be nearly linear. Since the propagator eigenvalues map to the decay coefficients in the relaxation, one scheme that could correct for this discrepancy would be to scale the eigenvalues of the (full-phase-space, and as described below, overdamped) *RTB* spectrum by the slope of this correlation.

The eigenvalues of all the blocked instances used in this study were found numerically to be real valued. By uniformly applying a scaling factor (e.g., the slope of the eigenvalue correlation) to each of the eigenvalues in the propagation equation (with no modification of the corresponding eigenvectors), one obtains the propagation equation:

$$\boldsymbol{X}(t) = [\boldsymbol{P}]_{2L\times 2L}[\exp((([\boldsymbol{D}]_{2L\times 2L}t)/\alpha_{(\text{slope})})]_{2L\times 2L}[\boldsymbol{P}]^{-1}_{2L\times 2L}$$
$$\boldsymbol{X}_0 \qquad (33)$$

which yields the three corrected responses presented in Figure 8.


dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142−12158

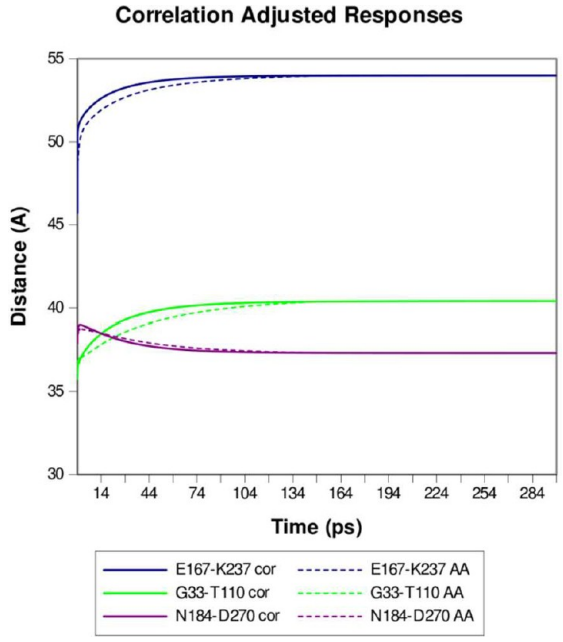



**Figure 8.** Corrected responses for the *AC* grouping: Response of the three distances with the eigenvalues in the propagation equation scaled by the (least-squares fitted) slope of the eigenvalue spectra of the *RTB*, (*FBP:AC*), indicated via solid lines, and *AA* instances (dashed lines). The correction factor used for this decomposition was $\alpha_{(slope)} = 4.2$.

The corrected relaxations demonstrate that using the slope of the eigenvalue correlation, measured at the low end of the spectrum, yields reasonable results (Figure 8). Here, the overall settling time for each distance correlates well with the all atom instance. In the earlier time dynamics, however, one sees that, for two of the distances, the relaxation is accelerated with respect to their nonblocked counterparts (Figure 8). The same correction is applied uniformly to each eigenvector and is assumed to apply for all directions in space. However, since the protein itself is highly anisotropic and the effect of the rigid blocking is also anisotropic, one can see that there will generally be variations in the eigenvectors. Intuitively, if the effective connectivity[27] within the constituent blocks is high, either through high force constants or high topological connectivity or both, the effective transmission of force through the block will be greater. This "force transmission" competence will depend on the size of each block and its effective connectivity. Blocks with higher connectivity that are smaller will transmit force more effectively.

**Representative RTB Langevin Normal Modes.** The full-phase-space Langevin formulation adds friction to the dynamics while retaining the inertial (acceleration) terms. For FBP, the dominant motions extracted from the full-phase-space formulation are similar to those derived from the ordinary vibrational normal modes except that each is induced by a weighted combination of two, similarly shaped, *Langevin* normal modes. In contrast, each of these motions corresponds to a single mode for the nonfrictive case.

To demonstrate this, the initial all atom displacement vector was projected onto the Langevin normal modes and then normalized. For this frictive case, two motions, each composed of two *Langevin* normal modes, induce almost 60% of the spectral weight of the closed state (Figure 9). The first induces 42% of the closed state. The second induces 16%. Motion 1, the cleft opening motion, formed from a combination of two Langevin modes (in a 3:1 ratio), represents the two domain, coordinated, unrolling action needed to expose the $Fe^{3+}$ binding site (Figure 9A). Motion 2, the registry shift motion, formed from two Langevin modes (in a 1.6:1 ratio), is a two domain shearing motion (perpendicular to motion 1) that is needed to position the side chains that interact with the $Fe^{3+}$ ion (Figure 9B). This shearing motion of the two lobes about the binding cleft occurs because of the weak interactions across the cleft. These interactions are weak because the cleft in the apo state is lined with negatively charged groups that stabilize the triply charged ferric ion. These negatively charged groups

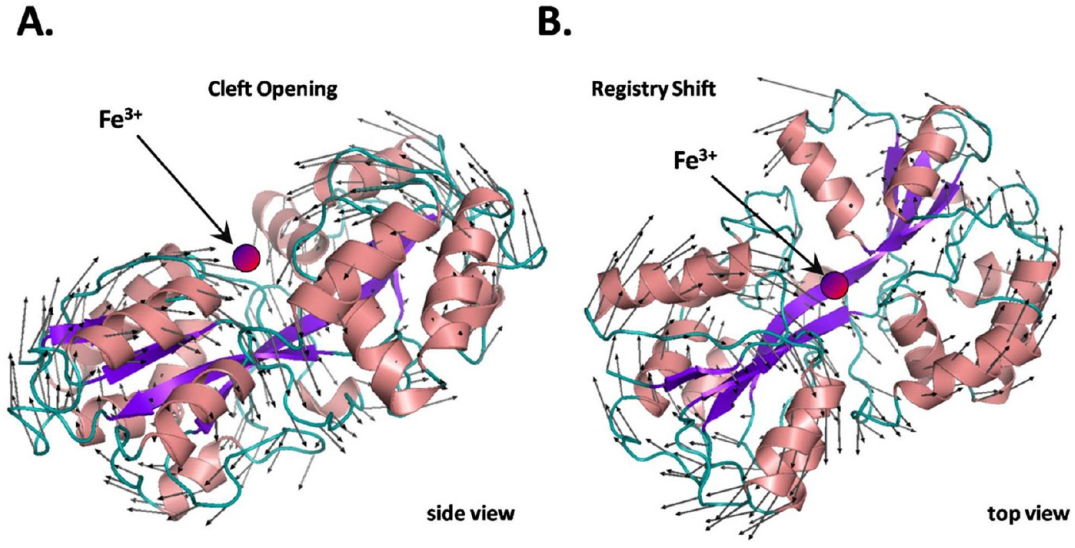



**Figure 9.** Dominant *rtb* motions for the grouping (*FBP:AC*): Two motions, each composed of two (similarly shaped) *Langevin* normal modes, combine to induce >50% of the closed state (see text for further details). (A) Cleft opening motion (side-view). (B) Registry shift motion (top-view). The location of the $Fe^{3+}$ ion is included for orientation purposes. The displacements are shown for the $\alpha$-carbons, where the positive direction is indicated by the black arrows and corresponds to the unbound to bound transition. The ribbon color indicates the secondary structure of the protein (purple, sheet; orange, helix; cyan, loop).


dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142−12158

are mutually repulsive and in the absence of the iron ligand will weaken the interaction across the cleft. This will stabilize the apo state because they weaken the interactions perpendicular to the cleft. In the ligand-bound state by contrast, these groups would articulate with the ferric ion ($Fe^{3+}$) to produce a more rigid region. The positive sense of these motions corresponds to the unliganded to liganded state transition (Figure 9). The cooperative displacement of the dominant modes is shown in Figure 10. Motion 1, in which more residues are displaced, is the more cooperative mode.

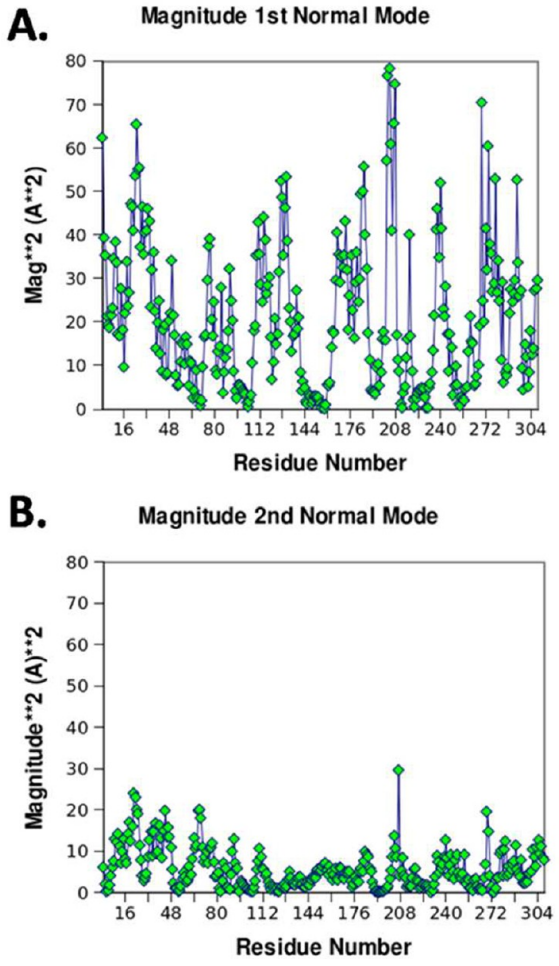



**Figure 10.** Cooperativity of dominant motions: squared magnitude of dominant *rtb ordinary* normal mode displacements. The magnitude of the $\alpha$-carbon displacement is projected out of each *RTB* normal mode to summarize the aggregate, residue level movement. (A) Magnitudes of *ordinary* normal mode 1 by residue. (B) Magnitudes of ordinary normal mode 2 by residue. The higher displacement per residue of the first motion reflects its higher cooperativity.

## 5. INTRINSIC PROPERTIES OF THE BLOCKED SYSTEM

This section examines the detailed mechanical properties of the blocked system such as (1) the dominant normal modes, (2) responses within different dynamic regimes, (3) kinematic error analysis, (4) the effects of friction, and (5) spectral analysis. We will also present relevant data needed to motivate the effect of blocking on the time contraction of the relaxation trajectories noted in the previous section.

**Representative Ordinary (Non-Frictive) RTB Normal Modes.** Ordinary normal modes, which consider only the eigenvectors of the mass weighted force matrix, ignore the effect of friction in the dynamics. The spectral projection of the ordinary normal modes of (*FBP:AC*) is concentrated in the lower frequency part of the spectrum. Expanding the spectral region containing the lowest 50 nonzero frequency normal modes reveals the dominant ordinary normal modes (Figure 11). The lowest frequency part of the spectrum is dominated by

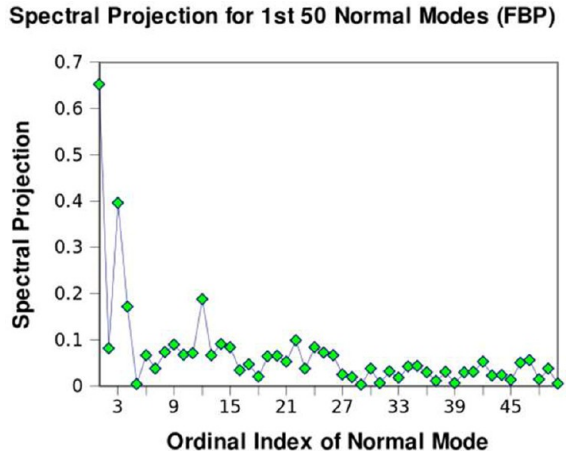



**Figure 11.** First 50 *RTB* normal modes: spectral magnitudes of the first 50 lowest energy nontrivial ordinary *RTB* normal modes of (*FBP:AC*).

two ordinary normal modes that correspond to two functionally relevant motions. These two dominant motions (Figure 9) induce >50% of the closed (bound) state. The most prominent motion (spectral projection ~0.65), representing the cleft opening motion, corresponds to the exposure of the FBP binding site. The next most prominent motion (spectral projection ~0.4), representing the registry shift motion, describes the positioning of the atoms within the binding cleft to best articulate with the ferric ion.

**Representative Overdamped RTB Normal Modes.** The overdamped formulation considers friction while ignoring the inertial contribution to the dynamics. Applying friction to a subset of atoms in the system can shift the spectral distribution of the normal modes of the propagation matrix.[18] In the overdamped limit, this effect is induced by pre- and post-multiplying the force matrix by the inverse half power friction matrix (see eq 20 above). This further shifts the spectral density of the overdamped case to the lower eigenvalue portion of the spectrum relative to the ordinary (nonfrictive) spectrum. This is achieved by diminishing the spectral intensity of the modes in which the nonfrictive atoms undergo the larger displacements. Depending on the distribution of friction in the molecule, this spectral broadening of the low energy frictive modes can produce more (similarly shaped) normal modes corresponding to a given motion (Figure 12). Interestingly, the modes having the largest proportional displacement of the nonfrictive atoms (atoms with very small friction constant situated in the protein interior) have a much shorter relaxation decay time (larger eigenvalue). This means that within the overdamped limit the nonfrictive atoms will tend to reorganize very rapidly in a way that will minimize the harmonic potential energy.

In the ordinary (nonfrictive) normal-mode analysis, we found that two dominant motions, each composed of one normal mode, account for the majority of the collective motion (Figure 11); these modes are similar in shape to those shown in


dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142−12158

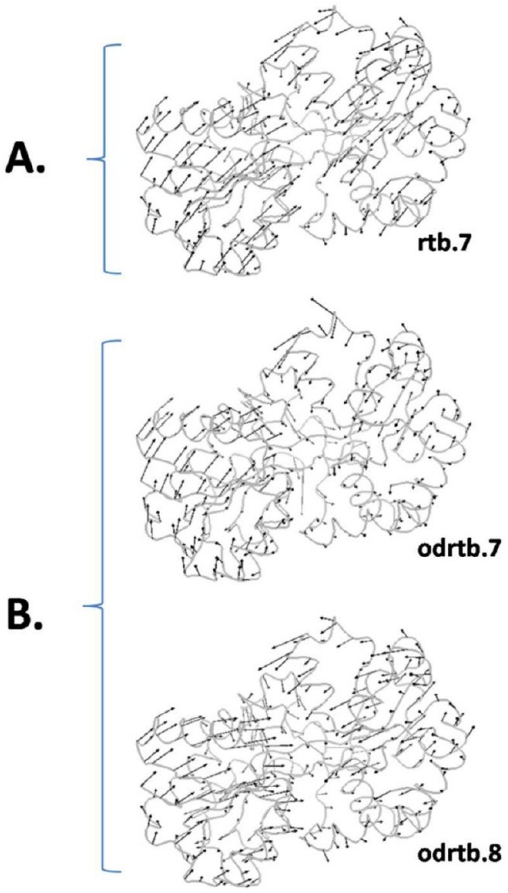



**Figure 12.** There are two (similarly shaped) overdamped *RTB* modes which linearly combine to produce one ordinary normal mode. (A) Dominant cleft opening ordinary normal mode (rtb.7). (B) Two corresponding normal modes from the overdamped Langevin propagator. Displacements are shown for $\alpha$-carbons; ribbon diagram provides protein context; viewing aspect is looking into the iron binding cleft. The mode odrtb.8 is weighted 1.5× greater than mode odrtb.7 in inducing the initial displacement.

(Figure 9). However, as in the case of full phase space Langevin dynamics described above, we find in the overdamped Langevin limit system that the two dominant motions are induced by two sets having two (similarly shaped) modes each. In the overdamped case, two modes combine (in a 1.5:1 ratio) to produce the cleft opening motion (Figure 12) and two modes combine (in a 3:1 ratio) to induce the registry shift motion.

**Overdamped Response.** Using the overdamped regime Langevin approximation, the relaxation curves obtained were nearly identical to the full phase space Langevin curves for times (>5 ps) (Figure 13). The only visible discrepancy occurred in the early stages of the response (see the Discussion for details). It is also worth noting that in the overdamped limit, which uses only the configuration components of the phase space vector, scaling each eigenvalue to adjust the time response (Figure 8) is equivalent to scaling each of the elements of the force matrix by the factor $1.0/\alpha_{(\text{slope})}$.

**Response along the Four Most Dominant Overdamped *RTB* Normal Modes.** The four overdamped modes that had the highest projection on the initial displacement were used to propagate the temporal relaxation response. The response of the tagged residues using only this set of overdamped normal modes compared favorably to the

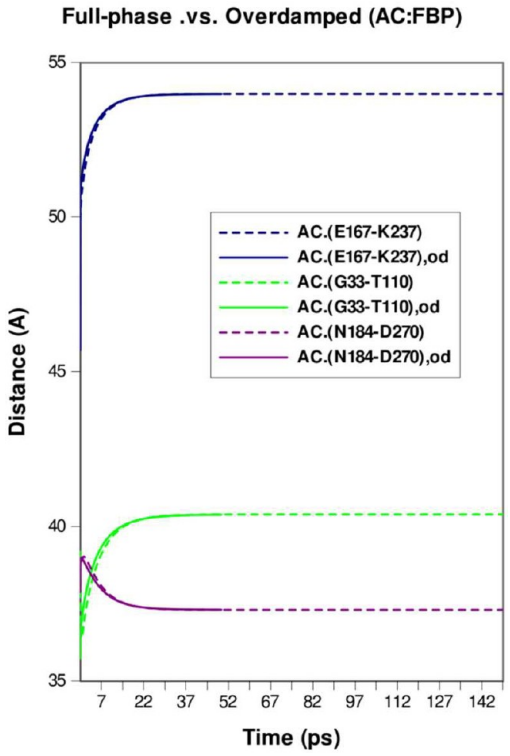



**Figure 13.** Comparison of the full phase space responses to the corresponding overdamped responses for the three tagged pairs within the AC decomposition.

response using the full spectrum of Langevin normal modes (Figure 14). The most apparent discrepancy occurred within the first few picoseconds of each response. This is where the highest eigenvalue modes (which were not among the four highest projection overdamped modes) will have the most pronounced effect. These high eigenvalue modes, however, dissipate quickly so that the response becomes represented by these four dominant modes.

**Properties of Overdamped Response.** One characteristic of the overdamped response is the apparent jump in the initial displacement that is visible at ($t$ = 0 ps) (Figure 14). This is due to the absence of the inertia term in the dynamical formulation. If the intervening atoms that mechanically connect a tagged pair are within a low friction region, this set of atoms will respond almost instantly (along fast nonfrictive Langevin modes) to the displacement perturbation. After the initial jump, the dynamics will be within the diffusive regime and relax on the slower decay components of the spectrum. It is interesting to note that, if the friction assigned to the interior atoms is lowered, the decay time for this *"jump"* becomes smaller (results not shown). One can conclude from this behavior that the atoms on the boundary that experience solvent friction become the rate limiting component of the dynamics, with the low friction interior atoms responding almost instantaneously in order to minimize the harmonic energy within the spring network.

**Projection Error Analysis.** The error metric $\alpha^{\perp}$ was used to estimate how well a blocked configuration replicated the all atom initial displacement. In this case, the relevant displacement was from the unbound to bound state. To test for systematic bias, we computed, on a block by block basis, the correlation of select block parameters with this error metric. A low correlation coefficient was evidence of a small or negligible


dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142−12158

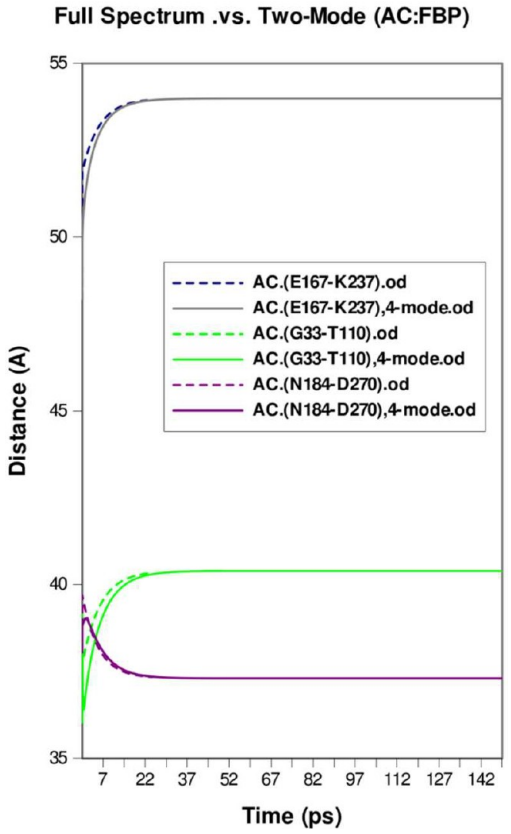



**Figure 14.** Comparison of the response using the full spectrum of overdamped Langevin modes (dashed) and the corresponding response obtained using only the four most dominant nontrivial overdamped *RTB* normal modes (solid) for the decomposition (*FBP:AC*). Tagged pairs: [*E167, K237*], [*G33, T110*], and [*N184, D270*].

systematic bias. The block level parameters examined were (1) the block size and (2) the magnitude of the Cartesian, block level, and angular displacements. The second of these indicated the kinematic error introduced by approximating the true rotation by a shear transformation (Figure 2).

The low correlation (0.19) of the error metric with block size indicated a negligible systematic error. Similarly, the low correlation (0.30) of the metric with the maximum block angular displacements indicated only a minor bias to this parameter.

The Hessian is generally only valid for infinitesimal displacements about an energy minimum. This is especially true for displacements along high eigenvalue eigenvectors, which displace the system along directions of high curvature in the underlying potential energy surface. Variations in the registry (contact topology) are evidence that high energy displacements may have occurred and that the Hessian may be invalid. Registry changes are kinematic defects that can be captured using the current error metric and are important to track because they can result in abnormal energy levels that indicate that the PES may be invalid. For instance, if two atoms connected by a (relatively stiff) covalent bond are exchanged (topologically inverted), the energy (as computed by the Hessian) could be incorrectly inflated. These inaccuracies are less critical if they occur toward the surface of the molecule where groups are free to undergo low energy rotations and translations. In this instance, a simple dihedral rotation or drift within the solvent can relax these to the topology of the computed Hessian. However, if they occur in the buried, densely packed regions of the molecule, they can result in an erroneous bias in the potential energy.

In the bound state, it was found that registry changes only occurred in the outer (solvent exposed) regions of the molecule and that no registry changes occurred on the interior of the protein (i.e., 7.0 Å below the solvent accessible region). Furthermore, these surface level inversions were found to rapidly revert back to the topology of the minimum energy state during the first 5 ps of the response.

**Spectral Correlation of Modes for Blocked and Non-Blocked Instances within the Overdamped Limit.** To better understand the relationship between the blocked and nonblocked instances, it is helpful to supplement the spectral correlation plot for the overdamped modes of *(FBP:AC)* and its all atom counterpart with adjoining plots. The central correlation plot in Figure 15 shows the correspondence of the sorted eigenvalues of the blocked and nonblocked spectra for the overdamped propagator matrix. The adjoining plots show the spectral projection of each of the eigenmodes for either the blocked (upper right) or nonblocked (lower left) instances. Linked through the central plot, one can appreciate how the component modes with the two instances are related.

By visual inspection, one can positively identify each mode in each spectral projection plot. For the two dominant motions observed in the spectrum of the non-friction-weighted, blocked, force matrix (cleft-opening and registry-shift), two modes in the overdamped blocked instance combine to produce the registry-shift mode and, similarly, two modes combine to produce the cleft opening mode, as marked in Figure 15. In the nonblocked instance, however, one mode of the nonblocked overdamped force matrix maps to the cleft opening mode and one mode maps to the registry shift mode. These data points can be identified on the central eigenvalue correlation plot, where it can be seen that they interlace with each other. This shows that the relationship of the blocked to nonblocked modes is few-to-few; i.e., a small set of modes in the blocked instance describes the same motion as a different small set of modes in the nonblocked case. For instance, the relaxation half-life (i.e., the time for the initial displacement to decay to half magnitude) for the representative modes of the blocked and nonblocked instances appears to be scaled by a factor of ∼4.

To further appreciate the nature of the correlation, one can use the data in Figure 15 to relate the response time frame of the two dominant motions. If one considers mode-2 (registry shift mode, m2), one can see that the average decay coefficient for the dual expression of this mode is 0.15. When this is divided by ∼4.2, one obtains 0.038 which is near the value of 0.04 which corresponds to the shift-registry mode in the associated nonblocked instance. In a similar fashion, one may take the average of the *RTB* modes comprising the cleft opening mode m1 (∼0.11) and divide by 4.2 to obtain 0.026. For the relaxation of the AAR pair [*E167, K237*], which undergoes the largest change in distance among the three AAR pairs considered here, this value is within the ballpark even though it results in a slightly faster relaxation than the associated decay of 0.02 of the nonblocked instance would predict (see Figure 5). In support of this, one can see from Figure 8 that the cleft spanning distances [*E167, K237*] and [*G33, T110*] relax at an accelerated pace relative to the single domain distance [*N184, D270*].

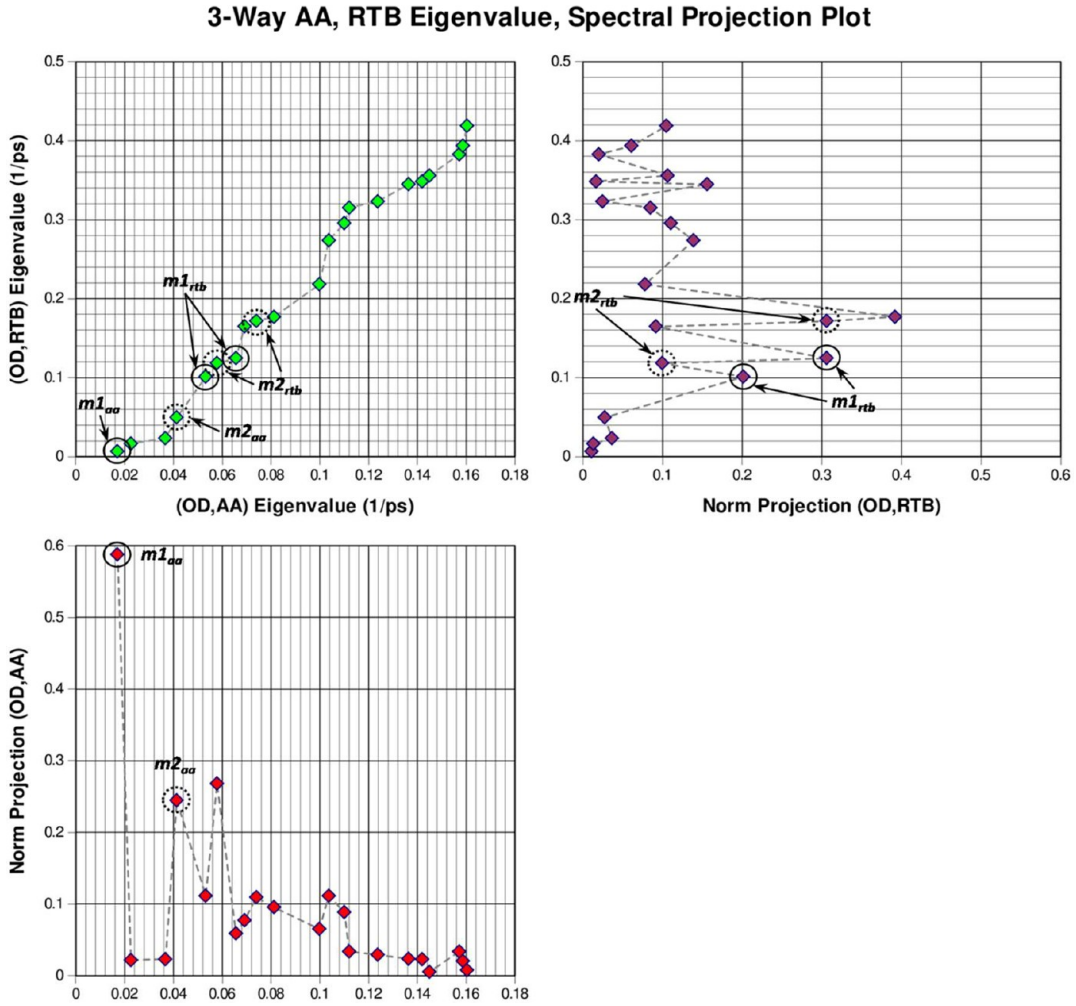



**Figure 15.** Spectral correlation and composition of eigenmodes of blocked and nonblocked instances within the overdamped limit for *(FBP:AC)*. The central correlation plot shows the correspondence of the sorted eigenvalues of the blocked and nonblocked spectra. The adjoining plots show the normalized projection coefficients $\beta_i$ vs the eigenvalues $\lambda_i$ of each of the modes for either the blocked (upper right) or nonblocked (lower left) instances. Linked through the central plot, one can appreciate how the component modes with the two instances are related. The modes corresponding to the cleft-opening (***m1***) and registry shift (***m2***) modes are indicated. The subscript indicates the mechanical context (*AA, RTB*) that the mode denotes.

## 6. DISCUSSION

**Computational Complexity.** *RT* blocking reduces the computation time from the corresponding nonblocked instance. Table 1 shows the total computation time as a function of the number of *DOFs*. This time is roughly cubic with the *DOFs* and corresponds to the time complexity of the diagonalization algorithm.

**Table 1. The Number of Rigid Blocks, Total Degrees of Freedom, and Computation Time (to Generate 1000 Time Steps) (on 3.2Gz Intel Pentium) for Each Decomposition of *FBP***

| full phase space decomposition statistics for *FBP* | | | |
|---|---|---|---|
| name | rigid blocks | DOFs | Calc-time (min) |
| *RES* | 310 | 1864 | 120 |
| *BM* | 38 | 4407 | 300 |
| *AC* | 254 | 5583 | 378 |
| *AA* | 4792 | 14376 | 1980 |

**Blocked Full Phase Space Langevin Normal Modes.** Only two functionally relevant motions, each composed of two similarly shaped full-phase-space *RTB* Langevin modes, dominate each blocked instance. For all decompositions tested, the granularity of the blocks proved to be fine enough to yield similar shaped curves relative to the nonblocked (all flexible atom) analogues for each of the AAR pair distance relaxations considered.

The second important mode (sliding to change registry of the two lobes) was not obvious before calculating it. The shearing type motion between the two lobes is a consequence of the weak interactions of the two lobes in this region. This motion, which is required to position the side chains that articulate with the ferric ion, is arrested once the ion binds. This is a good example of how tertiary components, such as the ferric ion, can modulate the normal modes of the underlying protein (FBP in this instance).

**Contrast with the Elastic Network Model (ENM).** An elastic network model (*ENM*) server[28] was used to generate the normal modes for the open state. The low frequency modes were qualitatively compared to the two most prominent


dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142−12158

Langevin normal modes (*LNM*). A low frequency *ENM* mode homologous to the *LNM* cleft opening mode was found; however, no equivalent mode or combination of modes was found to duplicate the shearing mode needed to position the amino acids that articulate with the ferric ion in the bound state. Since the interactions and the force constants used in each method intrinsically differ, it was not surprising that the shape of the eigenvectors varied. Specifically, for the all atom Hessian of the apo state, the interactions between the two domains at the cleft boundary are much weaker than those within each domain. These weaker interactions occur because the cleft in the apo state is lined with negatively charged groups that stabilize the triple ($Fe^{3+}$) charged ferric ion in the bound state. In the bound state by contrast, these groups articulate with the ferric ion to produce a more rigid region. These weak interactions are not accounted for in the *ENM* model, in which they are approximated by a uniform force constant for all pairwise atomic couplings within a prescribed cutoff distance. In the case of the all atom Amber Hessian, the lack of stiff reinforcement in this region allows the shearing motion to occur along a soft (i.e., low PES curvature) mode.

**Blocking Can Effectively Stiffen the Potential Energy Surface.** The gas phase PES of the molecule used in our analysis is equivalent to a network of harmonic springs. In this context, blocking can increase the overall potential energy because the deformations needed to accommodate the required displacement and satisfy the blocking constraint generally induce a higher potential energy. Blocked regions, by definition, do not store harmonic potential energy because the intrablock distances are fixed to those of the (intrablock) minimum energy configuration. Given a required all atom displacement, blocking forces other regions of the system to deform and absorb this displacement energy. Since the energy stored in a harmonic spring varies *quadratically* with displacement, this reapportionment can induce an increase in the harmonic energy of the nonblocked springs that can exceed the reduction of energy in the blocked springs, which are reset to their zero energy displacements. In all instances of the liganded state where blocking was applied, we observed a net increase in the potential energy relative to the ligand bound all flexible atom (nonblocked) instance. This effect will, however, be mitigated if the relative spring stiffness within the blocks is much larger than those within the nonblocked regions of the system, i.e., if the blocks approach true rigidity.

To illustrate this effect, we consider in Figure 16 a simple collinear system of springs with an outwardly applied axial load on both ends, as an idealization of a relaxing molecule (Figure 16). Note that a similar argument would apply if an inward load were applied to the end atoms. Such an inward load would be analogous to the load inducing the closed state in the FBP system. Since general loading configurations in an all atom system are some variant of axial tension and compression, the case considered here can be used to motivate more complex instances. It is also worth noting that the configuration displayed in Figure 16 is a mechanical analogue of the covalent backbone that threads through a single amino acid residue.

The internal block spring stiffness is $k_{int}$, and the stiffness connecting the two blocks is $k_{\partial B}$. The configuration displayed in Figure 16A is the minimum energy conformation, with the resting distance of the "bonds" (unstretched springs) connecting each atom set to 1.0 Å. We define two stretched configurations. The first is the all flexible atom configuration (Figure 16B), where each atom is free to move independently.

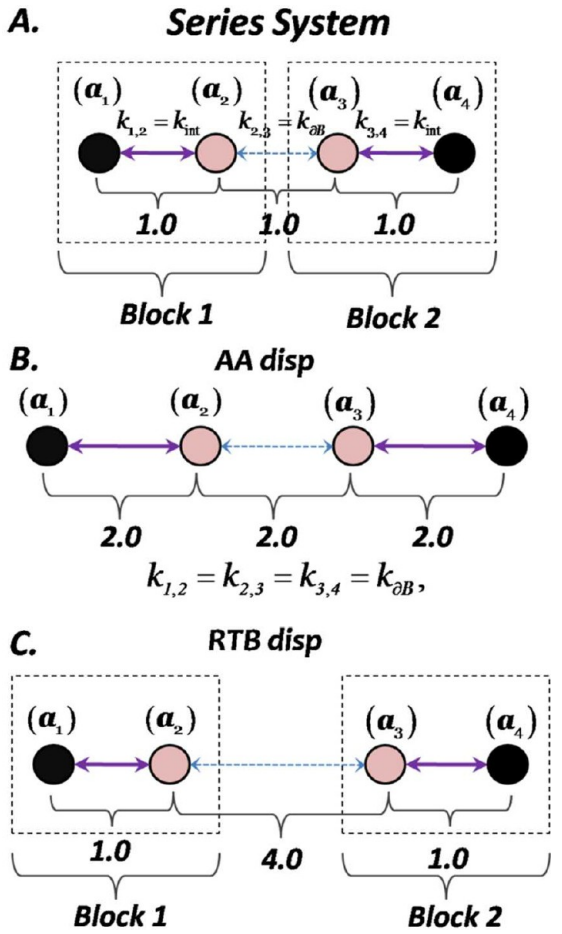



**Figure 16.** (A) Simple linear system having axial harmonic springs in a minimum energy (unperturbed) state. The black atoms are where the external load is applied. This configuration is the minimum energy conformation where the resting distance of the "bonds" connecting each atom is 1.0 Å. (B) The all atom equilibrium configuration due to perturbation applied at boundary atoms. In this instance, the interatom force constants are all equal. Note that the "bonds" are stretched to a length of 2.0 Å. (C) Equivalent blocked system with the applied displacement with the blocking constraint also applied.

The second is the blocked configuration (panel C) where the distances between atoms (1,2) and (3,4) are constrained.

A specified initial displacement places the end atoms (1,4) at the positions shown in panels B and C of Figure 16. Subject to this condition, the atoms will arrange themselves so as to achieve static equilibrium throughout the initial state. This is shown in panel B for the case where all spring constants are the same, and the prescribed distance between the two end atoms is 6 Å. Here each of the bond lengths have been stretched from 1.0 to 2.0 Å. On the other hand, the blocking constraints described above will force the distances between atoms (1,2) and (3,4) to contract, which will, in turn, force the middle spring to distend and absorb more energy than it would otherwise absorb in the all flexible atom instance. This adjustment in conjunction with the required displacement will position the blocks so that the boundary atoms (1,4) match the overall displacement (6 Å) and will result in the configuration shown in panel C. Since the spring energy is a quadratic function of the displacement, the total spring potential energy in the block constrained state will be higher than in the corresponding all flexible atom state.


dx.doi.org/10.1021/jp306030b | J. Phys. Chem. B 2012, 116, 12142−12158

Compute the potential energy for the all flexible atom instance given the applied displacement (panel B) (remember: the stable state has interatomic distances of 1.0 (panel A) so these interatomic distances will increase to 2.0 in order to provide the required overall end-to end displacement):

$$E_{aa} = (k_{1,2}(1)^2 + k_{2,3}(1)^2 + k_{3,4}(1)^2)/2 \tag{34}$$

Now compute the potential energy for the *blocked* configuration given the same applied displacement (panel C). Note that in this case the displacement of the central spring will be reapportioned to 3.0 (taking up the slack of the blocked elements that are reset to their minimum energy displacement):

$$E_{\mathrm{blk}} = (k_{2,3}(3)^2)/2 \tag{35}$$

If all the spring constants are uniform (i.e., $k_{1,2} = k_{2,3} = k_{3,4} = k_{\partial,\mathrm{B}}$), taking the ratio of these two cases shows that the blocked system has the higher energy:

$$E_{\mathrm{blk}}/E_{aa} = 9k_{\partial B}/3k_{\partial B} = 3 \tag{36}$$

It is reasonable to expect that this higher energy will drive a faster relaxation for the blocked case, as discussed in further detail below. Within the overdamped Langevin regime, for instance, where the friction profile is the same in both instances, this higher potential energy with the corresponding larger forces will contract the relaxation time scale (a detailed example is provided in the Appendix).

As a consistency check, it is also worth noting that the ratio of the energies matches the ratio of the effective force constants across the system when the all atom case is replaced by an equivalent series spring configuration. If the ensuing motion is overdamped (diffusive), the block change in inertia becomes zero so that the effective force imparted to a block is the limiting factor in its motion. Here the effective force constant from boundary to boundary becomes the determinant of motion. Under static conditions, the effective force constant through the series of springs shown above is related to the individual force constants by[29]

$$\frac{1}{k_{\mathrm{eq}}} = \frac{1}{k_{1,2}} + \frac{1}{k_{2,3}} + \frac{1}{k_{3,4}} \tag{37}$$

If the distance between atoms (1,2) and (3,4) is fixed (due to blocking) or equivalently through setting $k_{1,2}, k_{3,4} \rightarrow \infty$, then this becomes

$$\frac{1}{k_{\mathrm{eq}}} = \frac{1}{k_{2,3}} \tag{38}$$

which shows that the effective force constant across a block increases by applying the blocking constraint. In the overdamped regime of frictively damped motion, the frictional force will be balanced by the restoring spring force in a quasi-static fashion under conditions where the friction of the boundary atoms be much larger than the internal atoms. Under these conditions, the internal atoms will "instantaneously" adjust their positions to maintain equilibrium with the frictive forces on the boundary atoms. Increasing the effective force constant within this quasi-static limit will increase the potential energy for the same applied displacement. Thus, for the blocked instance, this increased force and potential energy will generally result in a faster relaxation (see the Appendix for details).

Blocking increases the harmonic strain energy when the blocked state is projected out from the corresponding all atom stationary state (strain energy minimum) that is induced by a fixed applied load. The total harmonic energy of the bound state is the sum of the harmonic strain-energy of the molecule and the work done to induce the bound state:

$$E_{tot}^{(a)} = \underbrace{\frac{1}{2}\delta\boldsymbol{q}^{(a)\,T}[\boldsymbol{F}]_{N\times N}^{(a)}\delta\boldsymbol{q}^{(a)}}_{E_{harm\ (strain)}} + \underbrace{\boldsymbol{f}_{bind}^{T}\delta\boldsymbol{q}^{(a)}}_{W_{binding}} \tag{39}$$

where $\boldsymbol{f}_{bind}$ is the equivalent applied binding force, which in the case of *FBP* is supplied by the $Fe^{3+}$ ion. The minimum occurs when the first variation in energy $\delta(E_{tot}^{(a)})$ due to perturbation $\delta(\delta\boldsymbol{q}^{(a)})$, about the stationary point (i.e., the static equilibrium displacement, $\delta\boldsymbol{q}_{stat}^{(a)}$) is zero:

$$\delta(E_{tot}^{(a)}) = ([\boldsymbol{F}]_{N\times N}^{(a)}\delta\boldsymbol{q}_{stat}^{(a)} + \boldsymbol{f}_{bind}^{T})\delta(\delta\boldsymbol{q}^{(a)}) = 0 \quad \text{or}$$
$$\delta\boldsymbol{q}_{stat}^{(a)} = -[[\boldsymbol{F}]_{N\times N}^{(a)}]^{-1}\boldsymbol{f}_{bind} \tag{40}$$

In the *FBP* analysis, the harmonic strain energy of the initial state was coincidentally minimized when the conjugate gradient method was applied to temper the maximum net atomic force to a modest, but nonzero, gradient rms value. This was done in part to relax any unrealistic forces (due to the removal of the $Fe^{3+}$ ion) that may occur at the outset of the relaxation. This strain minimized state was then projected to the blocked state by using $\delta\boldsymbol{q}_{init}^{(r)} = [\boldsymbol{B}]_{L\times N}^{T}\delta\boldsymbol{q}_{stat}^{(a)}$. Using the general formula for the all atom harmonic strain energy

$$\Delta E_{harm}^{(a)} = \frac{1}{2}\{\delta\mathbf{q}^{(a)}\}^{T}[\boldsymbol{F}]_{N\times N}^{(a)}\delta\boldsymbol{q}^{(a)} \tag{41}$$

the increase in energy due to blocking, relative to the strain-minimized (stationary) state $\delta\boldsymbol{q}_{stat}^{(a)}$, can then be calculated using

$$\delta\boldsymbol{q}_{(res)}^{(a)} = \left[\underbrace{[\boldsymbol{B}]_{N\times L}[\boldsymbol{B}]_{L\times N}^{T}}_{blocking} - [\boldsymbol{I}]_{N\times N}\right]\delta\boldsymbol{q}_{stat}^{(a)} \tag{42}$$

Substitution into eq 39, followed by the imposition of the condition indicated in eq 40, yields for the change in harmonic strain energy relative to the strain-minimized configuration

$$\Delta E_{harm}^{(blocking)} = \frac{1}{2}\delta\boldsymbol{q}_{(res)}^{(a)\,T}[\boldsymbol{F}]_{N\times N}^{(a)}\delta\boldsymbol{q}_{(res)}^{(a)} \tag{43}$$

Since the all atom Hessian matrix $[\boldsymbol{F}]_{N\times N}^{(a)}$ is positive semidefinite, the change in energy from the stationary state will necessarily be positive.

**Rigid Elements Must Be Load Bearing.** The atoms of a rigid cluster will tend to move in a coherent manner. However, to be truly rigid, they should also be able to support and transmit a load; otherwise, they will store or dissipate energy. Depending on the method of rigid decomposition, one may end up with blocks that cannot bear a load in a rigid fashion. If rigidity is assessed by coherent motions, one may mistake this "circumstantial rigidity" for true rigidity. For instance, one may have a relatively flexible structure adjoining a rigid structure which by virtue of its circumstantial configuration may appear rigid when analyzed using a "motion coherence" method as employed in the current study. If rigidity is incorrectly imparted to such blocks, this can bias the dynamics of nonequilibrium relaxation processes. Such a problem can be alleviated if intrablock flexibility is explicitly incorporated into the mechanical model.

## 7. CONCLUSIONS

**Impact of Blocking on System Responses.** Blocking provides a continuum description of a macromolecule that can

be used to understand and simulate complex mechanical properties. It is sufficient to model larger scale mechanical characteristics such as bending and shearing deformations that dominate lower energy molecular motions and is capable of modeling a diverse range of loading conditions present in macromolecules. In addition to this, blocking significantly reduces the modeling complexity.

Rigid blocking works best when the following conditions concurrently hold: (1) the number of degrees of freedom used to describe the system motion is minimized, (2) the blocks are compact, (3) the intrablock stiffness is much greater than the interblock stiffness, and (4) the decomposition is able to simulate functionally relevant motion. Condition 1 favors decompositions having the most atoms assigned to the largest blocks possible and ensures the most computationally efficient formulation. Condition 2 assures that each block is free to undergo realistic motions in response to an applied load. Condition 3 favors blocking configurations where stiff elements do not occur in the interblock regions. Condition 4 is necessary in order to maintain functional relevance. Therefore, to reduce the computational complexity of the problem and maintain dynamical accuracy, it is desirable to assign all the atoms within a protein to highly rigid groups of nontrivial size and compact shape, with a minimum number of stiff connections between blocks. When these criteria are jointly met, rigid blocking is expected to be efficient and to introduce minimal bias on the system dynamics.

We found that blocking can affect the potential energy surface by inducing a steeper harmonic energy well. If the stiffness criterion for well formed rigid blocks is not satisfied (condition 3 above), a bias may be introduced into the dynamics in which the effective stiffness may be greater than the control, i.e., the all flexible atom counterpart. For proteins, the ubiquitous distribution of covalent bonds within a globular protein domain suggests that nontrivial, compact, highly rigid clusters having soft intercluster connections are not prevalent. To illustrate this, note that for any complete (all atoms assigned to blocks) decomposition of a protein into regularly placed, compact blocks of nontrivial size (e.g., the residue-by-residue grouping) the pseudorandom thread of the stiff covalent backbone will cross the block boundaries (twice for non-terminal residues) and tend to negate the *optimal* stiffness criterion. As a result, the formation of rigid clusters within this context will force the stiffer elements within the interblock regions to take up the slack, inducing a steeper potential energy surface relative to the all-atom, fully flexible counterpart.

Fortunately, one is not necessarily limited to a purely rigid block formulation. It may be possible to impart some level of flexibility to a block in order to allow for energy storage. Such a quasi-rigid model would address the problems encountered in this study and provide more accurate dynamics within the scope of the overdamped regime. The quasi-rigid formulation would be more tolerant of systems permeated with stiff interactions such as covalent bonds. Here, one would include terms in the formulation that would allow potential energy storage (either explicitly or implicitly) within each block. This solution would become optimal if the number of additional *DOFs* needed to describe the internal energy storage is minimal.

To motivate a more explicit correction scheme, consider that, within the overdamped regime, the time rate of change of motion of the nonfrictive atoms/blocks within the interior of the molecule will be much faster than those on the surface where friction is dominant. Under these circumstances, the intrablock potential energy of the interior atoms/blocks will tend to adjust on a much faster time-scale compared with the aggregate (molecule-level) motions that (in the overdamped limit) are limited by the viscous forces on the frictive atoms/blocks. This may allow the change in the intrablock potential energy to be modeled implicitly as part of the interblock stiffness matrix and would not add any additional degrees of freedom to the problem. At the root of this correction would be an analysis of the effective stiffness of each block done in a fashion similar to the case study presented in the Discussion. The present results (which use a uniform scaling of the force matrix) suggest that some form of softening of each of the elements of the interblock force matrix may yield a reasonable refinement to each element of the stiffness matrix. To achieve this, an initial correction (similar to that used in the current analysis) could be applied and then refined by assessing the localized deformations about the 3d neighborhood of each block.

## APPENDIX

### Damped Collinear Spring System

Consider a one-dimensional system of $n$ collinear point masses connected by springs similar to the situation shown in Figure 16A. Consider the case where both ends are fixed and allow the other masses to move freely. Apply a simple change of variables such that

$$\delta x_i' = \delta x_i - \delta x_{i-1}, \quad i \in \{2...n\}, \quad i = 1, n; \quad \text{fixed ends} \tag{A1}$$

where $\delta x_i$ is the displacement of mass $i$ from its equilibrium position (when all springs in the system are unstretched). [Note: In this Appendix, we work in laboratory Cartesian coordinates; mass weighting is neither necessary nor convenient here.] If the load $f_{load}$ is applied in equal and opposite directions to the end masses 1 and $n$ and all the intervening masses are in mechanical equilibrium, then

$$\delta x_i' = \frac{f_{\text{load}}}{k_{i-1,i}} \tag{A2}$$

where $k_{i-1,i}$ is the spring constant of the spring connecting atoms $i-1$ and $i$. Since

$$\delta x_{tot}' = \sum_{i\in\{2...n\}} \delta x_i' \tag{A3}$$

then

$$\delta x_{tot}' = \sum_{i\in\{2...n\}} \frac{f_{load}}{k_{i-1,i}} \tag{A4}$$

The effective force constant across this system is

$$\frac{1}{k_{eff}} = \frac{\delta x_{tot}'}{f_{\text{load}}} = \sum_{i\in\{2...n\}} \frac{1}{k_{i-1,i}} \tag{A5}$$

which is the familiar relationship for the effective spring constant of a system of springs configured in series.[29]

Now, if the system is released at $t = 0$ from the static configuration just prepared (i.e., the external forces used to displace the end atoms in the preparation step are removed), and undergoes overdamped zero-temperature Brownian

motion, the atomic displacements will evolve in time according to

$$[\mathbf{\Gamma}]_{n\times n}\delta\dot{\boldsymbol{x}}(t) = -[\boldsymbol{F}]_{n\times n}\delta\boldsymbol{x}(t) \quad \text{(A6)}$$

where $[\mathbf{\Gamma}]_{n\times n}$ is the relevant friction matrix, which we will take to be diagonal, with diagonal elements given by the atomic friction coefficient appropriate to each atom, and $[\mathbf{F}]_{n\times n}$ is the force constant matrix (again, not mass weighted) associated with the collinear spring system under study. In the case where each internal friction constant $\gamma_{2,3,\ldots,n-1}$ is much smaller than either of the end atom friction constants $\gamma_{1,n}$, the internal atoms will experience a state of quasi-equilibrium. That is, the net force on each of these atoms will remain approximately zero for all times. In this quasi-static relaxation process, the forces on the end atoms prescribed by eq A5, i.e., $f_1 = k_{eff}(\delta x_n - \delta x_1) = -f_n$, hold *for all times*. To analyze a specific case explicitly, let us take the atomic friction constant describing both of the end atoms to be the same, i.e., $\gamma_1 = \gamma_n \equiv \gamma_0$, and focus on the end atom displacement difference $\delta x_n(t) - \delta x_1(t) \equiv D(t)$. (When the equilibrium chain length is added to $D(t)$, this variable represents the physical distance between the end atoms of the chain.) We then have $\gamma_0\dot{D}(t) = -2k_{eff}D(t)$, which integrates to $D(t) = D(0)\exp(-2k_{eff}t/\gamma_0)$. This illustrates the critical role of the effective force constant $k_{eff}$ in the dynamical process of interest. In particular, the time scale of the relaxation is inversely proportional to $k_{eff}$.

## ■ ASSOCIATED CONTENT

### Supporting Information

The detailed derivation of the working equations used in this study. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: coalson@pitt.edu.


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We gratefully acknowledge financial support from NSF grant CHE-0750332, BSF grant 2007296, and the computational resources of the Center for Molecular and Materials Simulations (CMMS) at the Univ. of Pittsburgh. We thank Dr. S. Essiz and Dr. M. Y. Niv for helpful discussions and Dr. Niv, M. Shudler, and M. Shavro for providing us with Blockmaster groupings for FBP.

## ■ REFERENCES


(1) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964−972.
(2) Ansari, A.; Berendzen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5000−5004.
(3) Pellicena, P.; Kuriyan, J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 702−709.
(4) Barry, P. H.; Lynch, J. W. *IEEE Trans. Nanobiosci.* **2005**, *4*, 70−80.
(5) Sherwood, P.; Brooks, B. R.; Sansom, M. S. P. *Curr. Opin. Struct. Biol.* **2008**, *18*, 630−640.
(6) Tama, F.; Sanejouand, Y. H. *Protein Eng.* **2001**, *14*, 1−6.
(7) Holger, G.; Thorpe, M. F. *Biophys. J.* **2006**, *91*, 2115−2120.
(8) Thorpe, M. F.; Lei, M.; Rader, A. J.; Jacobs, D. J.; Kuhn, L. A. *J. Mol. Graphics Modell.* **2001**, *19*, 60−69.
(9) Li, G.; Cui, Q. *Biophys. J.* **2002**, *83*, 2457−2474.
(10) Essiz, S. G.; Coalson, R. D. *J. Chem. Phys.* **2007**, *127*, 104109.
(11) Durand, P.; Trinquier, G.; Sanejouand, Y. H. *Biopolymers* **1994**, *34*, 759−771.
(12) Brooks, B. R.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A* **1983**, *80*, 6571−6575.
(13) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill: New York, 1955.
(14) Leach, A. *Molecular Modeling, Principles and Applications*, 2nd ed.; Pearson Education Limited: Harlow, U.K., 2001.
(15) Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press: New York, 2001.
(16) Ansari, A. *J. Chem. Phys.* **1999**, *110*, 1774−1780.
(17) Essiz, S. G.; Coalson, R. D. *J. Phys. Chem. B* **2009**, *113*, 10859−10869.
(18) Lamm, G.; Szabo, A. *J. Chem. Phys.* **1986**, *85*, 7334−7348.
(19) Weiss, S. *Nat. Struct. Biol.* **2000**, *7*, 724−729.
(20) Horn, R. A.; Johnson, C. R. *Matrix Analysis*; Cambridge University Press: Cambridge, England, 1987; p 25.
(21) Hubbard, S. J.; Thornton, J. M. *NACCESS, Department of Biochemistry and Molecular Biology*; University College: London, 1993.
(22) Kottalam, J.; Case, D. A. *Biopolymers* **1990**, *29*, 1409−1421.
(23) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R.; et al. *AMBER 7*; University of California: San Francisco, CA, 2002.
(24) Delano, W. L. *The Pymol Molecular Graphics System*; Delano Scientific: San Carlos, CA, 2002; http://www.pymol.org.
(25) Shudler, M.; Niv, M. Y. *J. Phys. Chem. A* **2009**, *113*, 7528−7534.
(26) Russell, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*; Prentice Hall: Upper Saddle River, NJ, 2010.
(27) Fiedler, M. *Combinatorics Graph Theory* **1989**, *25*, 57−70.
(28) Yang, L. W.; Rader, A. J.; Liu, X.; Jursa, C. J.; Ching, S. C.; Karimi, H.; Bahar, I. *Nuc. A. Res.* **2006**, *34*, 24−31.
(29) Rajasekaran, S.; Sankarasubramanian, G. *Computational Structural Mechanics*; Prentice-Hall: Cambridge, England, 2001; p 63.