COARSE-GRAINED TECHNIQUES TO STUDY DYNAMICS OF LONG-TIME SCALE CONFORMATIONAL CHANGES OF PROTEINS

by

Sebnem Essiz Gokhan

B.S., Koc Universirty, 2002

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

School of Arts and Sciences

This dissertation was presented

by

Sebnem Essiz Gokhan

It was defended on

April 6, 2009

and approved by

Kenneth D. Jordan, Ph. D.

Maria Kurnikova, Ph. D.

Craig Wilcox, Ph.D.

Dissertation Advisor: Rob D. Coalson, Ph. D.

COARSE-GRAINED TECHNIQUES TO STUDY DYNAMICS OF LONG-TIME SCALE CONFORMATIONAL CHANGES OF PROTEINS

Sebnem Essiz Gokhan, Ph.D.

University of Pittsburgh, 2009

The Rotation and Translation block (RTB) method of Durand et al. [1] [Biopolymers **34**, 759 (1994)] and Tama et al. [2] [Proteins **41**, 1 (2000)], which is an appealing way to calculate low frequency normal modes of biomolecules by restricting the space of motions to exclude internal motions of pre-selected rigid fragments within the molecule, is extended to a method for computing the Newtonian dynamics of a biomolecule, or any large molecule, with effective rigid-body constraints applied to a pre-chosen set of internal molecular fragments. This method, to be termed RTB-dynamics does not require the construction of the matrix of second spatial derivatives of the potential energy function, and can be used to compute the classical dynamics of a system moving in an arbitrary anharmonic force field with an efficient way of freezing out the high frequency motions within rigid fragments. Moreover, an approximation scheme is developed to compute Brownian motion according to the Langevin Equation for a molecular system moving in a harmonic force field and characterized by one or more rigid internal fragments by using the RTB methodology. To illustrate these methods elementary numerical applications to signal propagation in the small membrane-bound polypeptide gramicidin-A are presented.

Finally, Dynamic Linear Response Theory (DLRT) is adapted to the problem of computing the time evolution of the atomic coordinates of a protein in response to the unbinding

of a ligand molecule from a binding pocket within the protein. DLRT relates the non-equilibrium motion of the protein atoms which ensues after the ligand molecule dissociates to *equilibrium* dynamics in the force field, or equivalently, on the potential energy surface (PES) relevant to the unliganded protein. We numerically illustrate the application of DLRT for a simple harmonic oscillator model of the ferric binding protein, and for an analogous model of T4 lysozyme including the solvent effects on the motion using the Langevin prescription. Using a physically appropriate value of the viscosity of water to guide the choice of friction parameters, we find relaxation time scales of residue-residue distances on the order of several hundred ps. Comparison is made to relevant experimental measurements.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I am sincerely grateful to my advisor Prof. Rob D. Coalson for his guidance, suggestions and dedication throughout this study. He set such high standards in terms of being a good scientist and defining the phrase “decent research study”; I believe if I can imitate this side of him even with %10 percent yield, I will be in a good shape in the future. Special thanks to Prof. Maria Kurnikova for very helpful discussions and her valuable suggestions. I would like to thank my committee member, Prof. Ken Jordan for the all intellectual and moral support through this study. I also would like to thank Prof. Craig Wilcox for being in my doctoral committee and his suggestions.

I am greatly indebted to my dear colleagues and friends from chemistry department. To Mary Cheng, Richard Christie, Bill Kowallis and Artem Mamonov for their support about the various problems I encountered during the study in addition to their valuable friendship. I would like to extend my special thanks to Kadir Diri for his invaluable friendship and support. I never thought graduate school would be the place where I would find the big brother that I have been looking for. I also would like to acknowledge the support of Erikah Englund and Aparna Agarwal through this process. They were the best “Ph.D. Companions” that I could have.

I am also grateful to Sakine Batun and Zeynep Erkin for all of their support. If they were not standing by me during these last months, I wouldn’t be able to turn Pittsburgh into a home again.

Two precious people, Suda and Ismail Essiz have been there since the “beginning”. My mom and dad have spent all their life to make sure their only child gets the best possible education, and has the opportunity for success. I can not pay my debt to them by any means. All I am able to do is thank them. Mom, dad, this is for you.

Last but not the least, completion of this thesis would not be possible without my husband, Mehmet Gokhan, who went through every stage of this study together with me. I find it needless to thank him here and had I tried, I would not have known how to, either. Simply, I would say that he should be awarded with “half-of my degree”, but I am not sure whether he considered having a Ph.D. in Chemistry when proposing to me :)

.

1.0 INTRODUCTION

1.1 DYNAMICS OF BIOMOLECULES

Computational studies have proven to be valuable for understanding the dynamics of the proteins at a level of resolution that is necessary to get a complete picture of how they function. Since biomolecules such as proteins, nucleic acids, and polysaccharides change their structures or conformations both "on their own" and when they interact with other biomolecules, a single static structure does not provide enough information to fully explain how they function. Although the conformational substates of proteins (located in energy wells) and their rates of inter-conversion can in happy circumstances be detected experimentally, obtaining a structural description of the transition state configurations directly via experimental studies is very challenging because of the low probability and short life time of the high energy conformers on the transition pathway from one static substate to the other. Thus, computational simulation studies which follow the precise position of each atom at any instant in time are valuable in terms of connecting the static pictures of proteins to their action. When the time scale of the conformational changes of biomolecules is considered, it scans a very broad range from femtoseconds to hours (cf.Table 1). Dynamics on the $\mu s - ms$ time scale have received much attention recently, because many biological processes such as enzyme catalysis, protein-protein interactions, and signal transduction occur on this time scale. However, to be able to capture the

motions on the time scale of milliseconds or even microseconds for such processes is very challenging with today's computational power when the size of the biomolecule under study reaches 10^4-10^5 atoms, which is the case with most multi-subunit protein complexes[3].

Table 1 An Overview of Characteristic Motions in Proteins[4]

Type of Motion	**Functionality**	**Time and amplitude**
Atomic fluctuation Side chain motion	Ligand docking flexibility, temporal diffusion pathways	Femtoseconds to picoseconds (10^{-15}-10^{-12} s); less than 1 Å
Loop motion Terminal-arm motion Rigid-body motion(helices)	Active site conformation adaptation, binding specificity	Nanoseconds to microseconds (10^{-9}-10^{6} s); 1-5 Å
Domain motion Subunit motion	Hinge-bending motion, allosteric transitions	Microseconds to milliseconds (10^{-6}-10^{-3} s); 5-10 Å
Helix-coil transition Subunit association	Hormone activation	Milliseconds to hours(10^{-3} – 10^{-4} s); more than 10 Å

1.2 ION CHANNEL PROTEINS

Ion channels play an essential role in cell membranes by being the key component of many fundamental physiological functions such as electrical signaling in between muscle and nerve cells, muscle contraction, auto-regulation of cell volume, loss of ionized waste products, release of pancreatic insulin, regulation of blood pressure, and memory acquisition[5]. A cell is protected from its environment with a lipid bilayer which also acts like an electrical insulator by creating an impermeable barrier to many molecules (cf. Figure 1) and all ions. Because of the amphiphilic nature of the ion channel proteins, they provide a hydrophilic surface for the diffusion of ions which play an essential role in intra/inter cellular signaling. In response to the changes in their environment, they are able to change their structural conformations from a closed to an open state in which ions are driven through them by an electrochemical gradient. The conformation change between closed and open states is called gating and major stimuli which induce this conformational change are a change in the membrane potential (voltage-gated), mechanical stretching of the membrane (mechanosensitive ion channels), and ligand binding to extracelluar or intracellular domains (ligand-gated)[5].

Dysfunction of ion channels is related to important diseases such as epilepsy, migraine, arrhythmias, schizophrenia, and some muscular disorders[6-8], thus they are very important drug targets for pharmaceutical studies.

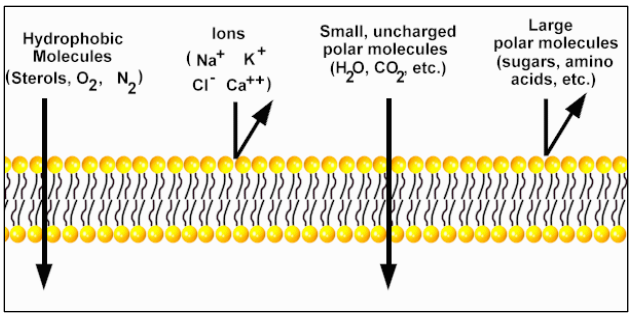

Figure 1. Diffusion of select molecules through phospholipid bilayer (http://saints.css.edu/bio/schroeder/carriers.html)

The first atomic structure of an ion channel was obtained by Arseniev's group in 1985 for Gramicidin A using solution state Nuclear Magnetic Resonance (NMR) [9]. Gramicidin A is a linear polypeptide consisting of 15 amino acids with alternating L- and D- configurations. The channel is formed via an amino-terminal to amino-terminal right-handed β-helix dimer via hydrogen bonding carbonyl, and amino groups six residues apart. Two monomers are held together by six hydrogen bonds and this configuration of the channel is selectively permeable to monovalent cations. The Gramicidin A dimer structure has been widely used as an ion channel model protein for computational studies, because it was the only available atomic resolution structure for ion channel proteins till 1998.

In general, when compared to the globular proteins, the present level of understanding of membrane proteins, and ion channels in particular, is more primitive. Intensive computational efforts together with advanced X-ray and NMR techniques [10-12] have been helpful in providing more detailed structural information of ion channel proteins. However, due to the difficulty of obtaining high resolution X-ray structures and the huge computational cost of full atomistic simulation techniques for such sized systems (lipid molecules + protein + water

molecules), there is a pressing need to develop coarse grained models for simulating dynamics of these systems. This is particularly the case when gating mechanisms are of interest. In addition to the system size, the time scale of gating makes full-atomistic simulation techniques infeasible.

1.3 GATING AND SIGNAL TRANSDUCTION IN NICOTINICOID RECEPTOR SUPERFAMILY

The members of the superfamily of nicotinicoid receptors belong to the neurotransmitter ligand-gated ion channel (LGIC) proteins which play an essential role in the propagation of electrical signals between cells at the chemical synapse. They are also referred to as the Cys-Loop superfamily of LGIC because of a conserved 15-amino acid-spaced disulfide loop in their extracellular domain. They are receptor-ion channel complexes with three domains: a large N-terminal hydrophilic extracelluar ligand binding site (ECD), a gated transmembrane pore (TMD), and smaller intracellular domain, forming an architecture around 160 Å in length along the direction perpendicular to the membrane. The members of this sequence related superfamily are gamma-aminobutyric acid type A (GABA-A), nicotinic acetylcholine, glycine, and the serotonin receptors, all of which are the targets of anesthetics and alcohols. The members of this superfamily have either homo- or hetero- pentameric structures. In the transmembrane segment, each subunit consists of 4 α-helices, M1-M4 (cf. Figure 2). Amphiphilic M2 helices from each subunit form the central pore of the channel, which is also the gate of the closed channel. The ligand binding sites are located at the interface between different subunits of the ECD and two to five ligand molecules can initiate the activation of the channel depending on hetero or homo pentameric configuration of the receptor [13]. The complete amino acid sequences of each of the

subunits of the nicotinicoid receptors are known, and there is high degree of sequence identity and similarity through different members of the family.

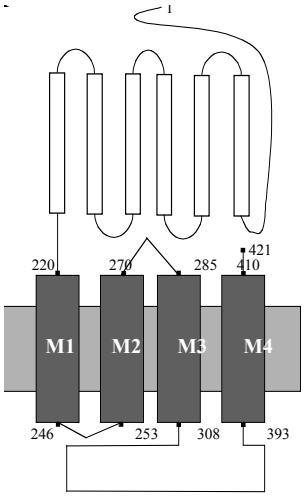

Figure 2. Nicotinicoid receptor structures. Schematic representation of the topology of the α1 subunit of GlyR [14].

Like the ion channel proteins in general, the major challenge in the field of ligand-gated ion channel studies is the lack of atomic resolution structures. Until relatively recently, most computational studies of nicotinicoid receptors have utilized the crystal structure of acetylcholine binding protein (AChBP) from Lymnaea stagnalis [15] which has around 20% amino acid sequence identity with ligand binding domain of the nicotinic acetylcholine receptor (nAchR) and the cryo-electron microscopy structure of the Torpedo receptor nAchR TMD at 4 Å resolution [16] as templates for building 3-D structures with comparative modeling. For GlyR, structures of the M2 helices[17] and of the M2-M3 bundle[18] have been obtained via NMR studies. In 2005, the structure of the entire nAchR in the closed-channel state, including the ligand-binding and intracellular domains, was resolved at 4 Å by cryo-electron microscopy by

Unwin *et al.* [19]. Even though the resolution of this structure is low, it provided essential information for modeling the regions connecting ECD and TMD of these receptors. Quite recently, in 2008 and 2009, three crystal structures of prokaryotic ligand gated ion channels have been obtained: two from bacterial Gloebacter violaceous (GLIC), one at 2.9 Å resolution [20] and another one at 3.1 Å resolution [21], plus one from bacterial Erwinia chrysanthemi (ELIC) at 3.3 Å resolution [22]. These are pentameric ligand gated ion channel homologues with ~20% amino acid sequence identity to the different members of the nicotinicoid superfamily. The ELIC structure, which presumably corresponds to a non-conducting state while GLIC structure with a 5 Å wide pore, was the first atomic-resolution open channel structure obtained. Although both ELIC and GLIC lack the signature conserved disulphide bonding in the ECD, the comparison of the open GLIC pore structure with the other, closed pore structures, will provide crucial information about the exact gate position and gating mechanism of the nicotinicoid receptor proteins.

Despite the difficulties due to the lack of even static pictures of different structural states of the LGIC receptors, many experimental and theoretical studies have been carried out successfully to understand crucial principles concerning how binding of neurotransmitter into the extracelluar domain induces large scale structural deformations in the transmembrane region which lets opening of the central pore. In the conformational pathway studies connecting the closed to open states of the receptor by using rate-equilibrium free-energy relationships, Auerbach *et al*. described the propagation of conformational changes through the receptor as a conformational wave. The mechanism suggested by these workers describes gating as the result of diffusion of a "defect", which is instigated in response to ligand binding to a well-defined binding pocket in the ECD, and through connecting loops ECD and TMD (this conformational

defect is distributed to the TMD) where a small change in the pore radius leads the passage of ions [23-25]. (That is, the term conformational defect is used for the initial disturbance to the equilibrium structure of the receptor which is created by the introduction of ligand into binding pocket.)

A number of more detailed structural models for the conformational changes leading to channel activation have also been proposed. In the cryo-electron microscopy studies of Unwin *et al.*, the proposed model for gating mechanism is the following: ligand binding induces conformational change which entails 15 º-16 º rotational movements of the α-subunit inner sheets relative to the outer sheets of the ECD domain. The rotational motion in the inner sheet of α-subunits is transmitted by their connected M2 helices to the hydrophobic girdle, thus weakening the hydrophobic interactions that hold the girdle together (cf. Figure 6 in ref.16).

In a 15 ns molecular dynamics simulation study of ECD+TMD of human α7 nAchR with explicit water and lipid bilayer molecules, Law and coworkers observed a twist to close motion in which movements of the C-loops in the ECD domain correlated with a 10 degree rotation and inward movement of two nonadjacent subunits in the TMD [26]. Interestingly, in another molecular dynamics simulation study of the same group, they modeled the receptor via recently available cryo-electron microscopy structure of Unwin *et al.* instead of using two separate structures for modeling ECD and TMD parts which is the case in their previous study, a lateral tilting of M2 helices initiating the channel to open has been observed [27].

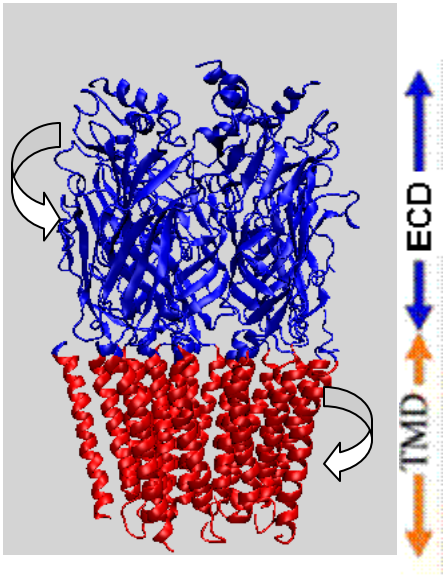

Figure 3. Quaternary Twist Model: Proposed gating model suggested for nAchR (The figure is created from pdb file 2BG9 [19] by using VMD software [98]).

Due to the large scale conformational changes and reorganization of different sub-structures within the molecule, the time scale of ligand binding is in the order of microseconds, whereas the transformation which is induced by that is in the time scale of milliseconds. Moreover, the size of the system is very large; specifically, there are ca. 1700 residues in the nAchR. Since all atom molecular dynamics studies when explicit water and lipid molecules are included for these large systems can generally be carried out to tens of nanoseconds (although a few all-atom MD simulations of large [$>10^4$] proteins propagated to ca.10 μs have been reported [29]), some other alternative analytical techniques have also been used to study the gating mechanism in nicotinicoid receptor proteins. In three of such studies that used normal mode analysis at various coarse-graining levels of description, two for nAchR [30, 31] and one for GlyR [28], the characteristic gating mechanism suggested is a "quaternary twist" (cf. Figure 3) model in which ECD rotates counterclockwise around the central axis and TMD rotates in the

opposite direction . This motion is observed in the single lowest frequency vibrational normal mode in all three studies. In the nAchR studies, an open channel state of the receptor has been proposed. In the study of Taly *et al*., out of 10 lowest frequency normal modes, only the first mode produced a structural reorganization compatible with a wide opening of the channel pore. In the study of Cheng *et al*., this twist motion of two domains with respect to each other was also observed in the lowest frequency normal mode; however they built the structural model for open pore as a linear combination of first and second lowest frequency modes (The main criteria for building an open state model was the observation of significant opening of the pore along the selected normal mode displacements without generating any steric clashes in the structure.) Even though there is still much controversy over this type of rotational motion for ligand-gated ion channel proteins, since substituted cysteine-accessibility (SCAM) experiments have shown that this type of rotational motion is disfavored for pore lining helices [32, 33], superposition of M1,M2, M3 helices in TMD and large portion of β-sandwich structure in ECD of putative open prokaryotic GLIC and the closed state ELIC structures displayed this type of quaternary twist motion of ECD with respect to TMD However, the lowest frequency normal mode is found to be not enough to explain the entire conformational transition, i.e. the structural displacement between the open and closed superimposed forms of GLIC and ELIC. The projection of the displacement vector between the open and closed superimposed forms of GLIC and ELIC onto each of the 100 lowest frequency normal modes shows that a basis set composed of the 100 lowest frequency normal modes only accounts for 50% of the whole transition [21] (Cf. reference [34] for further details of the projection technique).

1.4 LIMITATIONS OF CURRENT METHODS

1.4.1 Molecular Dynamics

Classical molecular dynamics is a highly useful simulation technique to investigate time evolution of the system which allows prediction of static and dynamic properties of substances. By numerically solving Newton's equations of motion for all atoms in the system, it propagates dynamics of molecules. However, due to the time scale of the motions present in the system, it requires a very fine time step. Ideally the time step in a molecular dynamics simulation should be 1/10 of the highest frequency vibration which is present in the system in order to satisfy the assumption that velocities and accelerations are constant during the time step taken. Since the temporal period corresponding to the highest frequency motion in biomolecules (X-H bond stretching) is ca. 10^{-14} seconds (X-H bond stretching), the proper time step should be on the order of 10^{-15} seconds. To be able to capture the motions on the time scale of milli or micro seconds (10^{-3} or 10^{-6}) with such a small time step is very challenging with today's computational power, especially for the case of membrane-bound proteins. In order to represent the system properly, the lipid bilayer molecules are needed to be included in addition to the solvent molecules. Consequently, the maximum realistic time scales for all-atom molecular dynamics simulations of such systems are typically just tens of ns [35].

Several techniques have been developed in order to overcome the problem of time step size in classical molecular dynamics simulations. One of those techniques, constraint dynamics, simply replaces high frequency vibrations by physical constraints. The SHAKE[36, 37] algorithm is representative of this type of dynamical constraint. In SHAKE, each constraint is imposed by adding an appropriate Lagrange multiplier term to the classical Lagrangian. The

Lagrange multiplier terms modify the Newtonian equations of motion in such as way as to constrain the desired bond lengths within a pre-set range (the smaller the tolerance in bond-length variations, the more iterations are needed to solve the constrained equations of motion). The additional work required to solve these constraint equations can significantly impede attempts to speed up molecular dynamics simulation methods [35].

Alternatively, instead of trying to increase the simulation time step by exclusion of calculation of high frequency motions present in the system, the dynamic process under the simulation study can be accelerated via introduction of a biasing external force in methods such as targeted MD[38] in which the force is used to drive the evolution of the simulation toward the targeted state, or steered molecular dynamics[39], in which the force is used to steer the system only along particular degrees of freedom. There are many other methods for accelerating the molecular dynamics studies. However, in most of these techniques the main assumption is that processes which are occurring on a long time scale are rare rather than slow: thus the aim is more towards sampling the rare events efficiently[40].

One other issue concerning all atom molecular dynamics studies is the inclusion of environment effects such as solvent (also lipid bilayer for membrane proteins). The simplest way to include solvent effects on dynamics of proteins is to explicitly introduce the solvent molecules into the simulation. However, simply adding 200 water molecules to the simulation box means additional 1800 degrees of freedom need to be simulated. An alternative way to incorporate solvent effects into MD-type simulations is to adopt a Langevin description [41] in which solvent water molecules are not explicitly included in the simulation, but rather their effect on protein is encompassed by friction and compensating random force terms that act upon the protein. Thus, the force on a particle arises from three sources. The first one is the standard systematic force

which is due to interaction between the particle and other particles. The second force on the particle is due to the motion of the particle through a viscous medium, i.e., the solvent: this retarding force is proportional to the particle's instantaneous velocity and a friction coefficient which depends on the intrinsic viscosity of the solvent and the particle's solvent accessibility. The final contribution to the force on the particle is due to random fluctuations, caused by interactions with solvent molecules, which obey a Gaussian distribution with zero mean value and delta-function two-time correlation function [42].

Incorporation of the Langevin scheme into molecular dynamics simulations is effective not only in terms of incorporating solvent effects or other environment effects such as those of the embedding lipid bilayer, but also it enables temperature control, since the prescribed system temperature is automatically maintained via a relationship, namely the Fluctuation-Dissipation Theorem [43], between the random force and the friction coefficient of the medium.

1.4.2 Vibrational Normal Modes and Related Techniques

The fact that some large-scale conformational changes are not accessible to standard molecular dynamics simulations makes vibrational normal mode analysis (NMA) potentially valuable in the study of biological systems, since NMA is not limited by the time scale of the process. NMA starts with calculation of the Hessian matrix which is comprised of the second derivatives of the potential energy matrix evaluated at the system's equilibrium configuration. From the Hessian, vibrational normal mode frequencies (obtained from the eigenvalues of the Hessian) and patterns of motion (given by the eigenvectors of the Hessian) can be extracted by standard matrix diagonalization. The major bottleneck is the difficulty in numerically evaluating the complete Hessian (which increases quadratically with system size) and diagonalizing it (direct matrix

diagonalization scales as N^3) [4, 44]. Special numerical methods such as the Lanczos algorithm can be employed to extract some of the lowest frequency normal modes by partially diagonalizing the Hessian matrix [45].

In a related but distinct method termed quasi-harmonic analysis, principal component analysis of the covariance matrix, which is extracted from very long MD trajectories on the full anharmonic potential surface, yields global modes of motion [46]. Berendsen and coworkers took one further step in this type of analysis by dividing up the configurational space into a very small size "essential subspace" (representing low frequency, global motions of the [protein] molecule) which is presumed to be enough to describe motions which are relevant for the function of protein, plus an irrelevant subspace which only includes local (high frequency) Gaussian fluctuations. In this method, an MD trajectory is run, and then from this MD data the covariance matrix of C-α atoms is constructed, i.e. $< \delta Q_i \delta Q_j >$, where δQ_i is the displacement of one cartesian component from the average equilibrium position of the i'th C-α atom. Diagonalization of this matrix yields the eigenvalues which are average square displacements. These are sorted in descending order. The essential subspace contains the modes corresponding to the k largest eigenvalues (with the cutoff determined by the condition that the sum of these k eigenvalues be a certain fraction of the sum of all the eigenvalues). This methodology [47], which is generally known as "essential dynamics", is able to identify slow, concerted motions of biological importance (e.g., hinge-bending [48]) from MD trajectories, but there are still some unresolved problems with it, such as the MD simulation length necessary to produce meaningful modes [49].

RTB normal modes analysis, developed by Sanejouand and collaborators, has been used as an alternative approach to normal mode analysis in order to determine low-frequency normal

modes of macromolecules[1]. In this method the molecule is partitioned into internally rigid fragments by using 3 translational and 3 rotational vectors for each rigid fragment. Thus, the number of degrees of freedom is significantly reduced in the hessian calculation since the internal vibrational motions within the rigid fragments are taken to be frozen out. Previous studies have been carried out in order to justify the RTB model by monitoring the agreement level between full all-atom normal modes and RTB-based normal modes [2, 44]. With the computationally improved implementation of the model by Li and Cui, the RTB normal modes method has been used to characterize large scale conformational changes in biomolecules by analyzing the collection of low frequency normal modes obtained via this method [2, 44, 50]. It has been shown that these studies RTB-based normal modes techniques can be used as an alternative approach to standard NMA if only low-frequency normal modes of the molecule are of interest, since it yields rather accurate results for low frequency modes of the system. In the RTB normal modes method, the interactions between atoms are calculated via an all atom force field. Thus, except excluding the internal degrees of freedom within rigid fragments there is no other simplification in the calculation of the interaction energies. In an alternative method called elastic network analysis[51], the interaction energy between particles (generally each residue is represented via C-α atoms) within a certain distance from each other is modeled as if they are connected via springs, and the spring force constant is taken to be uniform for all interactions. Unlike the RTB method, no detailed interaction potential energy function is employed in ENM models. The number of neighboring atoms, i.e. the structural organization of each C-alpha atom in protein's equilibrium structure, is the only factor effecting the calculation of interaction potential. This method is widely used to calculate only low frequency normal mode vectors of the protein systems for predicting conformational changes. Comparison of crystallographic B-

factors with the average root mean square fluctuation of C-α atoms which is calculated with ENM normal modes yields generally good agreement despite employing such a simplified model for representing the interaction potential energy in the system. However, the applications of this method are primarily restricted to guessing static pictures from normal mode eigenvectors [52-56], since the frequencies out of the analysis do not reflect the real frequencies of the system at all.

Even though the idea that a few low frequency normal modes are enough to capture a functionally relevant ("essential") subspace can be very useful in the study of long-time scale motions, it has not been widely used to coarse grain the propagation of time evolution of bio-molecular systems. Rabitz and coworkers developed a subspace integration method in order to study long time scale behavior of anharmonic crystals and glasses by solving Newtonian equations of motion in the reduced subspace which is spanned by calculating instantaneous low-frequency normal modes [57]. The major drawback of this method is the computational cost of calculating these low frequency normal modes "on the fly", i.e. at each instant of time calculation of the full Hessian matrix and partial diagonalization of this matrix [45] must be carried out.

1.5 OUTLINE OF THE DISSERTATION

In the RTB propagated dynamics which is presented in Section 2, we coarse grain the dynamics of the molecule by following RTB method developed for normal mode analysis for reducing the number of degrees of freedom. We do this by dividing the whole molecule into rigid bodies, or frozen blocks. In this manner, some local functionally irrelevant fluctuations of the molecule are

excluded, and the relative motion of these frozen blocks will help us to overcome the difficulty of simulating the "long-time scale" motion such as gating and signal transduction dynamics in ion channel proteins. A similar method called Multiple-Body O(N)[35] dynamics has been developed for propagating the dynamics of the molecule which is represented via rigid bodies and hinges connecting them. The orientation of each rigid body with respect to one another is described through series of vectors that span the hinge points between bodies, and through a corresponding series of angular coordinates that define the relative orientation of the bodies. These orientation vectors use Euler angles and spherical coordinates for defining the relative orientation of each rigid body. The rigid body representation scheme in RTB dynamics is much less complicated than the Multiple-BodyO(N) approach to partition molecule into rigid bodies. Namely, in the RTB partitioning scheme, there are several rigid blocks as well as a number of atoms which do not partition naturally into any rigid blocks. These latter atoms are treated explicitly, i.e., employing 3 Cartesian displacement coordinates per atom. By contrast, the motion of each internally rigid fragment is described by 6 Cartesian vectors, namely: 3 rotational displacement vectors plus 3 center of mass translational displacement vectors. Moreover, unlike the Multiple-BodyO(N) method, the RTB-dynamics partitioning scheme does not require explicit (and somewhat artificial) identification of hinge points between rigid blocks.

To test the RTB-dynamics method, gramicidin A embedded in a membrane mimetic is used. The type of motion which motivated us to develop these coarse-grained methods is the conformational wave motion suggested by Auerbach et al. in nAChR [23-25] , in which an initial disturbance to an equilibrium system from one end of the molecule is distributed through the molecule via loops and other interacting parts. Consequently, we developed a similar illustrative model utilizing gramicidin-A. In this toy model, initially gramicidin A dimer embedded in a

membrane mimetic is in an equilibrium state. Then a disturbance to the equilibrium system is introduced from one end of the molecule along the channel axis. This initial perturbation is employed by displacing the atoms of the first two residues from the top of the molecule from their equilibrium positions along the channel axis. The dynamics of the molecule propagated via classic all-atom Newtonian dynamics, RTB-dynamics, RTB normal modes, and ENM normal modes are monitored till the effect of initial disturbance reaches to the other end of the molecule along the channel axis. Only 8 indole groups are represented as rigid bodies in the gramicidin A dimer

In Section 3, the RTB method is used to freeze the local vibrations of internally rigid fragments in the propagation of Langevin dynamics of a molecule moving on a quadratic potential energy surface (PES). We term this method RTB-langevin dynamics: it represents a first step towards the future development of an algorithm for RTB-Langevin dynamics on general anharmonic PESs. Again, to test the validity of the method, gramicidin A dimer structure inside a membrane mimetic is studied, regarding the 8 indole groups as the only rigid bodies in the molecule. An initial perturbation is incorporated to the system by assigning an initial non-zero velocity to all the atoms of the third residue from the top of the molecule along the channel axis. The propagation is monitored via RTB-langevin dynamics on a quadratic PES and the all atom dynamics version until the effect of initial disturbance reaches to the end of the molecule or until the fluctuations of the molecule damp totally due to the friction imposed by solvent.

The concept of initial perturbation to an equilibrium system may be extended in general for any ligand binding process. In such processes the protein is in a certain equilibrium state with its environment, and introduction of a ligand molecule means a perturbation to this existing

equilibrium. The protein will go through conformational changes, in order to reach a new equilibrium with the existence of new forces due to ligand molecules. In this ligand induced mechanism , first proposed by Koshland [58] in 1958, the protein is a flexible molecule which will be able to adopt a new conformation when the ligand is bound to protein. In a recent study by Ikeguchi et. al.[59] ligand-induced binding model was employed in conjunction with classic linear response theory [60] in physics to ligand binding proteins. The structure of the ligand-bound conformation of the protein was predicted with the knowledge of only the ligand-free structure. Ligand-binding was modeled as an external perturbation force to the previously equilibrated protein system. In linear response theory, the core idea is that the equilibrium fluctuations of a system dictate the response to an external linear perturbation potential exerted onto the equilibrium system. By using the variance-covariance matrices calculated via a molecular dynamics trajectory or normal mode analysis in ligand free state of the protein, Ikeguchi et. al. effectively predicted the conformation of the ligand-bound form of three proteins by using the static LRT formalism. The perturbing force, representing ligand-protein interactions, was modeled by an attractive force vector between a selected atom in the ligand binding cite and the center of the ligand in the ligand-bound form of the protein [59].

As a generalization of the static LRT picture presented by Ikeguchi et al., we developed the dynamical analog of it (with the addition of a Langevin scheme to account for the effects of solvent) to monitor temporal relaxation of ligand-binding proteins going from ligand-bound to ligand-free state. These types of temporal relaxations can be monitored experimentally via spectroscopic techniques such as FRET[61] and ESR[62]. By tagging two sides of a protein with acceptor and donor fluorophores and by measuring the relative intensity of energy transfer in between them at different conformations of the protein, these techniques provide a

"spectroscopic ruler" for monitoring the conformational change of the solvated protein molecule in real time [63]. In section 4, a dynamic version of the LRT approach is developed to monitor the time scale required to go from ligand-bound to ligand-free state of two proteins by using the equilibrium time correlation functions on the final state PES. We used a harmonic oscillator model to describe the relevant PESs.

In order to test the dynamic LRT scheme effectively, we need the structure of ligand binding protein in each state, i.e. ligand-bound and ligand-free states. The proteins used for this study are both categorized as "hinge-bending" proteins with two domains due to the conformational change they go through upon ligand binding/unbinding events[64] The first example chosen for this study is the periplasmic ferric binding protein which has 309 residues; for this protein, both ligand-bound and ligand-free structures have been determined by X-ray crystallography [65, 66]. Upon binding of Fe+3 ions, the protein undergoes a conformational change of closure of two domains around ligand binding site. The second protein, T4 lysozyme is somewhat smaller globular protein with 164-residues. After binding to bacterial cell saccharide, it cleaves the glycosidic between the glycosidic bond between N-acetlymuramic and N-acetlyglucosamine. In the case of T4L, we utilized one of the X-ray structure of a mutant version of the protein as the open state [67] and the wild type X-ray structure for the liganded state [68]. T4 lysozyme is also composed of two domains connected by a long alpha-helix. As with the two domain closure observed in FBP, crystallographic studies of a T4L mutant in which a substrate is bound to the enzyme suggest that the substrate bound enzyme is locked in a state in which two domains have closed around the substrate binding site [69]. Finally, the time scales of relaxation upon ligand-dissociation for these two proteins calculated via dynamic LRT are compared to relevant experimental measurements.

2.0 A RIGID-BODY NEWTONIAN PROPAGATION SCHEME BASED ON INSTANTANEOUS DECOMPOSITION INTO ROTATION AND TRANSLATION BLOCKS

Essiz, S. and R.D. Coalson, J. of Chem. Phys., 2006. **124**: p:144116 .

2.1 ABSTRACT

The Rotation and Translation block (RTB) method of Sanejouand et al. [1, 2] provides an appealing way to calculate low frequency normal modes of large biomolecules by restricting the space of motions to exclude internal motions of pre-selected rigid fragments within the molecule. These fragments are modeled essentially as rigid bodies and the need to calculate high-frequency relative motions of the atoms that form them is obviated in a natural way. Here we extend the RTB approach into a method for computing the classical (Newtonian) dynamics of a biomolecule, or any large molecule, with effective rigid-body constraints applied to a pre-chosen set of internal molecular fragments. This method, to be termed RTB-dynamics, is easy to implement, conserves the total energy of the system, does not require the construction of the matrix of second spatial derivatives of the potential energy function (Hessian matrix), and can be used to compute the classical dynamics of a system moving in an arbitrary anharmonic force

field. An elementary numerical application to signal propagation in the small membrane-bound polypeptide gramicidin-A is presented for illustration purposes.

2.2 INTRODUCTION

Since the first attempts to perform Molecular Dynamics (MD) simulation on proteins [3] MD has developed into an exceedingly useful tool for elucidating structure/function relations in these and other bio-molecules. However, when explicit water solvent is added to the system (and in the case of membrane proteins, a lipid bilayer membrane as well), the time scale of all-atom MD simulations is restricted to 10's of nanoseconds -- sub-microsecond at most, even with the most powerful supercomputers currently available. This is, of course, quite short on biologically relevant time scales. To take one specific example, consider the case of membrane bound ion channel proteins [5]. Relevant dynamics of these molecules spans a wide range of time scales. The time scale of permeation of a single ion through such a channel is typically on the order of nanoseconds [70] thus MD treatment of ion permeation through channel proteins is at the edge of feasibility [71-73],while somewhat coarse grained Brownian Dynamics models are widely applicable. [74-81] However, processes such as the binding of a ligand to a Ligand Gated Ion Channel (LGIC) take on the order of microseconds, while channel gating (a deformation of the protein that opens or closes the pore to the passage of ions) typically takes milliseconds to seconds [13, 82-85], far beyond the all-atom MD time scale achievable with today's computers.

This situation motivates the development of coarse-graining computational strategies which leave out some, hopefully inconsequential, details of the overall atomic motions of the protein-environment system in order to simplify the system description so as to enable temporal

evolution to longer times scales [57, 86, 87]. In Normal Mode Analysis (NMA) [42, 88-90] a single evaluation of the 2nd derivative matrix at the equilibrium configuration of the protein followed by diagonalization of this ("Hessian") matrix generates a set of normal mode eigenvectors and eigenvalues, the former determining the vibrational pattern of each normal mode and the latter the corresponding frequency. From these one can synthesize the motion of the protein to arbitrarily long time scales. However, this type of analysis is intrinsically restricted to small amplitude motion around a single equilibrium reference configuration. Thus it is incapable of describing many interesting large scale conformational motions, such as the crossing of a transition state to go from reactant to product states of a chemical reaction. Nevertheless, the simplicity of harmonic oscillator models render them physically and mathematically appealing. Assuming that such a description is physically appropriate, there remains the difficulty of actually computing and manipulating the Hessian for large biomolecular systems. To simplify this aspect of the problem, approximations to the Hessian matrix have been introduced. As an extreme example, Elastic Network Models suppress all details of the underlying potential energy surface, except for the equilibrium configuration that the potential surface implies.[51, 52, 56] That is, all details of the chemical bonding and other intermolecular interactions are ignored, and subsumed by a network of springs connecting all pairs of essential atoms (which may be a subset of all the atoms in the systems, e.g., C-α atoms only). This network restores the system to a preset equilibrium configuration (generally taken from an X-ray crystal structure), while the values of the network spring constant and the cut-off distance beyond which atoms are presumed to be "unconnected" are taken as adjustable parameters. While this approach may describe the qualitative motion of large, tightly packed globular proteins [91], its applicability for describing critical motions of ion channels, particularly in the

transmembrane pore, where chemical detail is important to understanding the passage of ions through the narrow pore region, is as yet unclear.

The Rotation-Translation Block (RTB) method of Sanejouand et al.[1, 2] provides an alternative approach for reducing the complexity of the Hessian matrix required for normal modes calculations on protein molecules. In this approach, quasi-rigid fragments of the molecule are (pre)identified, and then displacement vectors that describe overall translation and rotation of each block are easily constructed. This generates a partial basis of displacement vectors for describing small vibrations about equilibrium. These displacements allow for the coupled motion of different blocks, but ignore small (high-frequency) vibrations within each rigid block. The full all-atom Hessian can then be projected in terms of this RTB basis, thus generating a reduced Hessian of much smaller dimension than the full version.

The goal of the present work is to extend some of the key ideas of the RTB normal modes procedure into a general Newtonian propagation scheme that uses the instantaneous RTB basis vectors to evolve the protein system in time so that the essential low-frequency large amplitude motions of the molecular are retained, but irrelevant high-frequency local vibrations are ignored. The method does not assume an underlying harmonic potential surface and thus in principle is applicable to complex reaction dynamics on anharmonic potential energy surfaces. The outline of this paper is as follows. The relevant theory is developed in Section 2.3. An illustrative example, namely, perturbation induced protein deformations (signal transduction) in an ion channel, is presented in Section 2.4, where we compare trajectory results obtained by integrating Newton's Equations for all atoms in the protein with the corresponding results obtained from RTB-dynamics and also via approximate normal modes analyses. As the degree of initial (and propagated) distortion increases, so do the effects of force field anharmonicity. RTB-dymamics

can account for these effects, whereas a normal modes description based on a single (equilibrium) reference configuration can not. Finally, Section 2.5 contains pertinent Discussion and Conclusions.

2.3 THEORY

2.3.1 Review of RTB decomposition

The motion of the atoms in a molecule can be described as Cartesian displacements around a reference configuration, in particular, the coordinates of the molecule at *t=0*. Given an initial configuration of the N atoms in the molecule, represented by the 3N-d Cartesian coordinate vector $\vec{x}_0$, we can write

$$\vec{x} = \vec{x}_0 + \delta\vec{x} = \vec{x}_0 + \boldsymbol{M}^{-1/2}\Delta\vec{x}\,. \tag{2.1}$$

Here $\delta\vec{x}$ is the 3N-d Cartesian displacement vector, and $\Delta\vec{x}$ is its mass-weighted equivalent ($\boldsymbol{M}$ being the 3Nx3N diagonal matrix whose non-zero elements contain the masses associated with each degree of freedom). We assume that we have identified groups of atoms that form rigid blocks (cf. Figure 4), and describe their displacements with particular linear combinations of the displacement vectors for those atoms. For concreteness, denote by N_j the number of atoms in block *j*. We construct six 3N-d vectors that describe rigid motions of the atoms of this block.

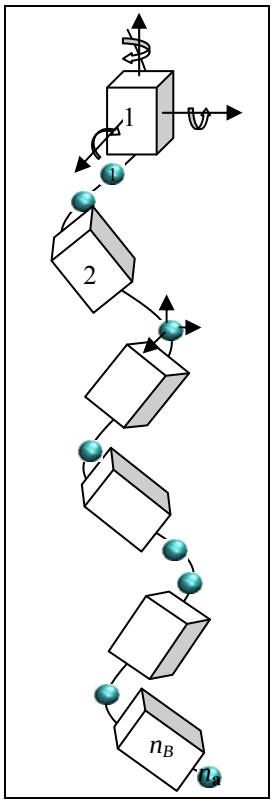

Figure 4. Schematic depiction of rigid blocks and independent atoms employed in the RTB decomposition procedure

Namely, there are three vectors that correspond to the overall translation of the center of mass of the block (the displacements of all atoms not in block *j* are set to *0*). Further, three vectors corresponding to overall rotations of the block are also easily constructed. All that is required is diagonalization of the 3x3 moment of inertia tensor for the block under consideration. This gives the 3 principal rotation axes, and a simple twist motion (linear transformation) about each of these axes generates the displacement corresponding to infinitesimal rotation about it. (Again, the displacements of all atoms not in the block under consideration are set to *0*.) A detailed description of the linear transformations underlying RTB decomposition can be found in [44, 88]. In mass weighted displacement coordinates, these 6 vectors are mutually orthogonal. This procedure is repeated for each rigid block. Any atom which is not part of a rigid block is represented simply by its 3 mass-weighted Cartesian displacements. If there are n_a independent atoms (not associated with rigid blocks), then the end result of this procedure is to generate $6n_B + 3n_a \equiv n_r$ mutually orthogonal "relevant" unit vectors in the 3N-d mass-weighted displacement space. This will in general be much fewer than the complete set of 3N mass weighted Cartesian displacement vectors, due to the fact that there are many atoms in each block, and displacements corresponding to internal vibrational motion within each block are suppressed. In the standard RTB normal modes procedure [1, 2], one considers the case of small vibrations around an equilibrium configuration $\vec{x}_0$, and takes matrix elements of the full 3Nx3N Hessian matrix (matrix of 2nd derivatives of the potential energy function) in the basis of the n_r vectors just constructed. This generates a reduced $n_r x n_r$ Hessian matrix, which in turn implies normal modes of vibration that correspond to relative motion of the rigid fragments (and the independent atoms). In the present development we follow a somewhat different route, exploring the possibility of computing Newtonian dynamics of these coupled rigid bodies

by propagating for a short time along the "instantaneous" relevant coordinate axes, then recalculating the RTB displacement axes based on the output configuration and repeating the procedure for many steps. Details of this scheme are developed in the following subsections.

2.3.2 Exact Propagation in local orthogonal RTB-based coordinates

Given the initial (*t=0*) coordinates of all the atoms, and the desired physical decomposition into rigid body fragments within the (N-atom) molecule, first switch to standard mass-weighted Cartesian displacements, $(\Delta x, \Delta y, \Delta z)$, for each atom. There are 3N such coordinates, denoted collectively as $\Delta\vec{x}$. Next, find the 3 overall translation displacement vectors (in the 3N-dim. space) and the 3 overall rotation displacement vectors for each of the n_B fragments using the RTB procedure: denote these unit-normed vectors as $\hat{u}^{(k)}$, $k = 1 - 6n_b$. Add to this set the $3n_a$ unit vectors corresponding to mass weighted Cartesian displacements of the n_a atoms which are not contained in any rigid block. Now, we need (formally) $3N - n_r$ other unit-normed vectors (all these vectors have length 3N), which are orthogonal to each other and to the n_r $\hat{u}^{(k)}$'s above. This can be done numerically, block by block, using standard linear algebraic projection techniques. [92] We now have a complete orthonormal set of vectors $\hat{u}^{(k)}$, $k = 1 - 3N$ that span the 3N-dimensional Cartesian vector space (hence the set of Cartesian displacements of all the atoms).

We can propagate classical mechanics (Newton's 2nd law) in the u-basis. We simply project all relevant 3N-d vectors, namely the system displacement vector, velocity vector, and force vector, onto the u-axes. This is easily achieved by taking the dot product with the appropriate $\hat{u}^{(k)}$. More specifically, let $\eta^{(k)}(t)$ be the displacement along unit vector $\hat{u}^{(k)}$. The

initial value of this coordinate is $\eta^{(k)}(0) = 0$ (by construction all displacements from the reference configuration are 0 at *t=0*). If $\dot{\vec{x}}_0$ is the initial velocity in the lab (or Cartesian displacement) frame, then $\vec{v}_0 = \boldsymbol{M}^{1/2}\dot{\vec{x}}_0$ is the velocity in the mass-weighted displacement coordinate system. We then project this 3N-d system velocity vector along the u-axes to obtain $\dot{\eta}^{(k)}(0) = \hat{u}^{(k)} \cdot \vec{v}_0$. Similarly, if the force on the atoms in the system in lab frame Cartesian space (again, 3N degrees of freedom in all) is $\vec{F}_0$, with the corresponding force in mass-weighted coordinates being $\boldsymbol{M}^{-1/2}\vec{F}_0 \equiv \vec{F}_0'$, then the force along the $\hat{u}^{(k)}$ axis is $F_u^{(k)} = \hat{u}^{(k)} \cdot \vec{F}_0'$ The equation of motion governing $\eta^{(k)}(t)$ is:

$$\ddot{\eta}^{(k)}(t) = F_u^{(k)}. \tag{2.2}$$

This is just Newton's 2nd Law in the u-axis coordinate system. Given the initial value of each $\eta^{(k)}$ and its initial time derivative (velocity), then the set of Equations 2.2 (one for each *k=1-3N*) can be integrated for any desired length of time. There is actually no need to re-expand around a different reference configuration after each time step; all that is needed is to be able to revaluate the force along the course of the trajectory (this is straightforward; see below). But, for the sake of developing a rigid-body dynamics propagation scheme, let us specify the following strategy for <u>exact</u> propagation (rigid body dynamics will emerge as an approximation to this scheme in Section. 2.3.3 below).

At *t=0*, $\vec{x}_0$ and $\dot{\vec{x}}_0$ are specified. We also assume that we have an explicit form for the Cartesian components of the force on each atom for any configuration of the system, i.e. $\vec{F}(\vec{x})$. Normally, we are provided with a potential energy function $V(\vec{x})$. Then, the y-component of the force on atom α is $-\partial V / \partial y_\alpha$,etc. Given $\vec{F}(\vec{x})$, the force in the mass-weighted coordinate

system $\vec{F}'$ can easily be obtained, as prescribed above, and then projected onto the u-axes as required for the present propagation scheme.

Based on the initial configuration $\vec{x}_0$, we find the $6n_B$ orthonormal RTB vectors plus the (trivially determined) $3n_a$ mass-weighted Cartesian displacement vectors for each independent atom in the molecule; then we construct $3N - n_r$ other orthonormal vectors, as described above. Now we have the complete set of basis vectors $\hat{u}^{(k)}$, $k = 1 - 3N$. Next, let $\mathbf{T}$ be the 3Nx3N matrix whose columns are the unit-normed $\hat{u}^{(k)}$ (expressed in the full 3N-d mass-weighted coordinate basis). Note the general linear transformations:

$$\vec{\eta}(t) = \mathbf{T}^{\mathbf{t}} \Delta\vec{x}(t)\,;\ \dot{\vec{\eta}}(t) = \mathbf{T}^{\mathbf{t}} \vec{v}(t)\,. \tag{2.3}$$

[$\mathbf{T}^{\mathbf{t}}$ is the transpose of the matrix $\mathbf{T}$; and $\mathbf{T}^{\mathbf{t}}\mathbf{T} = \mathbf{1}$.] The dynamical update procedure then goes as follows. By construction, $\Delta\vec{x}(0) = \vec{0}$; thus, from Equation 2.3, $\vec{\eta}(0) = \vec{0}$. Also, from Equation 2.3, $\dot{\vec{\eta}}(0) = \mathbf{T}^{\mathbf{t}} \vec{v}_0$. The components of the initial force along the u-axes are given similarly: $\vec{F}_{u,0} = \mathbf{T}^{\mathbf{t}} \vec{F}'(\vec{x}_0)$. (These conditions are consistent with the discussion centered around Equation 2.1 above.) For concreteness, we use the velocity Verlet scheme [42, 93] to dynamically update all 3N of the η-coordinates by a small time step δt :

$$\eta_f^{(k)} \equiv \eta^{(k)}(\delta t) = \dot{\eta}^{(k)}(0)\delta t + \tfrac{1}{2} F_{u,0}^{(k)} \delta t^2\,;\ k\text{=}1\text{-}3N\,. \tag{2.4}$$

This gives us the output coordinates. To obtain the output velocities according to the velocity Verlet algorithm, we also need to evaluate the force experienced at δt, that is, at the final configuration (in lab Cartesian coordinates), $\vec{x}_f = \vec{x}_0 + \boldsymbol{M}^{-1/2} \Delta\vec{x}_f$, where $\Delta\vec{x}_f$ is obtained from $\vec{\eta}_f$ via the backwards linear transformation implied in Equation 2.3. Thus, in the u-frame, $\vec{F}_{u,f} = \mathbf{T}^{\mathbf{t}} \vec{F}'(\vec{x}_f)$. Now:

$$\dot{\eta}_f^{(k)} \equiv \dot{\eta}^{(k)}(\delta t) = \dot{\eta}^{(k)}(0) + \tfrac{1}{2}[F_{u,0}^{(k)} + F_{u,f}^{(k)}]\delta t \quad ; \; k=1\text{-}3N \, . \tag{2.5}$$

The back linear transformation of these η-velocities to the corresponding velocities in mass-weighted lab Cartesian displacement coordinates is prescribed in Equation 2.3. (The Cartesian lab frame velocities are then obtained as $\vec{\dot{x}}_f = \boldsymbol{M}^{-1/2}\vec{v}_f$.) This completes one time step. To start the next step, regard the final lab Cartesian positions and velocities as the "initial" ones, and repeat (beginning with the recalculation of the RTB-based displacement coordinates).

For sufficiently small δt, this algorithm will clearly generate the exact Newtonian dynamics of the 3N-d (N-atom) molecule.

2.3.3 Approximate RTB-dynamics

In the space of small displacements around $\vec{x}_0$, linear combinations of the 6 RTB displacement vectors describe motions of the overall complex that keep the initial rigid fragments rigid. For block j, amplitude along any of the other $3N_j - 6$ local displacement vectors changes the distance between one or more pairs of atoms within this fragment, and hence corresponds to deviation from the rigid fragment model. Adding the full range of motions of the independent atoms to the relevant set of displacement vectors (bringing the total to n_r), let us organize the complete set of "rotated" basis vectors so that the n_r relevant ones are recorded first, i.e., vectors $n_r + 1, ..., 3N$ correspond to "irrelevant" internal vibrations within the various rigid blocks. Then, the basic idea is that when updating atomic displacements over a time step δt the output values of $\eta^{(k)}(\delta t)$ should be set to 0 for $k = n_r + 1, ..., 3N$, while updating the relevant coordinates in the fashion described above, i.e., by projecting the force vector, etc., along the appropriate u-axes. This

turns out to be part of the final prescription, but there is one additional subtlety that has to be dealt with, which concerns the manner in which the velocities are processed. The most naïve procedure would be to set $\dot{\eta}^{(k)} = 0$, $k = n_r + 1,...,3N$, identically. However, this would cause an unphysical "disruption" of the dynamics which is problematic (e.g., leads to energy non-conservation). At *t=0*, the Cartesian velocities of all the atoms are unrestricted. Thus when they are projected onto the u-axes, there will in general be non-zero components along the irrelevant u-axes, just as there will be along the relevant u-axes. Arbitrarily setting the velocities of the irrelevant η-coordinates to 0 actually violates the initial conditions, and leads to deviation in the propagated atomic trajectories over the course of many time steps from the exact Newtonian dynamics of the (rigid fragment) system. A more systematic approach is to add, at *t=0*, a specific additional force of constraint (which is constant over the propagation time interval δt). The effect of this force of constraint should be to cause the net displacement along all non-RTB u-coordinates to be zero at $t = \delta t$. The appropriate constraint force and its consequences for the irrelevant η-coordinate velocities turn out to relatively simple. Details are provided in Appendix A. Here we simply state the final computational prescription:

For a given initial lab Cartesian $\vec{x}_0$ and $\dot{\vec{x}}_0$,

1) Find the $6n_B$ RTB displacement vectors and add to these the $3n_a$ mass-displacement coordinates of the independent atoms: this brings the total of relevant basis vectors to n_r. (The other $3N - n_r$ internal displacement vectors never have to be constructed!) Project the initial mass-weighted velocity and force onto the relevant u-axes. Also calculate the "remainder" velocity $\vec{v}_\perp{}^{(0)}$, i.e., the component of the velocity perpendicular to the relevant displacement subspace,

$$\vec{v}_{\perp}^{(0)} = \vec{v}_0 - \sum_{k=1}^{n_r} \dot{\eta}^{(k)}(0)\hat{u}^{(k)} \ . \tag{2.6}$$

2) Dynamically update each of the n_r relevant coordinates and velocities by δt, using the velocity Verlet algorithm as described above.

3) The output mass-weighted Cartesian displacement coordinates are then given by

$$\Delta\vec{x}(\delta t) = \sum_{k=1}^{n_r} \eta^{(k)}(\delta t)\hat{u}^{(k)} \ , \tag{2.7}$$

and the output mass-weighted Cartesian velocities are given by

$$\vec{v}(\delta t) = \sum_{k=1}^{n_r} \dot{\eta}^{(k)}(\delta t)\hat{u}^{(k)} - \vec{v}_{\perp}^{\ (0)} . \tag{2.8}$$

Repeat these steps to achieve propagation over a macroscopic time interval. As shown in Appendix B, provided that the time step δt is small enough, this procedure will conserve energy.

The RTB-dynamics rigid body propagation scheme presented here bears some resemblance to the Multibody Order (N) Dynamics [MBO(N)D] scheme developed in Ref. [35]. Both methods produce approximate Newtonian dynamics of coupled rigid bodies which form a super-system of interest (e.g., a large protein molecule). However, there are significant differences in the details of implementation. The MBO(N)D scheme follows the motion of the pre-determined rigid bodies (coupled by hinges) using Euler angles, while our RTB-dynamics approximately freezes, over a short time step, the internal motion of pre-determined fragments within the overall molecule by suppressing modes of motion (the "irrelevant" degrees of freedom) that correspond to relative motion within each fragment. The RTB-method then has to re-characterize the rigid fragments after each time step by re-calculating the RTB coordinate unit vectors based on the updated configuration of the full molecule. It has the advantage that the

entire calculation is done in Cartesian coordinates, and thus the need to integrate Newton's (or Hamilton's) Equations [94] in non-Cartesian coordinates (e.g., Euler angles), is avoided.

2.4 ILLUSTRATIVE NUMERICAL EXAMPLE: APPLICATION TO SIGNAL PROPAGATION IN GRAMICIDIN

As a relatively simple but nontrivial application of the RTB-dynamics scheme to the process of signal propagation in an ion channel, we consider a gramicidin-A (GA) dimer, embedded in a membrane mimetic; cf. Figure 5. As indicated in the figure, the membrane model used here is simply a random array of 245 "pinning balls", i.e., Lennard-Jones centers of force that interact with the atoms of the GA polypeptide to hold it in place, in a membrane-spanning configuration. Similar minimalist mimetic models of the bilayer membrane have been used in studies of ion permeation through GA. [95] The NMR-derived GA structure of Arsen'ev et al. was used as a template. [96] This structure, embedded in the membrane mimetic described above, was subjected to energy minimization based on the Amber 94 force field [97] in order to determine the equilibrium configuration of the membrane-embedded GA dimer. (Lennard-Jones parameters $\varepsilon = 0.45$ *kcal/mol* and $\sigma = 1.78$ *Å* for the well depth and van der Waals diameter, respectively, were used to generate the pairwise interactions between each pinning ball and GA atom.) No water molecules were included in the calculation. The GA dimer protein was found to be mechanically stable in this environment, i.e., upon displacement of any atom from equilibrium, the system experienced a restoring force back towards the equilibrium configuration.

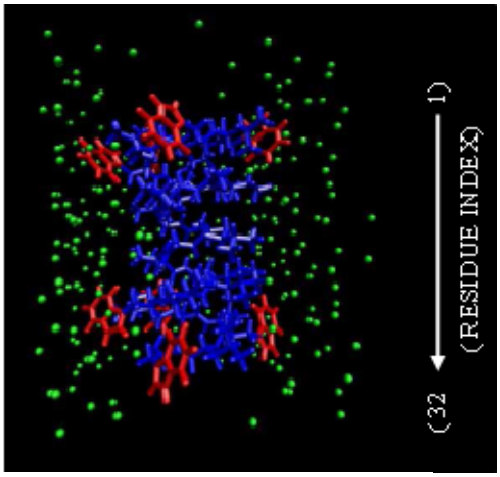

Figure 5. Equilibrium structure for a gramicidin-A dimer spanning an artificial lipid bilayer. The parts of the molecule highlighted in red are the indole groups which are treated as rigid bodies in the RTB rigid body decomposition procedure. Residue index runs from top to bottom, i.e., in the -z direction.

At *t=0*, the two residues at one end of the GA dimer were displaced from their equilibrium positions along the channel axis (all atoms of each amino acid [AA] residue being displaced by the same amount). The GA dimer was subsequently allowed to move according to Newtonian mechanics. The expectation was that the displacements would propagate from one end of the channel to the other, and this expectation was indeed borne out by direct integration of Newton's Equations for all atoms in the system, as depicted qualitatively in Figure 6. Furthermore, in addition to all-atom microcanonical (i.e., constant total energy) MD [42], calculations of the time evolution of this system were carried out using several approximate propagation schemes. Namely:

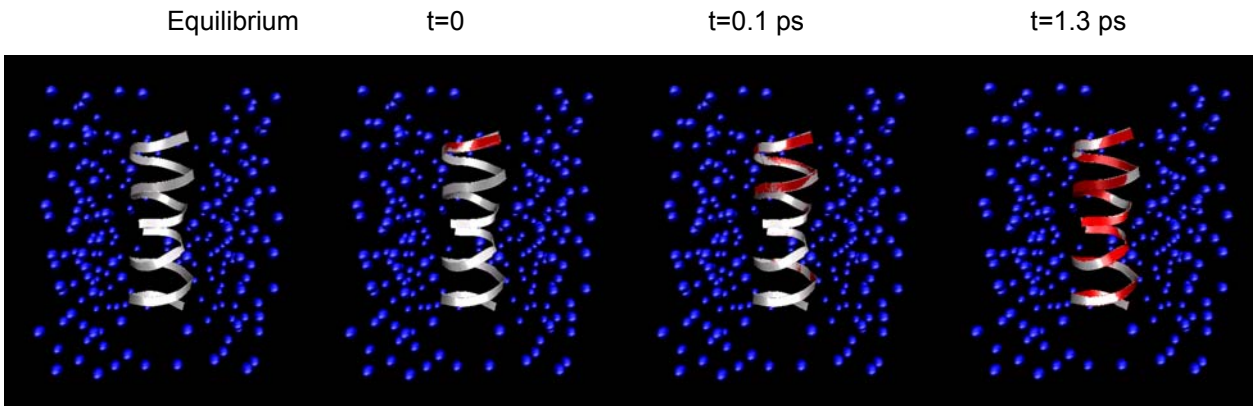

Figure 6. Schematic representation of signal propagation through the Gramicidin-A dimer. White ribbon depicts the equilibrium configuration of the gramicidin in the artificial lipid bilayer. At t=0, the two end residues at the "top" of the molecule are displaced downwards by 0.3 Å, as depicted in red. Time evolution is computed via classical mechanics. Red color indicates the protein structure as time progresses. The VMD [98] software was utilized to render these figures from MD data.

1) RTB-dynamics, the rigid body fragment propagation scheme developed in this paper. We treated each of the 8 indole groups (cf. Figure 5) in the dimer as a rigid fragment, and all other atoms as independent (each characterized by its 3 Cartesian components). Since each indole group really is, by the nature of its constitutive chemical bonds, rather rigid, this provides a good test of the performance of the RTB-dynamics method. That is, we expect that with this rigid body decomposition there should be good correspondence between all atom Newtonian dynamics and RTB-dynamics. Such behavior was in fact observed in our numerical studies, as described below.

2) RTB normal modes. To test the role played by potential surface anharmonicity on the motion of a mechanical deformation propagated through GA, we also studied the system's time evolution under an appropriate harmonic approximation. In particular, we calculated RTB normal modes [1, 2, 44] using the same indole rigid body decomposition employed in the RTB-dynamics scheme outlined above. Again, because the indole groups are rigid, the normal modes amplitude vectors and associated frequencies which arise from the RTB normal modes analysis (NMA) should accurately represent the low frequency normal modes of the full system (i.e., not assuming any partial rigid fragments). It is these modes which control large amplitude conformational changes in the molecule. Thus, if we superpose the initial displacements of interest to us here (see above) as a linear combination of RTB normal modes and then allow the system to propagate according to the usual tenets of normal modes theory [88, 90], we expect that the resultant dynamics will be essentially the same as would be obtained via full all-atom NMA. Further, we expect that for small initial displacements, this harmonic approximation should be accurate, i.e., give the same trajectory obtained by direct integration of Newton's equations (scheme 1 above). However, for sufficiently large displacements, the full Newtonian

dynamics of the GA system will be influenced by anharmonic terms in the potential energy surface (with respect to expansion about the configuration of mechanical equilibrium), and thus deviate from the predictions of the normal modes analysis inherent in the RTB normal modes model. In contrast, the RTB-dynamics method, which re-expands the RTB basis vectors that represent rigid body motions of the RTB fragments at each time step, does not suffer from this limitation. We expect that it can follow the true Newtonian dynamics of the system even for large amplitude motion. Indeed, we find this to be the case in the GA propagation example studied in this work.

3) ENM normal modes. In the spirit of the RTB normal modes propagation scheme 2, we considered the analogous procedure utilizing the Elastic Network Model framework. [51, 52, 56] The ENM is a harmonic oscillator model in which the underlying quadratic potential energy surface is assumed to arise from a network of springs which couple the motion of all pairs of atoms within a specified cutoff separation. This potential energy surface is *ad-hoc* except for its incorporation of an assumed mechanical equilibrium configuration – in our case, this would be the equilibrium configuration of the GA used in all four propagation schemes. In particular, the values of the cutoff separation and the spring force constant (which is assumed to be the same for all pairs of connected atoms in the system) are arbitrary. In the implementation presented here, we assumed a cutoff distance of 5 Å. Furthermore, to incorporate the confining effect of the pinning balls in our membrane mimetic on the motion of the GA protein, we included spring contacts between all pairs of pinning balls and GA atoms within the specified cutoff distance. (The pinning balls are immobile, but these spring contacts modify the Hessian which determines the normal modes of vibration of the GA atoms within the ENM model). Finally, we fix the value of the force constant in the ENM model by noting that since there is only one such

constant, the frequencies of all the ENM normal modes scale directly as the square root of this force constant, and thus by varying it we can adjust the overall time scale of the motion implied by ENM. This adjustment was done in order to make the time scale of the propagation of displacement amplitude from one side of the GA system to the other resemble the time scale obtained by full all-atom MD as closely as possible.

In our numerical studies we considered a range of initial displacements of the two end AA's (as described above). Figure 7 illustrates the behavior found in the small displacement limit. In particular, in this figure the end two AA's of GA are initially displaced by -0.05 Å (corresponding to a slight compression of the protein, qualitatively similar to the picture depicted in Figure 6**)**, and then the subsequent displacement of each C-α atom along the channel axis is monitored at several times. Propagation of the initial displacement from one side of the molecule to the other, followed by reflection back into the center of the system, is clearly apparent. It is also apparent that all-atom MD, RTB-dynamics and RTB normal modes are all substantially in agreement, while the ENM-dynamics produces noticeably different results. It is interesting to note that RTB-dynamics and RTB normal modes are in nearly perfect agreement, and deviate by a tiny amount from the all atom MD trajectory. This is actually to be expected, since in the limit of small displacements the Newtonian equations of motion become the same for RTB-dynamics and RTB normal modes schemes. (The "difference" in the two methods is simply in the way these equations of motion are integrated.) They deviate, formally, from the equations of motion implied by all-atom Newtonian mechanics in the treatment of the motion of the rigid fragments assumed in the RTB-dynamics scheme. (In the RTB-dynamics this motion is frozen out, while it is allowed to evolve fully in all-atom MD.) However, since the indole fragments are in fact quite rigid, the errors incurred in freezing out their internal motion are minor.

It is interesting to contrast the behavior just described with what transpires when the initial displacement of the end two AA's of the GA system is increased to -0.3 Å. In Figure 8, the time evolution of displacements of the backbone C-α's along the propagation axis is shown for all atom MD, RTB normal modes and ENM normal modes. Of course, any motion synthesized within the framework of single reference vibrational normal modes theory scales proportionally to the overall initial displacement of the system. That is, the results for the RTB normal modes and ENM normal modes schemes are identical except for an obvious scale factor in the displacement amplitude. However, full MD deviates from the corresponding RTB-normal mode predictions by a noticeable amount, particularly as time progresses. Again, this is a consequence of the fact that the atoms of the GA protein experience contributions to the full (AMBER) potential energy surface corresponding to cubic and higher terms in a Taylor series expansion about the configuration of mechanical equilibrium. In contrast, Figure 9 compares the behavior of all-atom MD vs. RTB-dynamics. RTB-dynamics can faithfully follow the full range of anharmonic motion, as explained above. (The small deviations between all-atom MD and RTB-dynamics can be attributed to errors made in the RTB-dynamics by assuming that the chosen rigid fragments are in fact completely rigid.)

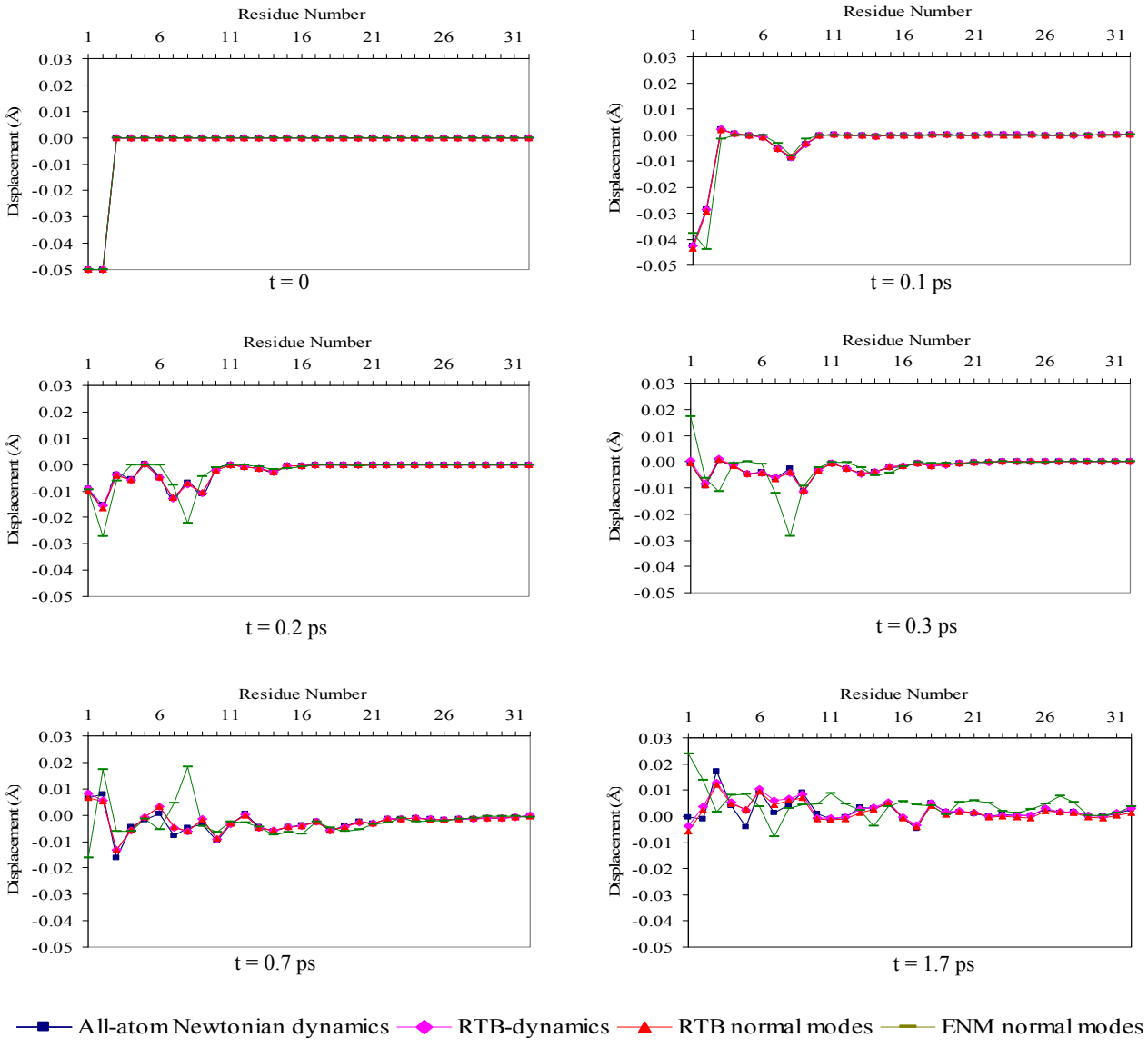

Figure 7. Signal propagation through the GA dimer. In each panel, the displacement from mechanical equilibrium along the pore axis of each C-α in the GA sequence is shown. At t=0, the first two residues from the top of the molecule are displaced from their equilibrium structure along the pore axis by -0.05 Å as indicated. The displacement of each C-α from the equilibrium structure is monitored until the effect of initial displacement reaches the other end of the molecule.

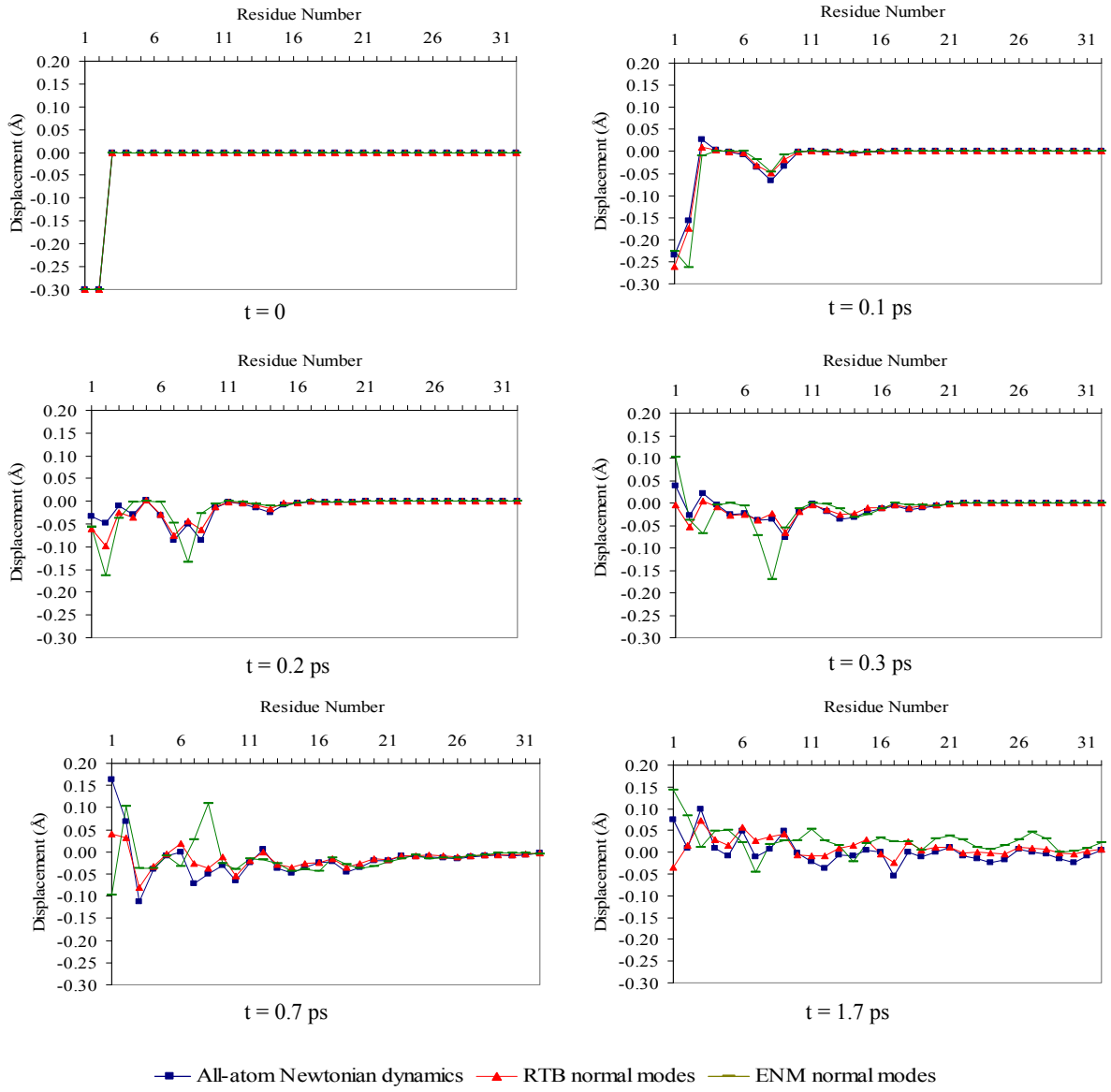

Figure 8. Signal propagation through the GA dimer. In each panel, the displacement from mechanical equilibrium along the pore axis of each C-α in the GA sequence is shown. At t=0, the first two residues from the top of the molecule are displaced from their equilibrium structure along the pore axis by-0.3 Å as indicated. The displacement of each C-α from the equilibrium structure is monitored until the effect of initial displacement reaches the other end of the molecule.

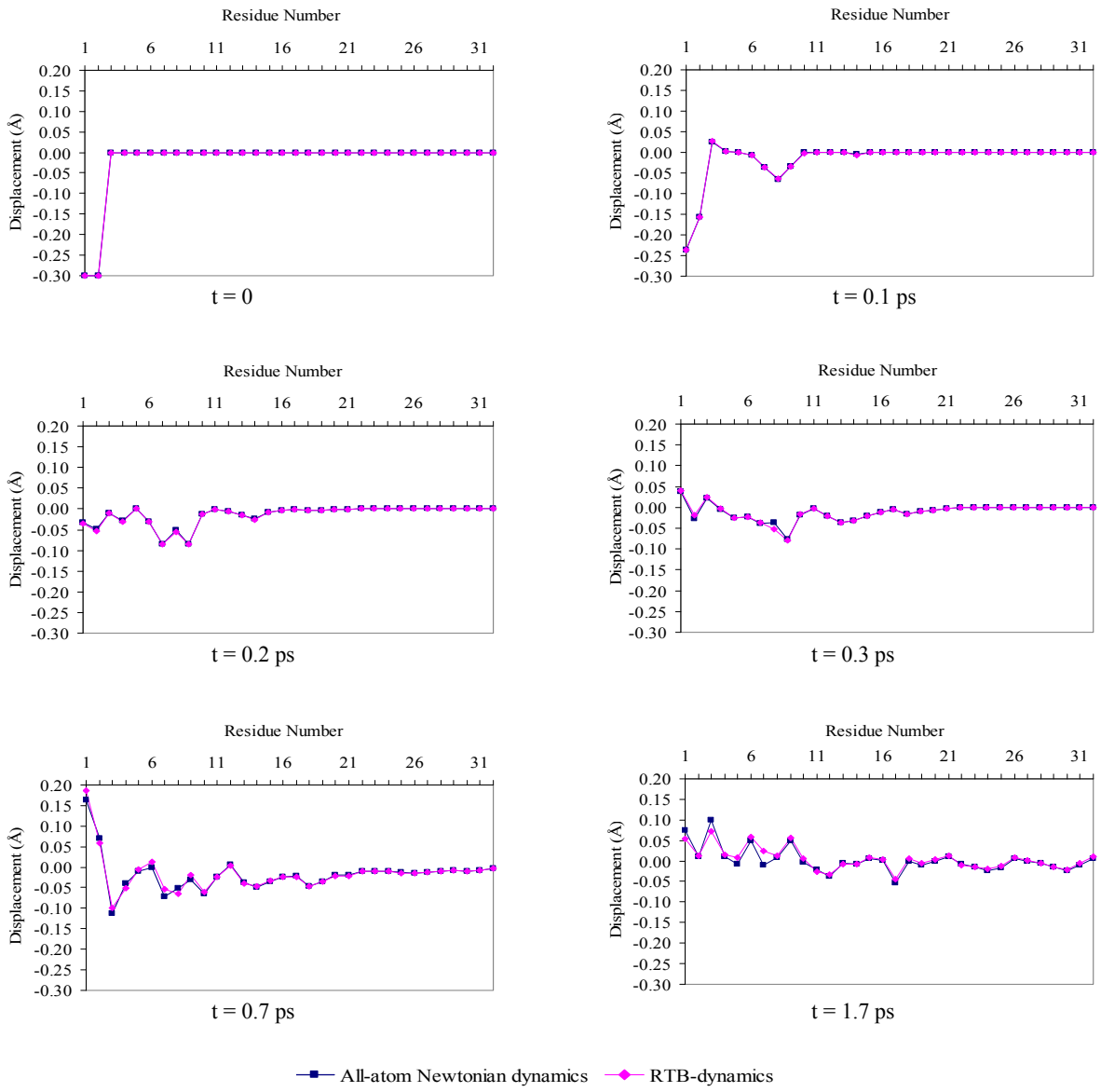

Figure 9. Signal propagation through the GA dimer. In each panel, the displacement from mechanical equilibrium along the pore axis of each C-α in the GA sequence is shown. At t=0, the first two residues from the top of the molecule are displaced from their equilibrium structure along the pore axis by -0.3 Å as indicated. The displacement of each C-α from the equilibrium structure is monitored until the effect of initial displacement reaches the other end of the molecule.

2.5 DISCUSSION AND CONCLUSIONS

The RTB-dynamics treatment presented here is rudimentary in certain aspects. To develop it into an MD technique for sampling from a canonical (finite temperature) ensemble, the classical dynamics should be thermostated so that a constant temperature can be automatically maintained. [42] Development of a barostat would be another useful extension. Further, it should be possible to adapt the ideas underlying the RTB-dynamics algorithm to a Langevin dynamics analog by replacing the short-time Newtonian (Verlet) update step with an analogous Brownian dynamics (drift-diffusion) or Langevin dynamics update. [42, 99]

A major issue that was not dealt with in this paper is how to choose the rigid fragments that are essential to the RTB strategy (either normal modes or full Newtonian dynamics). We illustrated our RTB-dynamics scheme with an example where obvious rigid fragments could be identified. In the general case this identification *a priori* (i.e., in the absence of performing all-atom MD simulations on the system) is difficult. Nevertheless, certain methodologies offer some promise of being able to identify rigid (vs. flexible) parts of a protein with far less computational effort than full MD would entail. These methods include the FIRST algorithm of Thorpe et al. [100], and the Elastic Network model introduced by Tirion [51] and recently utilized by many workers[52, 54, 101-103]. Both methods require only a reference structure – typically, the equilibrium structure of the protein, but do not attempt to generate or input any information about the all-atom potential energy surface, to run all-atom MD, etc. Thus they may be able to provide a rapid and qualitatively accurate zeroth order assessment of rigid parts of the protein, which can then be uploaded into the RTB-dynamics procedure. (In cases where the RTB-dynamics procedure is used to follow large amplitude motion on an anharmonic potential surface, it may be appropriate to repeat these rigidity analyses every few time steps, to obtain a

time-dependent rigid fragment decomposition that reflects the global course of the dynamics.) The incorporation of an effective (time-dependent) rigidity analysis of this type into the RTB-dynamics algorithm would boost its utility considerably.

2.6 ACKNOWLEDGMENTS

This work was supported in part by grants from NSF and ARO.

3.0 LANGEVIN DYNAMICS OF MOLECULES WITH INTERNAL RIGID FRAGMENTS IN THE HARMONIC REGIME

Essiz, S.G. and R.D. Coalson, J. of Chem. Phys., 2007. **127**: p. 104109

3.1 ABSTRACT

An approximation scheme is developed to compute Brownian motion according to the Langevin Equation for a molecular system moving in a harmonic force field (corresponding to a quadratic potential energy surface) and characterized by one or more rigid internal fragments. This scheme, which relies on elements of the Rotation Translation Block (RTB) method for computing vibrational normal modes of large molecules developed by Sanejouand and coworkers, provides a natural and efficient way to freeze out the small amplitude, high frequency motions within each rigid fragment. The number of dynamical degrees of freedom in the problem is thereby reduced, often dramatically. To illustrate the method, the relaxation kinetics of the small membrane-bound ion channel protein gramicidin-A, subjected to an externally imposed impulse, is computed. The results obtained from all-atom dynamics are compared to those obtained using the RTB-Langevin Dynamics approximation (treating 8 indole moieties as internal rigid fragments): good agreement between the two treatments is found.

3.2 INTRODUCTION

There is at present considerable interest in methods for computing large scale protein dynamics, including conformational changes which relate structure to the function of these molecules. Significant difficulties arise due to the long time scales (milliseconds and longer) on which functionally important conformational changes occur. Some progress can be made by recognizing that many large scale conformational changes are determined by collective motions in which internal quasi-rigid fragments (e.g., domains, or elements of the protein's secondary structure) move with respect to each other -- relatively slowly, since high frequency local atomic fluctuations corresponding to local vibrational motions of atoms within a fragment are naturally frozen out. Progress in numerical methods for rigid fragment Newtonian propagation has been made at the levels of both normal modes vibrational analysis [1, 2], which presumes that motion takes place on a a quadratic potential energy surface (PES), and fully anharmonic molecular (Newtonian) dynamics [35, 104] based on an arbitrary PES.

The time scale on which rigid-fragment molecular dynamics (MD) can be carried out is still limited. All-atom MD simulations on solvated proteins can typically be carried out for several nanoseconds on extant computer platforms. Much of the effort goes into the explicit inclusion of water molecules (as many as tens of thousands of these in modern protein dynamics simulations). Even a factor of 10 increase in time step which might come about by eliminating the need to accurately evolve high frequency local motions in the protein molecule would not allow one to access biologically relevant time scales of milliseconds and longer. One way to address these difficulties is to adopt a Langevin description [41], in which some atoms in the protein molecule are subjected to a frictional damping force proportional to their velocity, counterbalanced by a random force chosen so as to drive the entire protein to thermal

equilibrium at long times [105]. In this approach, solvent water molecules are not explicitly included. Rather, their effect on the protein motion is encompassed by the friction and random force terms that act upon the protein. Advantages of the Langevin Equation framework include: 1) a longer time step can be taken (the more so as the frictional damping becomes larger), 2) the natural incorporation of temperature (as embodied, for example, in the long-time relaxation to the Boltzmann distribution of atomic positions and momenta), and 3) inclusion of effects due to coupling with the environment (e.g., water solvent, and in the case of membrane bound proteins, a lipid bilayer membrane as well), which will not only modify the time scales of relaxation to thermal equilibrium, but also the *modes* of relaxation (e.g., by putting frictional contacts only on those atoms of the protein which are exposed to water, the dynamics of these parts of the protein will be preferentially and sometimes significantly modified).

Given the success of Newtonian dynamics schemes that exploit the partial rigid body decomposition noted above, it is natural to enquire whether these can be generalized to a Langevin dynamics scheme. Here we take a first step in this direction. Namely, we devise an algorithm for carrying out rigid-body Langevin dynamics when the PES governing the molecular motion is quadratic (i.e., contains no terms cubic or higher in the system coordinates). The theory is then similar in spirit to that of vibrational normal modes[1, 2], having the same advantages and disadvantages. Its advantages include an essentially analytic solution of the relevant equations of motion for any number of coupled coordinates ("degrees of freedom"), enabling the trajectory of the molecule to be computed out to infinitely long times. In practice, if the molecular system has D degrees of freedom, solution of the Langevin Equations of Motion (EOMs) involves linear algebraic manipulations in a $2D$ dimensional vector space: thus, a computer is needed for numerical solution in most cases. Still, the existence of a closed set of linear equations that

determines exact time evolution for all times is highly advantageous for obvious reasons: for example, thermally averaged correlation functions between various degrees of freedom can easily be extracted from solutions of the fundamental EOMs [105]. Of course, there are a number of inherent disadvantages of the harmonic oscillator model. All effects associated with inherent *an*harmonicities in the PES are missed (for example, barrier crossing from one quadratic basin of attraction to another). Also, determining the appropriate frictional damping coefficient for each degree of freedom is a nontrivial enterprise. A full reckoning would entail all-atom MD of the entire system (protein and environmental degrees of freedom), which would be computationally demanding, and in many cases intractable. There are useful shortcuts that can be employed, e.g., a molecular hydrodynamics approximation in which the friction coefficient associated with a given atom depends in a simple manner on its effective radius and the viscosity of the solvent that surrounds it [105-107], but the validity of these approximations is by no means incontrovertible.

The remainder of the paper is outlined as follows. In Section 3.3, we briefly review time evolution of a collection of harmonic oscillators under the Langevin Equation, which we shall also refer to as Langevin Dynamics, since it serves as the starting point for the development of an approximate solution to the same set of stochastic equations that incorporates pre-selected subsets of the particles (atoms) as rigid bodies. The rigid-body decomposition follows a strategy previously developed for analysis of vibrational normal modes of large protein molecules, known as the Rotation Translation Block (RTB) method [1, 2]. To provide explicit details of our partial rigid body Langevin dynamics procedure, we need to review the basic principles of RTB decomposition of the full space of D atomic displacements into a a set of n_R relevant orthogonal displacements (largely collective in nature), which span a subspace of the full D

dimensional displacement space, and a complementary set of $D - n_R$ irrelevant orthogonal coordinates. This is done in Section 3.4. Given an explicit specification of the relevant coordinates, the RTB-Langevin approximation is prescribed as the equations of motion of an n_R dimensional harmonic oscillator system characterized by $n_R \times n_R$ dimensional force constant and friction matrices which are obtained by projecting the full $D \times D$ dim. (all-atom) versions of these matrices onto the subspace of relevant displacements. (In the most frequently encountered case of unrestricted 3-dim. molecular motion, $D = 3N$, where N is the number of atoms in the molecule.) The number of irrelevant displacements is typically much larger than the number of relevant displacements, particularly if large quasi-rigid internal fragments can be identified. Thus the RTB-Langevin approximation will generally result in a significant reduction in the dimensionality of the dynamical problem that needs to be solved. The RTB-Langevin prescription is also provided in Section 3.4. To fully justify this prescription, we derive it in Section 3.5 as a mode decoupling approximation to the full dynamics, valid in the limit where the vibrational frequencies of the motions that are internal to a given quasi-rigid block are large compared to those which couple the collective motions of two or more such blocks. In Section 3.6, we illustrate the principles of RTB-Langevin dynamics for a prototypical two degree of freedom system, corresponding to the collinear motion of two masses (atoms) attached to an appropriate set of Hooke's Law springs. This simple system allows us to illustrate the basic principles of RTB-Langevin dynamics in transparent detail. To directly demonstrate the applicability of the method to protein systems, we present in Section 3.7 a more elaborate numerical example. There the gramicidin-A ion channel is embedded in a lipid membrane mimetic in a membrane-spanning orientation, and we follow the dynamical response of the

molecule to an external impulse applied at one end of the molecule at $t = 0$. Finally, Discussion and Conclusions are presented in Section 3.8.

3.3 REVIEW OF THEORY OF LANGEVIN DYNAMICS UNDER THE INFLUENCE OF A GENERAL MULTIDIMENSIONAL HARMONIC OSCILLATOR FORCE FIELD

This section closely follows the excellent exposition of Lamm and Szabo [105]. Consider motion in D Cartesian dimensions on a general quadratic potential energy surface (PES). Denote the coordinates as q_i, $i = 1,2...D$, and assemble these into the coordinate vector $\vec{q} = (q_1,...,q_D)^T$. Then the relevant PES is

$$V(\vec{q}) = \frac{1}{2}\vec{q}^T \mathbf{V}'' \vec{q}\,, \tag{3.1}$$

where $\mathbf{V}''$ is the relevant $D \times D$ force constant matrix. In the case of an N-atom molecule moving in 3-dim., each q_i represents the displacement from mechanical equilibrium along one of 3 cartesian dimensions, and hence $D = 3N$, as noted above.

The most general Langevin Equation consistent with this potential energy surface is

$$m_i \ddot{q}_i(t) = -\sum_{j=1}^{D} \gamma_{i,j} \dot{q}_j(t) - \sum_{j=1}^{D} V''_{i,j} q_j(t) + r_i(t)\,;\, i = 1 - D\,. \tag{3.2}$$

In Equation 3.2, m_i is the mass associated with coordinate i. The velocity dependent frictional damping (1st term on the r.h.s.) is, in the most general case, a linear combination of the velocities of all the coordinates, with appropriate superposition constants $\gamma_{i,j}$, i.e., friction coefficients. The

friction matrix $\gamma_{i,j}$ is always symmetric. Furthermore, the $r_i(t)$ in Equation 3.2 are Gaussian random force terms defined by the behavior of their expectation values, namely, $< r_i(t) >= 0$, $i = 1 - D$, and their two-time correlation functions, namely, $< r_i(t_1) r_j(t_2) >= 2k_B T \gamma_{i,j} \delta(t_1 - t_2)$, $i, j = 1 - D$, with k_B being Boltzmann's constant and T the absolute temperature. These properties of the random force components guarantee that the velocity distribution will approach Maxwell-Boltzmann form at long times, corresponding to thermal equilibrium.

It is useful to transform to mass-weighted coordinates, thereby removing the particle masses from the equations of motion. Defining $x_i(t) = \sqrt{m_i} q_i(t)$, we obtain the analog of Equation 3.2, which can be conveniently written in matrix-vector notation as

$$\ddot{\vec{x}}(t) = -\mathbf{\Gamma} \dot{\vec{x}}(t) - \mathbf{F} \vec{x}(t) + \vec{R}(t), \tag{3.3}$$

with the renormalized friction and force constant matrices $\Gamma_{i,j} = \gamma_{i,j} / \sqrt{m_i m_j}$ and $F_{i,j} = V''_{i,j} / \sqrt{m_i m_j}$, respectively, and the Gaussian random noise functions $R_i(t)$ characterized by $< R_i(t) >= 0$, $i = 1 - D$ and $< R_i(t_1) R_j(t_2) >= 2k_B T \Gamma_{i,j} \delta(t_1 - t_2)$, $i, j = 1 - D$. We shall regard Equation 3.3 as the fundamental vector EOM that we seek to solve, both exactly and approximately (via the RTB-Langevin prescription developed below).

The set of coupled Langevin Equations 3.3 admits stochastic solutions which are statistically equivalent to the phase space probability distribution associated with an appropriate Fokker-Planck Equation (see Section II of Lamm and Szabo [105]). If we designate the momentum conjugate to displacement coordinate x_i as $p_i \equiv \dot{x}_i$, and the 2D vector containing all phase space components as $\vec{X} = (x_1, ..., x_D, p_1, ..., p_D)^T$, then the phase-space probability density $P(\vec{X}, t)$ evolves according to

$$\frac{\partial P}{\partial t} = \sum_{j=1}^{D}\sum_{k=1}^{D} x_j F_{j,k} \frac{\partial P}{\partial p_k} - \sum_{j=1}^{D} p_j \frac{\partial P}{\partial x_j} + \sum_{j=1}^{D}\sum_{k=1}^{D} \Gamma_{j,k} \frac{\partial}{\partial p_j}(P p_k) + k_B T \sum_{j=1}^{D}\sum_{k=1}^{D} \Gamma_{j,k} \frac{\partial^2 P}{\partial p_j \partial p_k}. \qquad (3.4)$$

In particular, an initially Gaussian phase space distribution *remains* Gaussian for all times, namely,

$$P(\vec{X},t) = \frac{1}{(2\pi)^D \det[\mathbf{\Sigma}(t)]^{1/2}} \exp\{-\frac{1}{2}(\vec{X} - \vec{X}_t)^T \mathbf{\Sigma}^{-1}(t)(\vec{X} - \vec{X}_t)\}. \qquad (3.5)$$

In Equation 3.5, $\vec{X}_t \equiv (x_{1,t},...,x_{D,t}, p_{1,t},..., p_{D,t})^T$ is a 2D-dimensional vector of time-evolving parameters which specify the expectation values of position and momentum at time t. Furthermore, $\mathbf{\Sigma}(t)$ is a $2D \times 2D$ matrix of time-evolving parameters which determines the variance of the phase space probability distribution at a given instant of time, i.e., $\Sigma_{j,k}(t) = < (X_j - X_{j,t})(X_k - X_{k,t}) >$ for $j,k = 1-2D$, with the average being taken over the instantaneous probability distribution $P(\vec{X},t)$ given in Equation 3.5.

Substitution of the ansatz for the phase space probability density given in Equation 3.5 into the Fokker-Planck Equation 3.4 generates the following time evolution for X_t:

$$\dot{\vec{X}}_t = \mathbf{A}\vec{X}_t, \qquad (3.6)$$

where $\mathbf{A}$ is the $2D \times 2D$ matrix,

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{F} & -\mathbf{\Gamma} \end{pmatrix}. \qquad (3.7)$$

Equation 3.6 is formally solved as $\vec{X}_t = \exp(\mathbf{A}t)\vec{X}_0$, with $\vec{X}_0$ prescribing the desired initial values. In practice, the $2D \times 2D$ propagator matrix $\exp(\mathbf{A}t)$ can be synthesized from the eigenvalues and eigenvectors of $\mathbf{A}$, which can be computed easily via standard applied math

subroutines [108, 109]. The same substitution which generates Equation 3.6 generates an EOM for the spread matrix, namely,

$$\dot{\mathbf{\Sigma}}(t) = \mathbf{A}\mathbf{\Sigma}(t) + \mathbf{\Sigma}(t)\mathbf{A}^T + 2\mathbf{B}, \tag{3.8}$$

where $\mathbf{B}$ is the $2D \times 2D$ matrix given by

$$\mathbf{B} = \beta^{-1}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} \end{pmatrix}, \tag{3.9}$$

with $\beta = (k_B T)^{-1}$. The long-time value $\mathbf{\Sigma}(\infty)$ is unique, namely,

$$\mathbf{\Sigma}(\infty) = \beta^{-1}\begin{pmatrix} \mathbf{F}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}. \tag{3.10}$$

For an initial value matrix $\mathbf{\Sigma}(0)$, the time evolution implied by Equation 3.8 is formally obtained as

$$\mathbf{\Sigma}(t) = \mathbf{\Sigma}(\infty) - \exp(\mathbf{A}t)[\mathbf{\Sigma}(\infty) - \mathbf{\Sigma}(0)]\exp(\mathbf{A}^T t). \tag{3.11}$$

In practice, as noted above, evaluation of the matrices $\exp(\mathbf{A}t)$ and $\exp(\mathbf{A}^T t)$ is easily accomplished by computing their respective eigenvalues and eigenvectors (*vide supra*).

3.4 ROTATION TRANSLATION BLOCK DISPLACEMENTS AND THE RTB-LANGEVIN PRESCRIPTION FOR HARMONIC OSCILLATOR SYSTEMS

3.4.1 Relevant/Irrelevant Coordinate Decomposition based on Rotation-Translation Block (RTB) Displacement Amplitudes

The RTB decomposition is useful for spring systems in which certain groups of contiguous particles are connected by large spring constants, so that the group moves approximately as a rigid body. By constructing the 6 small displacement vectors that correspond to rigid body motion of the fragment (i.e., three overall translational modes and three orthogonal displacements corresponding to overall rotation of the rigid fragment), and ignoring all other internal motions of the fragment, the number of operative degrees of freedom is significantly reduced, and the rigid body character of the fragment's motion is naturally built into the ensuing classical mechanics of the full system. [Note: In this subsection we focus on the most common and important case that all atoms move in three dimensions and all rigid body fragments are non-collinear. Other special cases can be treated by direct analogy. One such example, namely, a collinear spring system, is studied in detail in Section 3.6.] For a given fragment (which in the case of immediate interest here will correspond to a molecular fragment within a larger molecule), the three overall translational basis vectors are trivial to construct. The basis vectors corresponding to small amplitude rotations of the rigid body can be obtained in a straightforward manner as appropriate ``twists" of the fragment -- one twist about each of the three principal rotation axes [44, 110]. The lab frame twist displacement vectors are easily converted to mass-weighted cartesian coordinates (*vide supra*), as are the overall translation vectors. Thus, each rigid body fragment gives rise to 6 mutually orthogonal unit vectors in the $D = 3N$ dimensional

space of small amplitude motion of the complete molecule. (All components of these $3N$-dim. vectors corresponding to motion of atoms that are not part of a given rigid fragment are set to zero.) It may well be the case that some atoms are not part of any internal rigid fragment. These must be treated in full atomistic detail (3 cartesian displacements for motion in 3 dimensions). The set of all RTB basis vectors for all rigid fragments in the overall ``macromolecule" plus all relevant cartesian coordinates for the atoms that cannot be assigned to a rigid fragment form n_R relevant basis vectors (mutually orthogonal and unit-normed) in the physical $3N$-dim. displacement space. They comprise a subset of all small physical displacements of all the atoms in the macromolecule. Let us denote these unit vectors as $\hat{x}'_i$, $i = 1,...,n_R$. In principle, we can construct $n_I = 3N - n_R$ other unit vectors, all mutually orthogonal to each other and the n_R relevant displacement vectors, e.g., by the Gram-Schmidt orthogonalization procedure [92]. (In fact, within the RTB approximation we will not have to do this, thus significantly reducing the computational effort required to evolve the desired dynamics, whether Newtonian or Langevin in nature.) Let us term these $\hat{x}'_i$, $i = n_R + 1,...,3N$. Now we have a complete set of unit vectors which spans the $3N$-dim. space of atomic displacements. By writing the general small-atom displacement as a linear combination of these, i.e. $\vec{x}(t) = \sum_{i=1}^{3N} x'_i(t)\hat{x}'_i$, we obtain equations of motion of the time-dependent amplitudes $\vec{x}'(t)$, i.e., in the primed basis. This procedure is valid for both Newtonian and Langevin dynamics schemes. Naturally, the EOM's have the same form in the $\hat{x}'$ frame as they do in the frame of lab-fixed cartesian displacements.

3.4.2 The RTB-Langevin dynamics prescription for harmonic oscillator systems

We focus here on the Langevin Equation 3.3, which when subjected to the orthogonal linear transformation introduced in the preceding subsection becomes

$$\ddot{\vec{x}}'(t) = -\mathbf{\Gamma}'\dot{\vec{x}}'(t) - \mathbf{F}'\vec{x}'(t) + \vec{R}'(t)\,. \tag{3.12}$$

Again, the form of the Langevin Equation is unchanged, but the friction matrix, force constant matrix and random force vector are transformed into the $\hat{x}'$ frame. Specifically, if we denote the matrix of $\hat{x}_{i'}$ unit vectors expressed in the mass-weighted lab frame as $\mathbf{T}$ (with each column of $\mathbf{T}$ containing one $\hat{x}_{i'}$), then

$$\mathbf{\Gamma}' = \mathbf{T}^T\mathbf{\Gamma}\mathbf{T}\,;\, \mathbf{F}' = \mathbf{T}^T\mathbf{F}\mathbf{T}\,;\, \vec{R}'(t) = \mathbf{T}^T\vec{R}(t)\,. \tag{3.13}$$

Note that the random force components in the primed frame naturally obey the relations

$$< R_i'(t) >= 0\,;i = 1,...,D$$

and

$$< R_i'(t_2)R_j'(t_1) >= 2k_BT\Gamma_{i,j}'\delta(t_2 - t_1)\,;i,\, j = 1,...,D\,.$$

The transformed Langevin Equation 3.12 describing motion in the primed frame are of course completely equivalent to the original Langevin Equations in mass-weighted lab cartesian displacement coordinates. [Note: In this subsection and the following section, we label the total number of degrees of freedom by D, with the understanding that the relevant coordinates have been identified via the procedure outlined in the previous subsection.]

When groups of atoms are tightly bound into a quasi-rigid body via springs with large force constants, RTB decomposition suggests a natural way to ``freeze out" high frequency motions from the relevant Langevin dynamics. This entails making a closed set n_R Langevin

equations out of equations $1,2,...,n_R$ in the primed frame, which in turn is accomplished by setting the remaining $D-n_R$ irrelevant components $x_i'(t)=0$, $i=n_R+1,...,D$ identically (hence their corresponding velocities $dx_i'/dt=0$, too). Such an approximation can be justified based on the following reasoning. The expectation values of irrelevant coordinates and momenta, i.e., $<x_i'(t)>$ and $<p_i'(t)>$, should become negligibly small as the springs in the quasi-rigid fragments become stiffer. (Again, they get ``frozen out'' of the frictively damped Newtonian dynamics that determines their time evolution.) To show in a rigorous fashion that the rapid fluctuations in these stochastic variables around their mean values have no net influence on the motion of the slow (relevant) coordinates, a more detailed analysis is necessary: this is provided in the next section.

Assuming for now that our assumptions about the separability of fast and slow variable motion in the limit of very stiff internal fragments is valid, the relevant coordinate Langevin Equations form a closed set of EOM's. Note that the $n_R \times n_R$ force constant matrix describing this motion is precisely the reduced Hessian matrix of standard RTB normal modes theory [1, 2]. These n_R Langevin Equations are equivalent to a $2n_R$ dimensional phase-space Fokker-Planck Equation which has the standard harmonic oscillator form, but utilizes the reduced $n_R \times n_R$ relevant coordinate force constant and friction constant matrices.

To state the operational EOM's for the relevant coordinates, let $\mathbf{F}^R$ be the $n_R \times n_R$ matrix obtained by projecting the full mass-weighted laboratory displacement coordinate force constant matrix $\mathbf{F}$ onto the subspace spanned by the n_R relevant coordinate unit vectors; specifically, $[\mathbf{F}^R]_{j,k} = \hat{x}_j'^T \mathbf{F} \hat{x}_k'$. Similarly, define the projection of the primed frame friction coefficient matrix

onto the relevant coordinate subspace as $[\mathbf{\Gamma}^R]_{j,k} = \hat{x}'^T_j \mathbf{\Gamma} \hat{x}'_k$. Finally, we need to project the complete set of initial conditions onto the relevant coordinate subset.

In the case of the expectation values of position and momentum, this is straightforward: the D-dimensional vector of mass-weighted position coordinates is projected via the appropriate dot products on the n_R relevant coordinate axes, and similarly for the momentum expectation value vector. [Any non-zero irrelevant coordinate and momentum components are presumed to be negligibly small at $t = 0$.] Arranging the expectation values of position and momentum of the relevant coordinates into the $2n_R$ dimensional vector $X^R_t = (x'_{1,t},...,x'_{n_R,t}, p'_{1,t},..., p'_{n_R,t})^T$, then $\vec{X}^R_t = \exp(\mathbf{A}^R t)\vec{X}^R_0$, with

$$\mathbf{A}^R = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{F}^R & -\mathbf{\Gamma}^R \end{pmatrix}. \tag{3.14}$$

Note that $\mathbf{A}^R$ is a $2n_R \times 2n_R$ dimensional matrix composed of the indicated $n_R \times n_R$ sub-matrix blocks.

Time evolution of the $2n_R \times 2n_R$ spread matrix for the relevant coordinates, $\mathbf{\Sigma}^R(t)$, is done using Equation 3.11, appropriately specialized to the relevant coordinate subspace; namely,

$$\mathbf{\Sigma}^R(t) = \mathbf{\Sigma}^R(\infty) - \exp(\mathbf{A}^R t)[\mathbf{\Sigma}^R(\infty) - \mathbf{\Sigma}^R(0)]\exp([\mathbf{A}^R]^T t) \ , \tag{3.15}$$

with

$$\mathbf{\Sigma}^R(\infty) = \beta^{-1}\begin{pmatrix} (\mathbf{F}^R)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \tag{3.16}$$

(thus implying that the long-time limit of $\mathbf{\Sigma}^R(t)$ is given uniquely by Equation 3.16). Of course, an initial relevant coordinate spread matrix is needed as input into the computation of $\mathbf{\Sigma}^R(t)$. This implies a decoupling of the relevant from the irrelevant coordinates in the full $2D \times 2D$

initial spread matrix. Usually construction of the appropriate $\mathbf{\Sigma}^R(0)$ will be clear on intuitive grounds.

3.5 MODE DECOUPLING ROUTE TO THE RTB-LANGEVIN HARMONIC OSCILLATOR DYNAMICS

We start from the prescription for exact all-atom Langevin dynamics monitored in the primed coordinate frame. As noted above, the full $2D$ dimensional phase space includes $2n_R$ coordinates/momenta corresponding to the subset of degrees of freedom termed ``relevant" and $2D-2n_R$ coordinates/momenta corresponding to the remaining, ``irrelevant" degrees of freedom. The full Langevin dynamics of this D dimensional harmonic oscillator system couples the motion of relevant and irrelevant degrees of freedom. We seek to identify the details of this coupling, and determine conditions under which it can be ignored. Our task is facilitated by transforming to a new basis in which the phase space variables are ``shuffled" so that the relevant coordinates and momenta are listed first, followed by their irrelevant counterparts. The decoupling of the motion of relevant and irrelevant coordinates/momenta then corresponds to the block diagonalization of all relevant phase-space matrices, e.g., $\mathbf{A}$, $\mathbf{\Sigma}(0)$, etc. Mathematically, the appropriate matrix transformations are easily accomplished using a $2D\times 2D$ shuffling matrix $\mathbf{T}_{sh}$, in which each column consists of an appropriate unit vector. (An explicit example is given in the next section.) The appropriate matrices will naturally be orthogonal in nature, i.e., $\mathbf{T}_{sh}^T=\mathbf{T}_{sh}^{-1}$. Using these matrices, we can transform to the shuffled coordinate system, which will be denoted as ``double primed", via $\mathbf{A}''=\mathbf{T}_{sh}^T\mathbf{A}'\mathbf{T}_{sh}$, $\mathbf{\Sigma}''=\mathbf{T}_{sh}^T\mathbf{\Sigma}'\mathbf{T}_{sh}$, etc.

The double-primed EOM's have precisely the same content as those in the primed frame, but the former suggest fruitful approximations due to their approximately block-diagonal structure. For example, $\mathbf{A}''$ can be presented as

$$\mathbf{A}'' = \begin{pmatrix} \mathbf{A}^R & \mathbf{a}_{1,2}{}^{RI} \\ \mathbf{a}_{2,1}{}^{RI} & \mathbf{A}^I \end{pmatrix}. \tag{3.17}$$

Here $\mathbf{A}^R$ is the $2n_R \times 2n_R$ submatrix which connects the relevant coordinates/momenta to each other (vide supra). Analogously, $\mathbf{A}^I$ is the $2n_I \times 2n_I$ submatrix which connects the irrelevant coordinates/momenta to each other. The off-diagonal matrices $\mathbf{a}_{1,2}{}^{RI}$, which has dimensions $2n_R \times 2n_I$, and $\mathbf{a}_{2,1}{}^{RI}$, which has dimensions $2n_I \times 2n_R$, couple relevant to irrelevant coordinates/momenta. The key feature of $\mathbf{A}''$ is that in the limit of nearly rigid fragments, all the eigenvalues of the submatrix $\mathbf{A}^R$ are well-separated from those of the submatrix $\mathbf{A}^I$, due to the large spring constants that characterize the irrelevant coordinate subset. These large spring constant values are entirely absent from the submatrix $\mathbf{A}^R$, and also from the off-diagonal blocks $\mathbf{a}_{1,2}{}^{RI}$ and $\mathbf{a}_{2,1}{}^{RI}$. Hence, the eigenvalues and eigenvectors of the full $2D \times 2D$ propagator matrix are well approximated by those associated with the block-diagonal approximant to $\mathbf{A}''$ obtained by setting the submatrices $\mathbf{a}_{1,2}{}^{RI}$ and $\mathbf{a}_{2,1}{}^{RI}$ to 0, or, equivalently,

$$\mathbf{A}'' \cong \begin{pmatrix} \mathbf{A}^R & 0 \\ 0 & \mathbf{A}^I \end{pmatrix}, \tag{3.18}$$

thus *decoupling* the (slow) relevant coordinate motion from that of the (fast) irrelevant coordinates.

To finalize the derivation of the RTB-Langevin approximation, we need to consider the time-evolution of the spread matrix $\mathbf{\Sigma}(t)$. Given the block-diagonalization of $\mathbf{A}''$ noted above,

then the propagators $\exp(\mathbf{A}''t)$ and $\exp(\mathbf{A}''^{T}t)$ factorize in a similar fashion. The only remaining issue is the structure of $\mathbf{\Sigma}''(\infty)$, which must also factorize appropriately in order to achieve decoupling of the time evolution of the spread matrices governing the relevant and irrelevant coordinates. Let us begin by expressing the exact infinite-time version of the spread matrix in the primed coordinate system, i.e.,

$$\mathbf{\Sigma}'(\infty) = \beta^{-1}\begin{pmatrix} \mathbf{F}'^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}. \tag{3.19}$$

Now, the high values of the force constants which enter (exclusively) into the irrelevant sub-block of $\mathbf{F}'$ allow us, by virtue of the usual arguments about the eigenvalues and eigenvectors of this matrix (*vide supra*), to approximate it as

$$\mathbf{F}' \cong \begin{pmatrix} \mathbf{F}^{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^{I} \end{pmatrix}. \tag{3.20}$$

Note that this is precisely the factorization which underlies the RTB-normal modes procedure for computing normal modes of vibration as solutions to Newton's Equations of motion. Given this essential factorization approximation, it follows immediately that

$$\mathbf{F}'^{-1} \cong \begin{pmatrix} (\mathbf{F}^{R})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{F}^{I})^{-1} \end{pmatrix}. \tag{3.21}$$

Upon ``shuffling'' to the double-primed coordinate system, we thus obtain

$$\mathbf{\Sigma}''(\infty) = \beta^{-1}\begin{pmatrix} \mathbf{\Sigma}^{R}(\infty) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}^{I}(\infty) \end{pmatrix}, \tag{3.22}$$

with $\mathbf{\Sigma}^{R}(\infty)$ given in Equation 3.16, and analogously for $\mathbf{\Sigma}^{I}(\infty)$. Now the entire propagation scheme for the spread matrix factorizes into relevant and irrelevant pieces, with relevant

submatrices constructed out of matrix elements of $\mathbf{F}$ and $\mathbf{\Gamma}$ with respect to relevant mode basis vectors, as anticipated above.

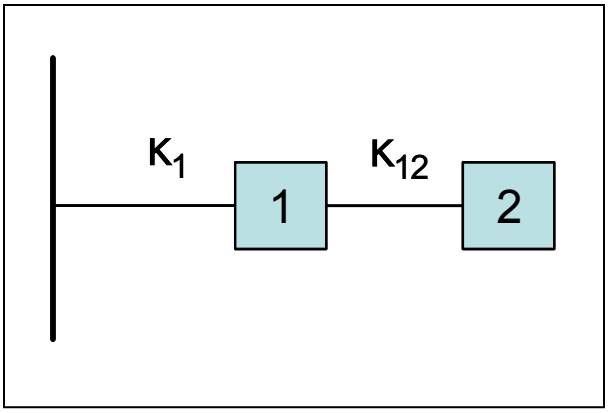

Figure 10. Two-atom collinear spring chain attached to a wall. Details are given in the text.

3.6 A SIMPLE 2 DIMENSIONAL EXAMPLE

We can illustrate the basic principles of the RTB-Langevin dynamics procedure by studying the two dimensional collinear spring system depicted in Figure 10. Taking the spring constant κ_{12} connecting masses 1 and 2 to be sufficiently large will effectively freeze the ``bond" connecting the two particles (``atoms"), rendering it a quasi-rigid diatomic molecule. The RTB-Langevin approximation generates a 1-D prescription for computing the motion of the ``slow" coordinate (essentially, the center of mass [c.o.m.] of the diatomic) without having to explicitly calculate the rapid small amplitude relative vibrational motion of the diatomic. Adopting a harmonic oscillator model of the underlying PES, we have

$$V(x_1, x_2) = \frac{1}{2}\kappa_1 x_1^2 + \frac{1}{2}\kappa_{12}(x_2 - x_1)^2 , \tag{3.23}$$

where x_1 is the position of coordinate 1 along the chain axis, $x_1 = 0$ being the position of atom 1 when the system is at mechanical equilibrium (i.e., all springs are unstretched), etc. Further, we take the masses of both atoms to be unity, with no loss of generality, and consider the case of arbitrary diagonal frictional couplings $\gamma_{j,j} > 0$, $j = 1,2$ on the two atoms (cf. Equation 3.2 above). Clearly, we can write the PES prescribed in Equation 3.23 in the matrix form $V(x_1, x_2) = \frac{1}{2}\vec{x}^T \mathbf{F} \vec{x}$, with $\vec{x}^T = (x_1, x_2)^T$ and

$$\mathbf{F} = \begin{pmatrix} \kappa_1 + \kappa_{12} & -\kappa_{12} \\ -\kappa_{12} & \kappa_{12} \end{pmatrix}. \tag{3.24}$$

Furthermore, the friction constant matrix in the same basis reads: $\mathbf{\Gamma} = diag(\gamma_{1,1}, \gamma_{2,2})$.

Next, we identify the relevant internal coordinates, namely the relative and center of mass motion of the diatomic. Specifically, let $x_{1'} = (x_2 + x_1)/\sqrt{2}$ and $x_{2'} = (x_2 - x_1)/\sqrt{2}$. Clearly, $x_{1'}$ is the center of mass coordinate for motion along the collinear axis, while $x_{2'}$ is the relative coordinate, specifying the degree of expansion/compression of the diatomic's bond length. The normalization factors are chosen to make the transformation between unprimed and primed coordinate frames orthonormal. Thus, $\vec{x}'(t) = \mathbf{T}^T \vec{x}(t)$, with

$$\mathbf{T} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}. \tag{3.25}$$

Then we can compute $\mathbf{F}' = \mathbf{T}^T \mathbf{F} \mathbf{T}$, namely,

$$\mathbf{F}' = \begin{pmatrix} \kappa_1 / 2 & -\kappa_1 / 2 \\ -\kappa_1 / 2 & \kappa_1 / 2 + 2\kappa_{12} \end{pmatrix}. \tag{3.26}$$

Analogously,

$$\mathbf{\Gamma}' = \begin{pmatrix} \Gamma_{1,1}' & \Gamma_{1,2}' \\ \Gamma_{2,1}' & \Gamma_{2,2}' \end{pmatrix} = \begin{pmatrix} (\Gamma_{2,2} + \Gamma_{1,1})/2 & (\Gamma_{2,2} - \Gamma_{1,1})/2 \\ (\Gamma_{2,2} - \Gamma_{1,1})/2 & (\Gamma_{2,2} + \Gamma_{1,1})/2 \end{pmatrix}. \tag{3.27}$$

Now we can synthesize $\mathbf{A}'$,

$$\mathbf{A}' = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{F}' & -\mathbf{\Gamma}' \end{pmatrix}. \tag{3.28}$$

Next, we introduce a shuffling transformation that simply exchanges the order in which the 4 relevant phase-space coordinates are labeled, from (x_1', x_2', p_1', p_2') in the primed basis to (x_1', p_1', x_2', p_2') in the double primed basis,

$$\mathbf{T}_{sh} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.29}$$

In the double primed basis, the propagation matrix is given by $\mathbf{A}'' = \mathbf{T}_{sh}\mathbf{A}'\mathbf{T}_{sh}$. Explicitly, it takes the form of Equation 3.17, with

$$\mathbf{A}^R = \begin{pmatrix} 0 & 1 \\ -\kappa_1/2 & -\Gamma_{1,1}' \end{pmatrix}; \mathbf{A}^I = \begin{pmatrix} 0 & 1 \\ -[\kappa_1/2 + 2\kappa_{1,2}] & -\Gamma_{2,2}' \end{pmatrix} \tag{3.30}$$

and

$$\mathbf{a_{1,2}}^{RI} = \begin{pmatrix} 0 & 0 \\ \kappa_1/2 & -\Gamma_{1,2}' \end{pmatrix}; \mathbf{a_{2,1}}^{RI} = \begin{pmatrix} 0 & 0 \\ \kappa_1/2 & -\Gamma_{2,1}' \end{pmatrix}. \tag{3.31}$$

Note that the stiff spring constant κ_{12} appears only in the block $\mathbf{A}^I$. Thus, as $\kappa_{12} \to \infty$ the eigenvalues of $\mathbf{A}^I$ become large in magnitude (in the present example they approach the values $\pm i\sqrt{2\kappa_{12}}$), while the eigenvalues of $\mathbf{A}^R$ remain of order 1, as do the elements of the matrices

$\mathbf{a}_{1,2}^{RI}$, $\mathbf{a}_{2,1}^{RI}$. Thus, the matrix $\mathbf{A}''$ becomes well represented by the block diagonal approximant in which $\mathbf{a}_{1,2}^{RI} \to 0$ and $\mathbf{a}_{2,1}^{RI} \to 0$. This achieves the desired decoupling between the time-evolution of the relevant (R) and irrelevant (I) subsets of phase space variables. Since the irrelevant phase space coordinates do not affect the motion of the relevant ones, the irrelevant variables do not have to be explicitly tracked (or even constructed) if we are only interested in the motion of the slow (``relevant") degrees of freedom.

We turn now to some representative numerical calculations. Let us focus first on the time evolution of the frictively damped positions and momenta of the two particles (cf. Equation 3.6). For concreteness, consider the initial condition that the center of mass (c.o.m.) coordinate $x_{1'}(t)$ is displaced from its equilibrium position, while the relative coordinate is *not* displaced, and the initial velocities of both coordinates are zero. When the system is released from this configuration at $t = 0$ the c.o.m. coordinate will oscillate, and in doing so will induce motion in the diatomic relative coordinate. As $t \to \infty$, the oscillations of both coordinates will damp out, with the particles ultimately coming to rest at their equilibrium positions, i.e. $x_1(\infty) = x_2(\infty) = 0$. The central issue is to what extent the motion of the c.o.m. coordinate is independent of that of the relative coordinate.

The answer depends directly on the value of κ_{12} (all other parameters being held constant). To illustrate this, we set the value of $\kappa_1 = 1.0$. Further, we take the values of the atomic friction constants in mass weighted coordinates as $\gamma_{1,1} = 0.15$ and $\gamma_{2,2} = 0.20$. Then, for a diatomic spring constant of $\kappa_{12} = 5.0$, the exact time-evolution of the center of mass position $x_{1'}(t)$, center of mass momentum $p_{1'}(t)$, relative coordinate position $x_{2'}(t)$, and relative coordinate momentum $p_{2'}(t)$ are shown in Figure 11a. These results should then be compared to

the corresponding results obtained for the case that $\kappa_{12} = 20.0$, all other parameters being held fixed, which are presented in Figure 11b. The main difference between Figure 11a ($\kappa_{12} = 5.0$) and Figure 11b ($\kappa_{12} = 20.0$) is that the relative coordinate motion is significantly reduced in Figure 11b. The time evolution of the c.o.m. coordinate and momentum appears to be rather similar in the two cases. This is confirmed more quantitatively in Figure 12, where we compare the exact c.o.m. trajectory $x_{1'}(t)$ for $\kappa_{12} = 5.0$ and $\kappa_{12} = 20.0$ with the corresponding TB approximant, i.e., 1-dim. effective harmonic oscillator motion of the c.o.m. corresponding to force constant $F_{1,1}' = \kappa_1/2$ and friction constant $\Gamma_{1,1}'$. (We term this the TB approximant because it is obtained using the RTB-Langevin strategy, but for a quasi-rigid fragment confined to collinear motion such as the diatomic considered here, there is only one rigid body collective motion, namely, translation of the center of mass.) Note that $x_{1,TB}'(t)$ is, by construction, independent of the value of κ_{12}. The full all-atom trajectory $x_{1'}(t)$ *does* of course depend on the value of κ_{12}. Clearly, it approaches the TB approximant curve uniformly as $\kappa_{12} \to \infty$.

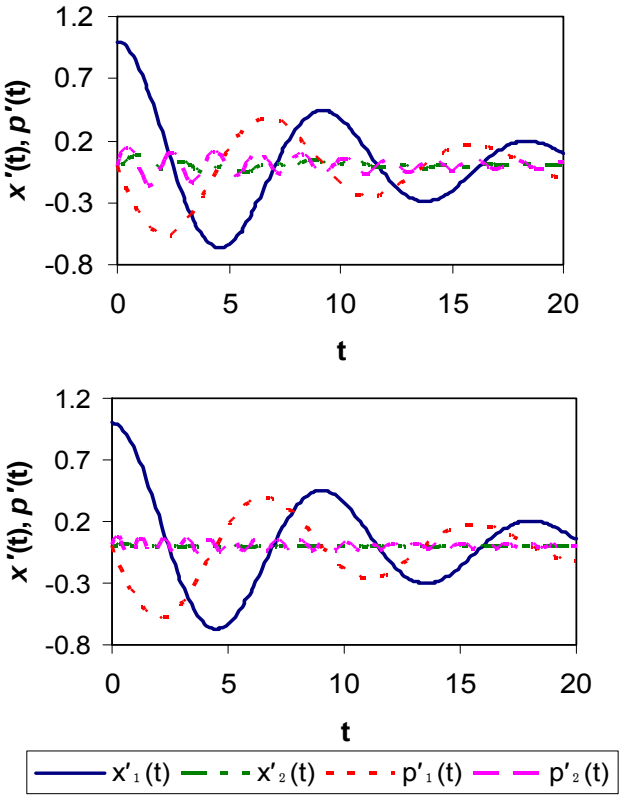

Figure 11. Exact all-atom $\vec{x}'(t), \vec{p}'(t)$ for initial displacement conditions specified in text when: a) $\kappa_{12} = 5.0$, b) $\kappa_{12} = 20.0$.

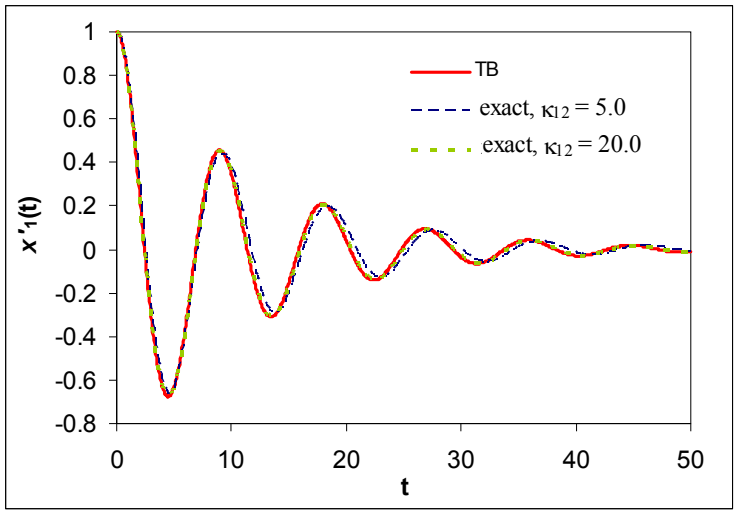

Figure 12. Time evolution of c.o.m. coordinate $x_{1'}(t)$ for initial displacement conditions specified in text, comparing TB approximant with exact all-atom Langevin dynamics when $\kappa_{12} = 5.0$ and when $\kappa_{12} = 20.0$.

Next, we analyze the time-evolution of the spread matrix $\mathbf{\Sigma}(t)$ of this system. Again, there are many possible choices for the initial values of the elements in this matrix. For concreteness and simplicity, let us consider the case that all of them are zero, i.e., $\mathbf{\Sigma}(0) = 0$, which corresponds to the situation where the positions and momenta of both atoms are precisely specified. At $t = 0$, these atoms are coupled to the heat bath inherent in the Langevin dynamics model, so that at long time they relax to their Boltzmann equilibrium phase space distribution of values. The time course of this relaxation process is of interest. For concreteness, we set the temperature $k_B T = 1$. Figure 13a then shows for the case that $\kappa_{12} = 5.0$ the time evolution of second moments of the probability distribution involving the center of mass position and/or momentum, comparing the all-atom (exact) values to the corresponding TB approximants. In

both the exact and TB approximation scenarios, we expect that $< p_{1'}(\infty)^2 >= k_B T$. Furthermore, the long time value of $< x_{1'}(t)^2 >$ reflects the underlying potential energy surface. In particular, since the TB approximation represents a decoupling of c.o.m. from relative coordinate motions, the long time value of $< x'_{1,TB}(t)^2 >$ obeys the standard formula for a 1-d unit mass harmonic oscillator with the effective spring constant $\kappa_{eff} \equiv \kappa_1/2$, namely $< x'_{1,TB}(\infty)^2 >= k_B T/\kappa eff$. Finally, the cross correlation function $< x_{1'}(t)\, p_{1'}(t) >$ is expected to evolve from a value of 0 at $t = 0$ to non-zero values for $t > 0$, and then return to 0 as $t \to \infty$. The agreement between the full dynamics and their TB approximants is reasonable. The most pronounced discrepancy between the TB approximate dynamics and the exact dynamics appears in the mean squared position expectation value, $< x_{1'}{}^2(t) >$. The TB approximation goes to a slightly different long-time asymptotic value, as, indeed, it should. From the discussion above, the exact value of $< x_{1'}{}^2(\infty) >$ is

$$< x_{1'}{}^2(\infty) >= k_B T \mathbf{F}'^{-1}_{1,1} = k_B T(\kappa_1/2 + 2\kappa_{12})/\left[(\kappa_1/2 + 2\kappa_{12})\kappa_1/2 - \kappa_1^2/4\right].$$

Clearly, this approaches the value of the corresponding TB approximant noted above as $\kappa_{12} \to \infty$, keeping all other parameters fixed. In Figure 13b, we show the analogous results for a system with $\kappa_{12} = 20.0$: here the correspondence between all-atom and TB approximant dynamics is very good.

While the 2-dim. collinear example considered above nicely illustrates the theoretical underpinnings of the RTB-Langevin scheme, it is perhaps not totally convincing, in that the motion of a real protein is non-collinear, and we do not have ``tunable'' control

over the fragments of the protein which we designate as quasi-rigid. Consequently, it behooves us to study a more realistic example in detail, as is done in the next section.

.

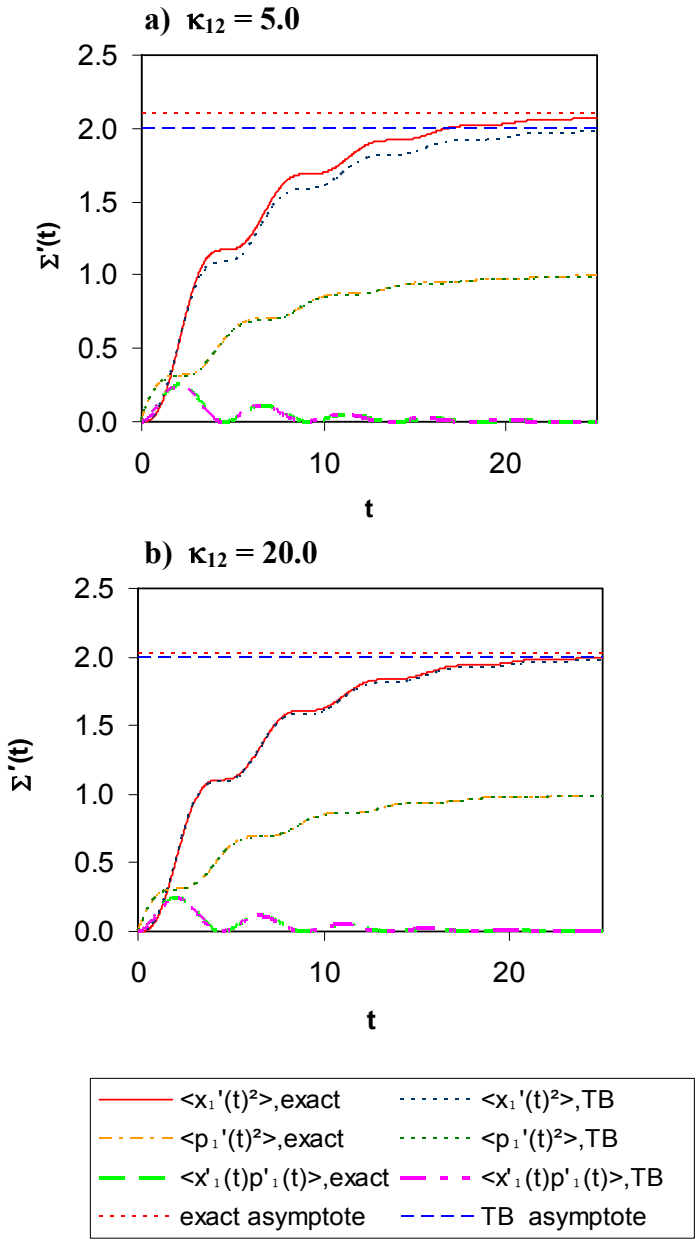

Figure 13. $< x_1'^2(t) >$, $< p_1'^2(t) >$ and $< x_1'(t)\, p_1'(t) >$ for: a) $\kappa_{12} = 5.0$, b) $\kappa_{12} = 20.0$. The results of exact all-atom Langevin dynamics are compared with the corresponding TB approximants as noted in the figure legend. Long-time asymptotic values of $< x_1'^2 >$ associated with exact all-atom Langevin dynamics and TB-Langevin dynamics are also indicated.

3.7 APPLICATION OF THE RTB-LANGEVIN DYNAMICS METHOD TO RELAXATION DYNAMICS OF GRAMICIDIN-A IN RESPONSE TO A SUDDEN IMPULSE

Gramicidin-A (GA) is a small antibiotic peptide, 15 amino acids in length, that dimerizes in lipid bilayer membranes so as to form a membrane-spanning pore. Ions, primarily cations, are then able to permeate through the protein pore, and hence to pass through the membrane. For present purposes, GA serves as a model system for illustrating the dynamical response of a protein to an external perturbation. Our goal is to identify an appropriate collection of internal rigid fragments within the GA dimer molecule, and then to compute RTB-Langevin dynamics arising in response to an initial disturbance of the molecule. Because the GA dimer is modest in size, we can also perform all-atom Langevin dynamics for the same system and thus assess the accuracy of RTB-Langevin dynamics. The specific disturbance we will consider here is a sudden velocity kick at $t = 0$ applied near one end of the GA dimer. This choice is motivated in part by the observation that in some ion channels, most notably ligand gated ion channel (LGIC) proteins[13, 26, 30], binding of an appropriate ligand to the extra-cellular domain (ECD) of the protein induces a conformational change in the ECD which is subsequently propagated into the transmembrane domain, ultimately causing the latter to gate open or shut. The simple model system and dynamical scenario investigated here thus bear some resemblance to signal transduction in LGICs.

We embedded the GA dimer in a membrane mimetic consisting of a random array of 245 pinning balls, as illustrated in Figure 5. These interact with the gramicidin peptide via Lennard-Jones pair potentials to hold the peptide in place. For the initial GA dimer structure we used Arseniev's NMR-derived coordinates [96]. This structure was subjected to energy minimization

based on the Amber 94 force field [97] in order to determine the equilibrium configuration of the membrane-embedded GA dimer. (Lennard-Jones parameters ε =0.15 *kcal/mol* and σ =1.78 *Å* for the well depth and van der Waals diameter, respectively, were used to generate the pairwise interactions between each pinning ball and GA atom.) No explicit water molecules were included in the calculation and the pinning balls which form the mimetic membrane bilayer were not allowed to move.

For the RTB decomposition of the molecule, each of the eight indole groups was treated as a rigid fragment (the parts highlighted in red in Figure 5), and all other atoms in the system as independent atoms. For each rigid body fragment, RTB basis vectors corresponding to the 3 c.o.m. translational degrees of freedom and the rotations about each of the 3 principal moments of inertia were constructed, and each independent atom was represented via its 3 Cartesian displacement coordinates. We chose indole groups as the rigid fragments within the molecule because by virtue of their chemical composition they are indeed relatively rigid, as was borne out in our previous study of RTB-Newtonian dynamics in the same system [104].

3.7.1 Properties of the Friction Matrix

The friction matrix used here was modeled as a *diagonal* matrix of friction coefficients with numerical values determined by applying Stokes's formula to the individual atoms. In mass weighted Cartesian displacement coordinates $\Gamma_{i,i} = 6\pi\eta a_i/m_i$, where η is the solvent viscosity, a_i is the hydrodynamic radius of the atom and m_i is its mass; furthermore, $\Gamma_{i,j} = 0$ for $i \neq j$. Hydrodynamic interactions [99] between the atoms of the GA dimer were neglected.

The hydrodynamic radii of the atoms in the GA dimer were determined by the relative degree of surface exposure of each atom [106] to the solvent in the presence of the membrane mimetic. The array of pinning balls of the membrane were included in this calculation to account for the shielding of gramicidin-A by the membrane. The GETAREA [111] molecular surface program was used with a probe radius of 1.4 Å to determine the surface area exposed to solvent for each atom in the GA dimer model shown in Fig 5. Van der Waals parameters for the atoms were obtained from the AMBER 94 force field. Following Pastor and Karplus [107] , for a fully exposed atom a hydrodynamic radius of 0.77 Å was employed, except for hydrogen atoms. For a fully exposed hydrogen atom the value of the hydrodynamic radius was taken as 0.2 Å . After the degree of surface exposure was ascertained, the hydrodynamic radius of each atom was scaled to these fully exposed atomic radii using the square root of the fraction of the solvent accessible surface area, which was obtained via the GETAREA program.

3.7.2 Time evolution of the position and momentum expectation values

After determining the force constant and friction constant matrices, the eigenvalues and eigenvectors of the propagator matrix **A** (Equation 3.7) are computed using the HQR2, BALANC and QRTHES routines in EISPACK [109]. These have the following general properties. All the eigenvalues have negative or zero real parts. All complex-valued eigenvectors occur in conjugate pairs (with their corresponding eigenvalues also being complex conjugates). As noted in section 3.3, the eigenvalues and eigenvectors of the **A** matrix can be used to propagate the relaxation kinetics of a GA system subjected to an externally imposed initial perturbation. Denoting the diagonal matrix of eigenvalues of **A** as $\mathbf{\Lambda}$ and the corresponding matrix of eigenvectors as **U** (with one eigenvector in each column of **U**), we can write

$\vec{X}_t = \mathbf{U}e^{\Lambda t}\mathbf{U}^{-1}\vec{X}_0$. The initial phase space expectation values of the molecular coordinates/momenta, $\vec{X}_0$, can be chosen in any manner desired. In the numerical examples presented below, the initial conformation of the GA dimer is chosen to represent the effect of a sudden external impulse on the molecule. At $t = 0$, starting with all the atoms in the GA dimer at rest in their equilibrium positions, the atoms in the third residue from the top of the molecule are given a non-zero velocity along the channel axis (all atoms in the amino acid are assigned the same velocity value). This velocity kick is the mechanical impulse applied to the molecule: we expect that this impulse will be transmitted along the channel axis via inter-atomic interactions until the system relaxes back to its equilibrium state (the GA dimer is held in place via the presence of pinning ball membrane mimetic).

To summarize the chosen initial conditions for $\vec{X}_0$: In the illustrations of the propagation scheme presented here (cf. Figures. 14-16), an initial velocity kick 0.1 Å /ps in magnitude is imparted to all atoms in the third amino acid from the top of the molecule, which does not contain any of rigid indole groups, the direction of the kick being vertically downward (from top to bottom in Figure 5). We subsequently compute the time evolution of the phase space distribution function that describes the relaxation of the perturbed molecule back to thermal equilibrium, via both all-atom Langevin dynamics and approximate RTB-Langevin dynamics. For illustrative purposes, we present here a variety of details concerning the displacement of the C-α atoms along the channel axis.

We consider three different values of solvent viscosity, namely i) $\eta = 0$, the zero friction limit, ii) $\eta = 1cp$, a value which is representative of the magnitude of the friction typically experienced by a protein molecule in the presence of water solvent [112] , and iii) $\eta = 10cp$, which represents a hypothetical highly viscous medium. For each case, the RTB-Langevin

dynamics propagation is compared with all-atom Langevin dynamics. In Figures. 14-16, we focus on the time evolution of the z-coordinate displacement of the C_α atoms in the peptide (presented in laboratory displacement coordinates, i.e., the q_i's of Equation 3.1).

Figure 14 displays the results for the z-coordinate displacement (note: the direction the z-axis is upward [from bottom to top] in Figure 5) of the C_α atoms in the peptide when there is no solvent, i.e., $\eta = 0$. Even for relatively long times, reasonable agreement between the exact all-atom and RTB-Langevin dynamics is sustained. As a consistency check, we have compared the results for both all-atom and RTB-Langevin dynamics with the corresponding results obtained by standard vibrational normal modes analysis [1, 2, 88]. Since $\eta = 0$, the Langevin dynamics should reduce to Newtonian dynamics, and we found (results not shown) that it indeed does so.

Figure 15 displays results for a solvent characterized by viscosity $\eta = 1cp$. The initial time dynamics is very similar to that obtained when there is no solvent damping (Figure 14). For longer times, the effect of damping by solvent becomes clearly visible. The initial velocity kick is still transmitted through the molecule, but the interaction of the protein molecule with solvent causes the propagating deformation of the protein to damp to zero after a finite period of time. For the parameter choices and initial conditions invoked here, RTB-Langevin dynamics is seen to be in nearly perfect agreement with all-atom Langevin dynamics for all times.

In Figure 16, we present results for the high viscosity solvent characterized by $\eta = 10cp$. As is clear from the figure, solvent damping is discernably stronger in this case. The energy in the initial velocity kick nearly dies out before it reaches the other end of the molecule due to the high solvent viscosity. From Figure 15 and Figure 16, it is apparent that taking the indole groups in the GA dimer to be rigid bodies leads to a smaller error in the RTB-Langevin dynamics

trajectory with respect to exact all-atom results when frictional damping is applied than in the zero friction case (Figure 14).

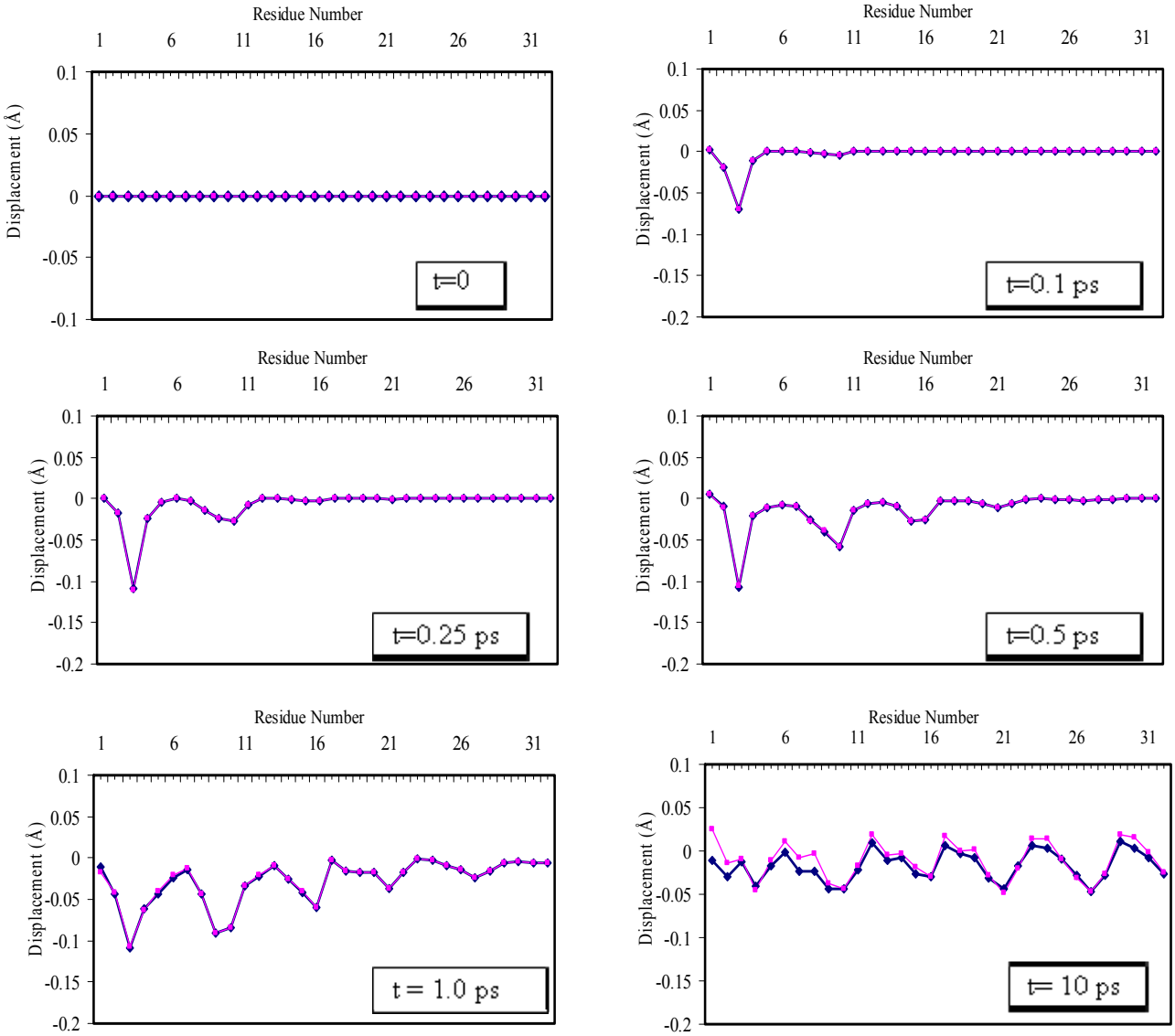

Figure 14 Time evolution of z (channel axis) displacements of C_α atoms from the equilibrium structure of gramicidin-A dimer for $\eta = 0$ cp. Two propagation schemes are compared: RTB-Langevin dynamics (filled squares) and all-atom Langevin dynamics (filled diamonds). At t=0, the third amino acid from top of the molecule is kicked with a velocity of -0.1 Å / ps. Residue number presented here is along the channel axis, numbering from top to bottom (cf. Figure 5).

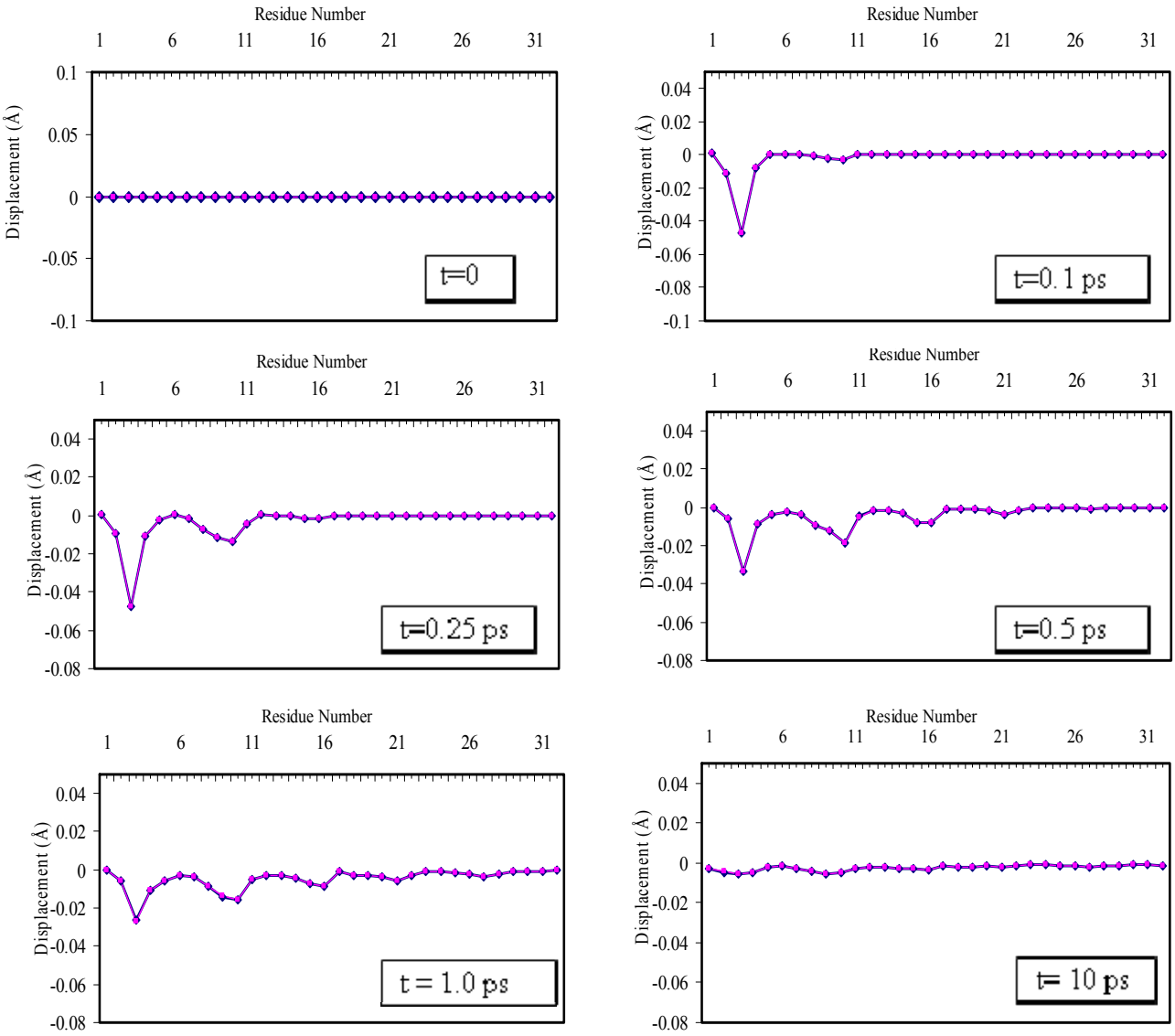

Figure 15. Time evolution of z (channel axis) displacements of C_α atoms from the equilibrium structure of gramicidin-A dimer for $\eta = 1$ cp. Two propagation schemes are compared: RTB-Langevin dynamics (filled squares) and all-atom Langevin dynamics (filled diamonds). At t=0, the third amino acid from top of the molecule is kicked with a velocity of -0.1 Å /ps. Residue number presented here is along the channel axis, numbering from top to bottom (cf. Figure 5).

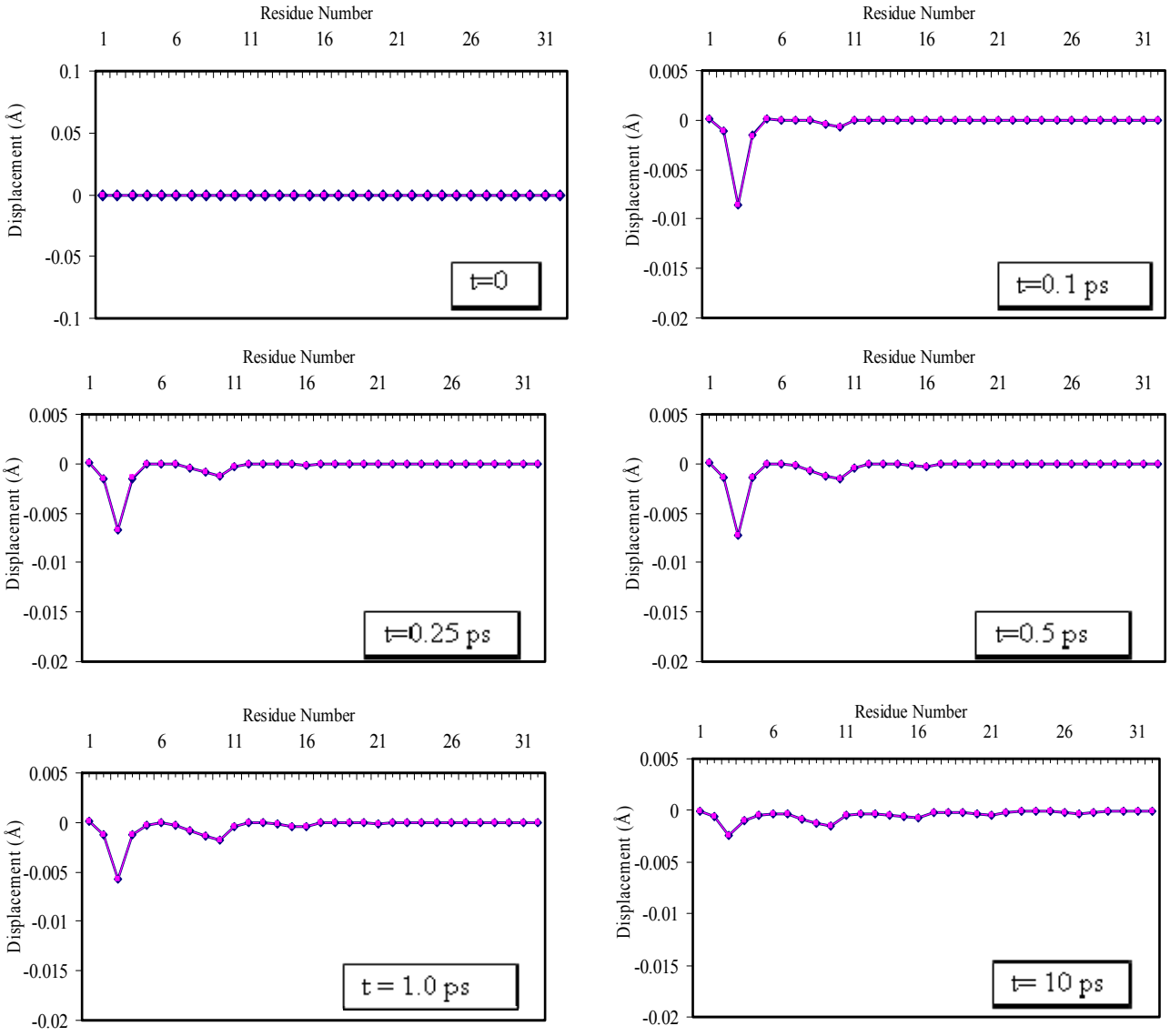

Figure 16. Time evolution of z (channel axis) displacements of C_α atoms from the equilibrium structure of gramicidin-A dimer for $\eta = 10$ cp. Two propagation schemes are compared: RTB-Langevin dynamics (filled squares) and all-atom Langevin dynamics (filled diamonds). At t=0, the third amino acid from top of the molecule is kicked with a velocity of -0.1 Å /ps. Residue number presented here is along the channel axis, numbering from top to bottom (cf. Figure 5**).**

3.7.3 Time evolution of the spread matrix

Next we analyze the time evolution of the spread matrix, as prescribed by Equation 3.11. A wide range of possible initial values of the spread matrix can be contemplated, depending on the physical scenario of interest. As was done in our study of the 2-dim. collinear oscillator system in Section. 3.6, in the calculations presented here we chose for concreteness and simplicity $\mathbf{\Sigma}(0) = 0$, which corresponds to the case where all position and momentum values are precisely specified (have zero dispersion about their mean values) at *t=0*. Because of the coupling to a heat bath in the Langevin model, they will relax to the appropriate Boltzmann distribution at long times.

We compare the time evolution of the spread matrix as calculated via all-atom Langevin dynamics to that obtained within the RTB-Langevin approximation for solvents with two different viscosities, namely, again, $\eta =$ 1 *cp* and 10 *cp*. In the all-atom case, the matrix $\mathbf{A}$ is built using the full atomic force constant matrix $\mathbf{F}$ and friction matrix $\mathbf{\Gamma}$ (expressed in the basis of mass-weighted Cartesian displacements) which are both $3N \times 3N$ in size. Consequently, for all-atom dynamics $\mathbf{\Sigma}(\infty)$ is the $6N \times 6N$ matrix given by Equation 3.10. To compare these results to RTB reduced space results in which we neglect the $3N - n_R$ irrelevant degrees of freedom, a transformation from all-atom to RTB reduced space is carried out by projecting the all-atom $\mathbf{\Sigma}(t)$ onto the subspace spanned by the n_R relevant coordinate RTB unit vectors; the resultant $2n_R \times 2n_R$ dimensional matrix will be denoted as $[\mathbf{\Sigma}(t)]^R$. The long time steady state value of $[\mathbf{\Sigma}(t)]^R$ is given, independent of its initial value, by the $2n_R \times 2n_R$ matrix

$$[\mathbf{\Sigma}(\infty)]^R = \beta^{-1} \begin{pmatrix} (\mathbf{F}^{-1})^R & 0 \\ 0 & 1 \end{pmatrix}, \tag{3.32}$$

where each of the indicated sub-blocks is an $n_R \times n_R$ matrix.

For the RTB-Langevin case, the matrix $\mathbf{A}^R$ is built using the RTB reduced force constant and friction matrices, which are both $n_R \times n_R$ in size (cf. Section 3.4.2 for details of RTB projection). Time evolution of the spread matrix $\mathbf{\Sigma}^R(t)$ leads to the long time steady state prescribed by Equation 3.16.

In Figure 17a, one representative diagonal element of the upper left sub-block of $[\mathbf{\Sigma}(t)]^R$ (designated in the figure legend as ``exact") and $\mathbf{\Sigma}^R(t)$ (designated in the figure legend as ``RTB") is shown for the viscosity values $\eta = 1\,cp$ and $10cp$. Here we illustrate the time evolution of the variance in the motion of the z coordinate of the C_α carbon in residue 16 (cf. Figure 5 for the residue index labeling convention). Note that for the initial condition $\mathbf{\Sigma}(0) = 0$, the entire time-evolved matrix $\mathbf{\Sigma}(t)$ scales proportional to $k_B T$ (cf. Equations 3.10,3.11). Thus, we plot in Figure 17a $< \delta z_J(t) \delta z_J(t) > / k_B T$, where $\delta z_J(t) \equiv z_J - z_J(t)$, with z_J being the mass-weighted z coordinate of the C_α carbon in residue 16, which we label here as J (the average being taken over the phase space distribution at time t). Note that the long-time asymptotic values of the all-atom Langevin and RTB-Langevin results are discernably different. (We checked that the asymptotic values of the spread matrix elements shown in 17a are consistent with the formulae for $[\mathbf{\Sigma}(\infty)]^R$ and $\mathbf{\Sigma}^R(\infty)$ given above. Similar consistency checks were performed for the long-time steady state dynamics displayed in Figure 17b and Figure 17d below.) This is expected since the indole groups in the GA dimer system are not absolutely rigid: for the ideal case in which the frozen groups are completely rigid the steady state values would be identical, as illustrated for the two-atom collinear spring system analyzed above (cf. Figure 13).

In Figure 17b, we plot the time evolution of the quadratic correlation of the mass-weighted displacement of the z components of the residue 16 C_α carbon and the residue 15 C_α carbon atom, once again normalized by $k_B T$, viz. $< \delta z_J(t) \delta z_K(t) > / k_B T$, where K represents the mass-weighted z coordinate of the C_α atom in residue 15. This element of the matrix illustrates the general trend which is seen in the off-diagonal elements within the upper left sub-blocks of $[\mathbf{\Sigma}(t)]^R$ and $\mathbf{\Sigma}^R(t)$.

In Figure 17c, we show the early-time evolution of the variance in the mass-weighted momentum of residue 16 C_α in the z direction, normalized again by $k_B T$, i.e., $< \delta p_J(t) \delta p_J(t) > / k_B T$. It is clear from this figure that the momentum distribution reaches Maxwell-Boltzmann equilibrium more quickly as the viscosity of the solvent is increased. Figure 17d presents the same spread matrix element out to much longer times. This element of the matrix illustrates the general trend which is seen in the diagonal elements from the lower right sub-block of the spread matrix of $[\mathbf{\Sigma}(t)]^R$ and $\mathbf{\Sigma}^R(t)$.

We have also computed the time evolution of lower left and upper right sub-blocks of the spread matrix, which detail the time evolution of bilinear correlation between momentum and position values. The long-time steady-state values for those parts of the spread matrices (not shown) go to zero, as expected.

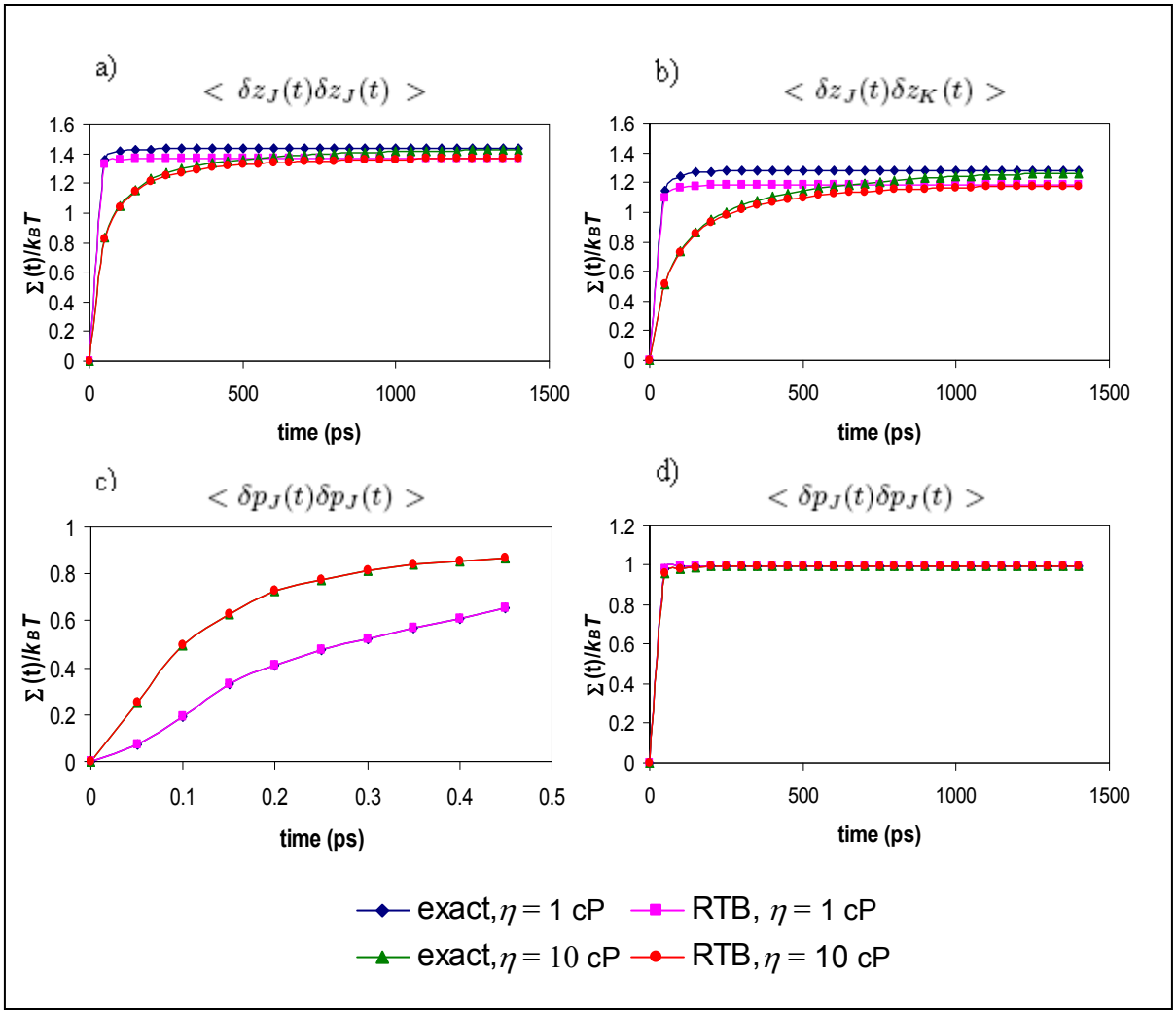

Figure 17. Time evolution of selected elements of the phase space distribution spread matrix $\boldsymbol{\Sigma}(t)$ for the gramicidin-A dimer. Panel: a) The mean squared displacement of the z displacement of the C_α atom of residue 16. b) The cross-correlation of the z displacements of the C_α atoms of residues 16 and 15. c) The mean squared displacement of the z component of the momentum of the C_α atom of residue 16. d) The same element as c) but out to 1500 ps.

3.7.4 Accuracy and Efficiency Issues

The primary purpose of the numerical illustrations presented in this section is to establish the accuracy of the RTB-Langevin approximation for computing Langevin dynamics in the harmonic regime of a protein characterized by a set of nearly rigid internal fragments. We chose the GA dimer as a test system because it is small enough that exact all-atom Langevin dynamics can be computed to benchmark results obtained by the approximate RTB-Langevin method. In the same spirit, we identified the 8 indole groups as rigid fragments, because these are guaranteed to be nearly rigid by virtue of their chemical bonding. Using this system as a test case, we were indeed able to demonstrate reasonable accuracy of the RTB-Langevin dynamics approximation, as demonstrated in Figures 14-17.

The efficiency gain in freezing the 8 indole groups in this system is modest. The number of dynamical degrees of freedom goes from 1656 (corresponding to 552 atoms in the peptide) to 1344 relevant coordinates utilized in the RTB-Langevin calculation, resulting in a time savings of roughly a factor of two. However, dramatically improved scaling (reduction of storage space and cpu time) is inevitable when the procedure is applied to large proteins with many frozen internal fragments ranging from 1 amino acid in size up to entire helices, domains, etc. In closely related RTB normal modes calculations, there are several examples in the recent literature where breaking a large protein into fragments consisting of 1-5 contiguous amino acids enabled a normal modes analysis (NMA) to be performed successfully, whereas the analogous computation incorporating all 3 Cartesian coordinates of every atom in the NMA would not have been feasible [30, 91]. From a computational standpoint, RTB-Langevin analysis is essentially identical to RTB NMA, except that the relevant vector space is twice as large for the Langevin case (since one works in phase space rather than configuration space, as done in NMA): thus the

basic scaling properties with system size and number/size of internally rigid blocks must be similar.

In the GA dimer example, we could not control the stiffness of the indole moieties, but we asserted that these were stiff, and indeed found from direct comparison with all-atom exact results errors of ca. 10 % in the long-time values of quadratic positional correlation functions. In the case of a large protein, where exact all-atom calculations are intractable, there is no simple prescription for computing the ``percentage accuracy" of the RTB-Langevin approximation. The only simple systematic check is to reduce the size of the originally assumed rigid blocks, and see if the predicted large time and length scale dynamics changes significantly [2]. If not, the blocks are ``small enough".

3.8 DISCUSSION AND CONCLUSIONS

In this paper we have developed an approximate solution of the Langevin Equation for a system comprised of many coupled harmonic oscillators which is designed to accelerate computations of the Langevin dynamics of large molecules characterized by one or more quasi-rigid fragments. The essence of our method is to ``freeze" the rigid internal fragments in a conceptually appealing and computationally straightforward manner, namely, using the RTB decomposition procedure previously introduced by Sanejouand and coworkers [1, 2] in the context of standard vibrational normal modes of protein molecules. This can significantly reduce the number of dynamical degrees of freedom in the problem, allowing one to focus on the low frequency collective modes that drive large scale conformational changes in the protein. We were able to derive the RTB-Langevin prescription in a controlled fashion, i.e., as a mode decoupling approximation to the

full all-atom Langevin dynamics in the limit that the internal vibrational frequencies corresponding to relative motion of atoms within each internal fragment become very high (as the springs holding these fragments become very stiff). In the present work we focused exclusively on systems governed by quadratic (harmonic oscillator) potential energy surfaces (PESs). While this model has proven to be quite useful for studying many types of protein motions, it has obvious limitations, namely, it cannot account for any effects due to inherent anharmonicities in the molecular force field. However, even for an anharmonic system, the PES becomes *locally* harmonic about any reference point in configuration space, and thus, if we take small enough time steps, the analysis presented in this paper holds for each small time step (with the potential energy surface expanded locally about the instantaneous configuration). This observation can potentially be exploited to develop a stochastic propagation algorithm, directly in the spirit of standard (all-atom) Langevin dynamics algorithms [99], that identifies at each time step the RTB displacements corresponding to overall translation and rotation of each internal rigid-fragment, and propagates only the relative motion between fragments, automatically ``freezing" the internal motion of the fragments.

3.9 ACKNOWLEDGMENTS

This work was supported in part by NSF Grant No. CHE-0518044 and ARO-MURI Grant No. DADD19-02-1- 0227.

4.0 DYNAMIC LINEAR RESPONSE THEORY FOR CONFORMATIONAL RELAXATION OF PROTEINS

4.1 ABSTRACT

Dynamic Linear Response Theory is adapted to the problem of computing the time evolution of the atomic coordinates of a protein in response to the unbinding of a ligand molecule from a binding pocket within the protein. When the ligand dissociates from the molecule, the protein molecule finds itself out of equilibrium, and its configuration begins to change, ultimately coming to a new stable configuration corresponding to equilibrium in a force field which lacks the ligand-protein interaction terms. Dynamic Linear Response Theory relates the non-equilibrium motion of the protein atoms which ensues after the ligand molecule dissociates to *equilibrium* dynamics in the force field, or equivalently, on the potential energy surface (PES) relevant to the unliganded protein. In general the connection implied by Linear Response Theory hold only when the ligand-protein force field is small. However, in the case where the PES of the unliganded protein system is a quadratic (harmonic oscillator) function of the coordinates, and the force of the ligand upon the protein molecule in the ligand-bound conformation is constant (independent of the positions of the protein atoms), dynamic LRT is *exact* for any ligand-protein force field strength. An analogous statement can be made for the case where the atoms in the protein are subjected to frictional and random noise forces in accord

with the Langevin Equation (to account for interaction of the protein with solvent, for example). We numerically illustrate the application of dynamic LRT for a simple harmonic oscillator model of the ferric binding protein, and for an analogous model of T4 lysozyme. Using a physically appropriate value of the viscosity of water to guide the choice of friction parameters, we find relaxation time scales of residue-residue distances on the order of several hundred ps. Comparison is made to relevant experimental measurements.

4.2 INTRODUCTION

Protein conformational transformations induced by an externally applied perturbation are of great biological importance. Consider, for example, the response of a protein to ligand binding. Introduction of a ligand molecule into a binding pocket generates new forces on the protein, causing it to distort in a fashion that has direct functional implications. Specific types of ligand binding include: i) binding of agonist molecules to a pocket in the extra-cellular domain of a ligand-gated ion channel (ultimately this causes the transmembrane pore to gate open)[13]; ii) binding of an oxygen molecule to a binding pocket in myoglobin for storage purposes[113]; iii) binding of a drug molecule to a specific site or pocket in a protein, thus inducing conformational changes in the protein which are significant enough to alter its function [114-117]. Numerous approaches to calculating these effects (as well as those associated with the reverse process, i.e., ligand dissociation from a binding pocket) have been proposed [56, 118-122]. One promising approach for predicting the final configuration given the initial configuration and the applied perturbation force is static Linear Response Theory [59]: all relevant quantities can be calculated from static equilibrium correlation functions based on the Boltzmann distribution associated with

the final state potential energy surface (PES). Such averages can be computed via Metropolis Monte Carlo or molecular dynamics (MD) for general PESs, including effects of protein force field anharmonicities and water solvent.

Also of considerable interest are the *temporal* signatures of the relaxation process, the most obvious being the essential time scale needed for the transformation to take place. Indeed, with modern spectroscopic methods (e.g., FRET [61] or ESR [62]) finer details of the relaxation process can be gleaned. In particular, by tagging two residues of the protein with appropriate chromophores, the time-evolution of the distance between these residues can be directly monitored over the course of the relaxation from initial to final conformations. In the present paper we develop the dynamical analog of static LRT, namely dynamical LRT, which enables such temporal and intrinsically non-equilibrium details to be computed from e*quilibrium* time correlation functions on the final state PES. We then illustrate the basic principles of dynamic linear response theory by applying it in a simplistic manner to ligand binding/unbinding processes associated with two proteins, namely the ferric binding protein (FBP), which binds Fe^{3+} ions, and the enzyme T4 lysozyme (T4L), which binds polysaccharide molecules prior to cleaving them. For FBP, X-ray crystal structures have been obtained for both the unliganded ("apo", hence "open") and liganded states . In the case of T4L, we utilize the X-ray structure of a mutant version of this protein as the open state and the wild type [WT] X-ray structure as our template for the liganded state. For concreteness, we focus on the scenario in which the ligand is initially bound in the binding pocket, and the protein is in thermal equilibrium. At time $t=0$, the ligand dissociates from the binding pocket, and the protein, now out of equilibrium because the "perturbation" force supplied by the ligand has been removed, begins to relax towards its equilibrium structure in the absence of ligand. Thus, we consider the unliganded ("apo") form of

the protein as the final state. Using standard Hessian matrix techniques, we construct a quadratic approximation to the final state PES which encodes mechanical equilibrium at the appropriate protein conformation. Then we deduce the constant force perturbation which would shift the equilibrium configuration to that of the initial (liganded) state, thus mimicking the effect of the bound glutamate molecule on the protein to which it binds. This establishes initial and final state PESs, and hence the perturbation potential connecting them. We then couple the protein motion to the solvent in an approximate way, i.e., using a Langevin model (adding friction and random force to all the atoms in the system, with the atomic friction values chosen based on the solvent-accessible surface area of each atom). We show that the dynamical LRT prescription holds for such a dissipative system. Numerical solutions of the harmonic oscillator-Langevin Equation are presented for each of the two proteins noted above, in particular for the time-evolution of pairs of C-alpha atoms that lie along the protein's exterior (and hence could in principle be tagged in a FRET or ESR experiment). Insights are thereby gained into the time scales of the relaxation process, and the degree of validity of harmonic oscillator PESs for describing them.

In Section 4.3 of this paper we present the general theory of dynamic Linear Response Theory (LRT) for protein conformational changes induced by an appropriate perturbation potential, and show that it contains static LRT as a specific limit. In Section 4.4 we develop the details of dynamic LRT for two D-dimensional Harmonic Oscillator Hamiltonians related by a linear potential shift, corresponding to a constant perturbation force. For such a system dynamic LRT is exact for any perturbation force, however large. Explicit formulae for calculating the relevant relaxation dynamics are provided. Then, in Section 4.5, we illustrate the principles of dynamic LRT for the same multidimensional harmonic oscillator system, but under the additional influence of friction and a compensating fluctuating random force, i.e., the Langevin

Dynamics analog of the isolated Hamiltonian dynamical system considered in Section 4.4. We show that dynamic LRT is also exact for this system for any perturbation potential strength, and provide explicit formulae for computing the time evolution of the coordinate relaxation from equilibrium on the initial to the final state potential energy surfaces when the protein system is coupled to a heat bath in accord with the Langevin Equation In Section 4.6, we apply the principles developed in Sections 4.4 and 4.5 to study a simplistic model of the conformational change induced when Fe^{3+} bound to the appropriate pocket in the FBP protein escapes. In Section 4.7, we provide an analogous analysis T4L, treating the polysaccharide substrate molecule as the ligand. Finally, Discussion and Conclusions are presented in Section 4.8.

4.3 DYNAMIC LINEAR RESPONSE THEORY OF PROTEIN RELAXATION DYNAMICS: GENERAL THEORY

Consider a high dimensional system governed by the reference Hamiltonian $H_F = \mathrm{T} + V_F$, where T is the system's kinetic energy and V_F is the relevant multi-dimensional potential energy surface. Now imagine that a perturbation potential δV is applied to the system, which is then allowed to equilibrate in the perturbed potential, i.e., under $H_I = H_F + \delta V$. At time $t=0$ the perturbation is turned off. The classical phase space density describing the system will eventually relax back to equilibrium on V_F. [Note: The subscript "F" indicates final state, i.e., the one to which the system ultimately relaxes after the perturbation δV is removed. Analogously, the subscript "I" indicates initial state, i.e., the one in which the system equilibrates when the perturbation potential is present, up until t=0.] Consequently, the average of the phase space

function $A(\vec{x}(t), \vec{p}(t)) \equiv A(t)$, where $(\vec{x}(t), \vec{p}(t))$ denotes time evolution of the initial phase point $(\vec{x}, \vec{p})$ *under Hamiltonian* H_F, should relax to $\bar{A}_F$, the (static) equilibrium average on V_F:

$$\bar{A}_F = < A >_F \equiv \int d\vec{x} d\vec{p} e^{-\beta H_F(\vec{x},\vec{p})} A(\vec{x}, \vec{p}) / \int d\vec{x} d\vec{p} e^{-\beta H_F(\vec{x},\vec{p})} , \tag{4.1}$$

with $\beta \equiv (k_B T)^{-1}$. Let us denote by $\bar{A}(t)$ the non-equilibrium average of A of the initial phase space swarm of trajectories distributed according to $\exp(-\beta H_I)$, and evolving under H_F. Then, according to LRT, $\overline{\delta A}(t) \equiv \bar{A}(t) - \bar{A}_F$ evolves to lowest order in the perturbation strength as [43]

$$\overline{\delta A}(t) = -\beta < (A(t) - \bar{A}_F)(\delta V - \overline{\delta V}_F) >_F . \tag{4.2}$$

Note that for long times we expect $\overline{\delta A}(t) \to 0$.

To take a concrete example, let $\delta V = -\sum_j f_j x_j$, where the sum runs over all system coordinates x_j and the f_j's are constants. Further, choose the coordinate origin such that $< x_k >_F = 0$, i.e. the equilibrium position of x_k on V_F is 0. Then Equation 4.2 takes the form,

$$\overline{\delta A}(t) = \beta \sum_j < (A(t) - \bar{A}_F) x_j >_F \; f_j . \tag{4.3}$$

Specializing further to $A = x_k$ implies:

$$< x_k(t) >_I = \beta \sum_j < x_k(t) x_j >_F \; f_j . \tag{4.4}$$

Note that we require as input on the r.h.s. of Equation 4.4 a set of quadratic equilibrium time correlation functions. These are calculated entirely based on H_F: the classical dynamics is generated by this Hamiltonian, and the relevant phase-space average is taken over the relative

thermal equilibrium distribution $\exp(-\beta H_F)$. Note also that at $t=0$ a previously derived [59] static version of the LRT formula is recovered. That is,

$$< x_k(0) >_I = \beta \sum_j < x_k x_j >_F \; f_j \,, \tag{4.5}$$

where $< x_k(0) >_I$ prescribes the change in the equilibrium displacement of coordinate x_k on V_I relative the corresponding value on V_F, and the r.h.s. of the static LRT formula (Equation 4.5) shows how this change can be obtained from static equilibrium correlation functions associated with V_F only.

4.4 DYNAMIC LINEAR RESPONSE THEORY IS EXACT FOR A HARMONIC OSCILLATOR POTENTIAL ENERGY SURFACE WITH A LINEAR PERTURBATION POTENTIAL.

4.4.1 Formal Theory

While Equation 4.4 holds to first order in δV for any reference potential V_F, for reasons that will become clear below it is of particular interest to consider the multi-dimensional harmonic oscillator reference Hamiltonian,

$$H_F = \sum_j p_j^2 / 2m_j + \frac{1}{2} \vec{x} \cdot \boldsymbol{K} \cdot \vec{x} \,. \tag{4.6}$$

Here the index *j* runs over all D degrees of freedom. (Typically, we will be concerned with the motion of an N-atom molecule in three dimensions, in which case D=3N. For the sake of generality we will not make this specialization until specific applications are considered in

Section 4.6.) If the same system is subjected to a constant additional (perturbation) force, the Hamiltonian is then modified to

$$H_I = \sum_j p_j^2 / 2m_j + \frac{1}{2}\vec{x}\cdot\boldsymbol{K}\cdot\vec{x} - \vec{x}\cdot\vec{f}\ . \tag{4.7}$$

The subscript I designates this as the Hamiltonian, generated by adding the perturbation force (last term on the r.h.s. of Equation 4.7), that determines the *initial* thermal phase space distribution, $\exp(-\beta H_I)/Z_I$, with the normalization integral (partition function) Z_I given by $Z_I = \int d\vec{x}d\vec{p}\exp(-\beta H_I)$. Since $\vec{f}$ is a D-dimensional vector of constants, by completing the square Equation 4.7 can be written equivalently as

$$H_I = \sum_j p_j^2 / 2m_j + \frac{1}{2}(\vec{x}-\vec{x}_{eq})\cdot\boldsymbol{K}\cdot(\vec{x}-\vec{x}_{eq}) - \frac{1}{2}\vec{x}_{eq}\cdot\boldsymbol{K}\cdot\vec{x}_{eq}, \tag{4.8}$$

where $\vec{x}_{eq} = \boldsymbol{K}^{-1}\vec{f}$ gives the equilibrium positions of all degrees of freedom on potential energy surface V_I (i.e., H_I minus the kinetic energy term) relative to the analogous quantities on surface V_F. The quantity $\vec{x}_{eq}$ can be approximated via static LRT. Recalling that the elements of the positional two-point static correlation function are given for the Hamiltonian H_F in Equation 4.6 by $< x_j x_k >_F = \beta^{-1}[\boldsymbol{K}^{-1}]_{jk}$, then according to LRT (cf. Equation 4.5):

$$\beta(\beta^{-1}\boldsymbol{K}^{-1})\cdot\vec{f} = \boldsymbol{K}^{-1}\cdot\boldsymbol{K}\vec{x}_{eq} = \vec{x}_{eq}\ . \tag{4.9}$$

Thus, static LRT correctly predicts the equilibrium shifts induced by the perturbation in the model under consideration. In fact, it predicts the correct answer no matter how large $\vec{f}$ is.

Proceeding now to the time-dependent case, the 2D-dim. Gaussian phase space density corresponding to thermal equilibrium on H_I is the same as that for H_F except that the

Boltzmann distribution for H_I is centered in position space at the displacements $\vec{x}_{eq}$ noted above, whereas the Boltzmann distribution corresponding to H_F is centered at $\vec{x} = 0$. Designating the normalized thermal density distribution corresponding to H_I as $\rho_I(\vec{x}, \vec{p}) = \exp(-\beta H_I)/Z_I$, then the nonequilibrium average of one of the Cartesian components, say $x_j(t)$, is given by

$$< x_j(t) >_I = \int d\vec{x}_0 d\vec{p}_0 \rho_I(\vec{x}_0, \vec{p}_0) x_j(\vec{x}_0, \vec{p}_0, t), \tag{4.10}$$

where $x_j(\vec{x}_{0,} \vec{p}_0, t)$ is the time propagation of $x_j(t)$ according to Hamilton's Equations based on Hamiltonian H_F, starting from the initial phase point $(\vec{x}_0, \vec{p}_0)$.

The quantity $x_j(\vec{x}_{0,} \vec{p}_0, t)$ is formally determined as follows. The relevant Hamiltonian equations of motion (fully equivalent to Newton's 2nd Law here, but expressed as a set of coupled 1st order ODE's) are linear. Specifically,

$$\begin{pmatrix} \dot{\vec{x}} \\ \dot{\vec{p}} \end{pmatrix} = \begin{bmatrix} 0 & \boldsymbol{m}^{-1} \\ -\boldsymbol{K} & 0 \end{bmatrix} \begin{pmatrix} \vec{x} \\ \vec{p} \end{pmatrix} \equiv \boldsymbol{A} \begin{pmatrix} \vec{x} \\ \vec{p} \end{pmatrix}. \tag{4.11}$$

The formal solution of these EOMs is given by $\begin{pmatrix} \vec{x}(t) \\ \vec{p}(t) \end{pmatrix} = \boldsymbol{U}(t) \begin{pmatrix} \vec{x}_0 \\ \vec{p}_0 \end{pmatrix}$, with the 2Dx2D propagator matrix $\boldsymbol{U}(t) \equiv \exp(\boldsymbol{A}t)$. The essential points are i) $\boldsymbol{U}(t)$ is independent of initial conditions, and ii) it can be used to construct the time evolution of any initial conditions as a linear combination of those initial conditions. Namely,

$$x_j(\vec{x}_0, \vec{p}_0, t) = \sum_{k=1}^{D} [U_{j,k}(t) x_{k,0} + U_{j,k+D}(t) p_{k,0}]. \tag{4.12}$$

Substituting into Equation 4.10 above, and using the properties that $\int d\vec{x}_0 d\vec{p}_0 \rho_I(\vec{x}_0, \vec{p}_0) x_{k,0} = x_{k.eq}$ while $\int d\vec{x}_0 d\vec{p}_0 \rho_I(\vec{x}_0, \vec{p}_0) p_{k,0} = 0$, we obtain the following time evolution for the average position of coordinate *k*:

$$< x_j(t) >_I = \sum_{k=1}^{D} U_{j,k}(t) x_{k,eq} \quad , \tag{4.13}$$

where $< >_I$ indicates that the average is performed over the Boltzmann distribution corresponding to H_I (although, again, the indicated time evolution of x_j is carried out on the PES $V_F = \vec{x} \cdot \boldsymbol{K} \cdot \vec{x} / 2$). An explicit prescription for $U_{jk}(t)$ in terms of vibrational normal mode eigenvalues and eigenvectors is given below.

Now, let us examine the related equilibrium fluctuations associated with H_F, as detailed on the r.h.s. of Equation 4.4 above, with $\vec{f} = \boldsymbol{K}\vec{x}_{eq}$. The key quantity that needs to be calculated here is $< x_k(t) x_j >_F$, where Hamiltonian dynamics is evolved on H_F *and* the relevant phase space density for performing the average is the classical Boltzmann distribution corresponding to H_F, i.e., $\rho_F(\vec{x}_0, \vec{p}_0) = \exp(-\beta H_F) / Z_F$, with $Z_F \equiv \int d\vec{x} d\vec{p} \exp(-\beta H_F)$. We then have

$$< x_k(t) x_j >_F = \int d\vec{x}_0 d\vec{p}_0 \rho_F(\vec{x}_0, \vec{p}_0) x_k(\vec{x}_{0,} \vec{p}_0, t) x_{0,j} , \tag{4.14}$$

where $x_{j,0}$ is the jth component of $\vec{x}_0$. Using the expression for $x_j(\vec{x}_0, \vec{p}_0, t)$ in Equation 4.12 above, and the properties that $< x_j p_k >_F = 0$ and $< x_j x_k >_F = \frac{1}{\beta}[\boldsymbol{K}^{-1}]_{jk}$, we find that

$$< x_j(t) x_m >_F = \frac{1}{\beta} \sum_{k=1}^{D} U_{j,k}(t) [\boldsymbol{K}^{-1}]_{km} . \tag{4.15}$$

Combining appropriate equations above, we find

$$\beta\sum_{j=1}^{D} < x_n(t)x_j >_F \; f_j = \sum_{k=1}^{D} U_{n,k}(t)\left[\boldsymbol{K}^{-1}\boldsymbol{K}\boldsymbol{x}_{eq}\right]_k = \sum_{k=1}^{D} U_{n,k}(t)x_{k,eq} \tag{4.16}$$

Comparing Equations 4.13 and 4.16, we see that the dynamic LRT relation given in Equation 4.4 is indeed exactly satisfied for arbitrary $\vec{f}$. As will be shown in Section 4.5 below, the same demonstration of the connection of non-equilibrium relaxation dynamics based on ρ_I and equilibrium fluctuations based on ρ_F can be carried out in the case that time evolution proceeds via the analogous Langevin Equation.

4.4.2 Explicit Prescription for Computing Relaxation to Equilibrium on a Harmonic Oscillator PES

We wish to compute the quantity $< \vec{x}(t) >_I$, where the classical (Hamiltonian) dynamics is carried out on H_F and the average is over the Boltzmann distribution relevant to H_I. To do this, we decompose the mass-weighted Hessian matrix $\boldsymbol{W} \equiv \boldsymbol{m}^{-1/2}\boldsymbol{K}\boldsymbol{m}^{-1/2}$, where $\boldsymbol{m}$ is the (diagonal) mass matrix, in the usual manner, i.e.,

$$\boldsymbol{W} = \boldsymbol{T}\begin{pmatrix} \omega_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \omega_D^2 \end{pmatrix}\boldsymbol{T}^t. \tag{4.17}$$

Here the $D \times D$ matrix $\boldsymbol{T}$ contains the D unit-normed eigenvectors of $\boldsymbol{W}$, one per column. $\boldsymbol{T}$ is an orthogonal matrix, and thus $\boldsymbol{T}^{-1} = \boldsymbol{T}^t$; furthermore, $\omega_j^2 \geq 0$ is the eigenvalue corresponding to the eigenvector in the j'th column of $\boldsymbol{T}$.

Using the eigenvectors and eigenvalues prescribed in Equation (4.17), one determines that

$$< \vec{x}(t) >_I = \boldsymbol{m}^{-1/2}\boldsymbol{T}\begin{pmatrix} \frac{\cos(\omega_1 t)}{\omega_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\cos(\omega_D t)}{\omega_D^2} \end{pmatrix}\boldsymbol{T}^t\boldsymbol{m}^{-1/2}\,\vec{f}\ . \tag{4.18}$$

Note the reduction at $t = 0$:

$$< \vec{x}(0) >_I = \boldsymbol{m}^{-1/2}\boldsymbol{T}\begin{pmatrix} \frac{1}{\omega_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\omega_D^2} \end{pmatrix}\boldsymbol{T}^t\boldsymbol{m}^{-1/2}\,\vec{f} = \boldsymbol{m}^{-1/2}\boldsymbol{W}^{-1}\boldsymbol{m}^{-1/2}\,\vec{f} = \boldsymbol{K}^{-1}\,\vec{f}\ ,$$

as expected.

4.5 LANGEVIN EQUATION ANALOG: RELAXATION ON A HARMONIC OSCILLATOR PES

For motion according to the harmonic oscillator Hamiltonian H_F introduced in Section 4.4, the relevant Langevin Equation reads:

$$\begin{pmatrix} \dot{\vec{x}} \\ \dot{\vec{p}} \end{pmatrix} = \begin{bmatrix} 0 & \boldsymbol{m}^{-1} \\ -\boldsymbol{K} & -\boldsymbol{\gamma}\boldsymbol{m}^{-1} \end{bmatrix}\begin{pmatrix} \vec{x} \\ \vec{p} \end{pmatrix} + \begin{pmatrix} 0 \\ \vec{R}(t) \end{pmatrix}\ , \tag{4.19}$$

where $\boldsymbol{\gamma}$ is the relevant DxD friction constant matrix and $\vec{R}(t)$ is a vector of random force components imbued with the appropriate statistical properties [105, 123, 124] . As above, we wish to compute the time development of the coordinate x_k on H_F, given the initial phase space density $\rho_I(\vec{x}, \vec{p})$. In fact, the time evolution of this entire phase space density on H_F , i.e.,

$\rho_I(\vec{x},\vec{p},t)$, can be calculated in a straightforward manner. One finds [105] that the initial Gaussian density remains Gaussian, with the expectation values of position and momentum given by the solution of the Langevin EOMs in Equation 4.19 in the zero temperature (T=0) limit, i.e., setting $\vec{R}(t)=0$. These equations then have the same generic structure as the Hamiltonian equations in Equation 4.11. Specifically,

$$\begin{pmatrix}\dot{\vec{x}}\\ \dot{\vec{p}}\end{pmatrix}=\begin{bmatrix}0 & \boldsymbol{m}^{-1}\\ -\boldsymbol{K} & -\boldsymbol{\gamma m}^{-1}\end{bmatrix}\begin{pmatrix}\vec{x}\\ \vec{p}\end{pmatrix}=\boldsymbol{A}'\begin{pmatrix}\vec{x}\\ \vec{p}\end{pmatrix}. \tag{4.20}$$

Equation 4.20 has the formal solution $\begin{pmatrix}\vec{x}(t)\\ \vec{p}(t)\end{pmatrix}=\boldsymbol{U}'(t)\begin{pmatrix}\vec{x}_0\\ \vec{p}_0\end{pmatrix}$, with $\boldsymbol{U}'(t)\equiv\exp(\boldsymbol{A}'t)$. The time evolution of any given initial conditions can then be expressed as a linear combination of those initial conditions:

$$x_j(\vec{x}_0,\vec{p}_0,t)=\sum_{k=1}^{D=3N}[U'_{j,k}(t)x_{k,0}+U'_{j,k+D}(t)p_{k,0}], \tag{4.21}$$

where, as in the Hamiltonian dynamics case, the coefficients $U'_{jk}(t)$ are independent of initial conditions. Using the same averaging analysis, we obtain:

$$<x_j(t)>_I=\sum_{k=1}^{D}U'_{j,k}(t)x_{k,eq}\ . \tag{4.22}$$

Note as a practical aside that this provides us with a straightforward way to compute $<x_j(t)>_I$. Namely: i) Determine the 2D eigenvalues and eigenvectors of $\boldsymbol{A}'$; ii) synthesize $\boldsymbol{U}'(t)\equiv\exp(\boldsymbol{A}'t)$ from these eigenvalues/vectors; iii) then:

$$<x_j(t)>_I=\left[\boldsymbol{U}'(t)\begin{pmatrix}\vec{x}_{eq}\\ 0\end{pmatrix}\right]_j\ . \tag{4.23}$$

This procedure is implemented numerically in Section 4.6 below.

Next, we consider fluctuations about equilibrium on H_F. The essential quantity that we need to compute is again $< x_k(t)x_j >_F$ except now $x_k(t)$ is the evolution of x_k from the initial phase point $(\vec{x}_0, \vec{p}_0)$ based on the Harmonic oscillator Langevin Equation above, and the average includes both an average over initial phase points *and* the random noise vector $\vec{R}(t)$. The formal solution of the Langevin Equation in the case of a non-zero random noise vector, which is obtained by integrating Equations 4.19, is of the form:

$$x_j(\vec{x}_0, \vec{p}_0, t) = \sum_{k=1}^{D}\left[U'_{j,k}(t)x_{k,0} + U'_{j,k+D}(t)\, p_{k,0} + \int_0^t dt' D_{jk}(t-t')R_k(t') \right], \qquad (4.24)$$

where the $D_{jk}(t)$ can be calculated from the eigenvalues/vectors of $\boldsymbol{A}'$ if desired. Full details are not needed here: one simply has to note that the $D_{jk}(t)$ are *in*dependent of initial phase space values and the specifics of the random force vector function $\vec{R}(t)$. Now, we consider the average $< R_j(t)x_k >_F$ where the average over system coordinate x_k is taken over the canonical Boltzmann distribution based on H_F, and the average over the random force function $R_j(t)$ is taken over realizations of this noise function with properties prescribed by Langevin theory. We find that $< R_j(t)x_k >_F =< R_j(t) >< x_k >_F = 0$ (in fact, both factors in the factorized correlation function evaluate to 0), and thus the third (inhomogeneous) term in Equation 4.24 does not contribute to the two point correlation functions of interest. Only the 1st two terms on the r.h.s. of Equation 4.24 are relevant, and these are computed from the T=0 solution of the Langevin Equation, i.e., the friction-damped relaxation of the initial phase point $(\vec{x}_{eq}, 0)^t$ given above. From this point on, the derivation follows that of the zero-friction case (ordinary Hamilton's Equations; cf. Equation 4.11) exactly, leading to the equality:

$$< x_n(t) >_I = \beta \sum_{j=1}^{D} < x_n(t) x_j >_F \; f_j \; . \tag{4.25}$$

as in the case of non-dissipative Hamiltonian mechanics. Also as in the case of Hamiltonian mechanics, the equality indicated in Equation 4.25 holds for an arbitrarily large perturbation force $\vec{f}$.

At first glance it may seem purely coincidental that there exists a version of dynamic LRT for a multi-dimensional harmonic oscillator system coupled to a heat bath via the Langevin Equation which is completely analogous to the version formulated in Section 4.4 for the same, isolated, harmonic oscillator system (not coupled to a heat bath, and evolving according to deterministic Hamilton's Equations). There is, however, a more direct connection between the two. If the protein "system" is bilinearly coupled to a large set of auxiliary harmonic oscillator coordinates (a harmonic oscillator "bath") in an appropriate fashion, and the entire system+bath super-system is then evolved via Hamilton's Equations, the motion of the system coordinates can be shown to obey a Generalized Langevin Equation (GLE) with memory friction kernel determined by the detailed bilinear coupling coefficients. Since the overall super-system is of harmonic oscillator type and is time evolved as an isolated Hamiltonian system, the results of Section 4.4 apply directly. But because the system motion obeys a GLE, these results can be cast into the form arrived at in the present subsection. Full details of this equivalence are provided in the Appendix C.

4.6 NUMERICAL EXAMPLE 1: AN ELEMENTARY MODEL OF CONFORMATIONAL CHANGE IN THE FERRIC BINDING PROTEIN (FBP) INDUCED BY LIGAND DISSOCIATION FROM A BINDING POCKET.

We wish to provide a tractable example that illustrates some of the basic principles of dynamic LRT. Since all-atom MD of large proteins in H_2O solvent is extremely time-consuming, with maximum time scales attainable on extant computational platforms being 10s of ns, we invoke here a harmonic oscillator (quadratic PES) model of the relaxation of a globular protein from one (crystallized) stable state to a second one upon unbinding of an appropriate ligand from the binding pocket in the protein. A harmonic oscillator Hamiltonian can be evolved to arbitrarily long times in a relatively painless fashion, i.e., by standard linear algebraic manipulations performed on finite dimensional matrices. As part of this computational strategy, we will treat the solvent implicitly, i.e. as a source of friction and random buffeting that influences motion of the explicitly treated mechanical degrees of freedom, which are the atoms of the protein. Thus, for a protein consisting of N atoms, there are 6N dynamical phase space degrees of freedom, and this is the dimensionality of the relevant matrices needed for time propagation.

Consider the bacterial Ferric Binding Protein (FBP), which is used by pathogenic bacteria to absorb iron from the host environment. This protein, comprised of 309 residues, provides an easily appreciated example of a ligand-binding induced protein conformational transition. Furthermore, X-ray crystal structures have been obtained for both ligand-bound and apo states. Thus we know the equilibrium conformations in both states (after minor refinements described below), and therefore have a reasonable way to construct Hamiltonians that approximately govern fluctuations around both equilibrium configurations, as discussed in detail below.

In the present paper we will utilize the X-ray structure FBP in the apo ("open") and in the liganded ("closed") state. X-ray structures for ligand-free (PDB code: 1D9V) and Fe^{3+}-bound (PDB code: 1MRP) forms were downloaded from the Protein Data Bank. Both structures were energetically minimized using the AMBER 94 force field. in vacuum (no solvent molecules were included in the minimization process) by gradually decreasing harmonic restraints on the backbone atoms, using a mixed steepest descent and conjugate gradient minimization protocol, until an RMS gradient on the order of 10^{-5} kcal/(mol·Å^2) was reached. No cutoff distance for non-bonded interactions was enforced. For the closed state structure, the ligand, i.e. an Fe^{3+} ion, was also removed. (Including the Fe^{3+} ion bound inside the pocket was found to have little effect on the final geometric configuration.) Then these two structures were superimposed by standard fitting techniques [98].

As noted above, we will consider the ligand-free (apo) form as the final state. A perturbation potential supplied by the bound Fe^{3+} ion ligand shifts the equilibrium conformation to that associated with the ligand-bound conformation, which serves as the initial state of our model. At *t=0*, the ligand dissociates from the binding pocket, and the protein begins to relax to the final state conformation. (See Figure 18 for a schematic depiction of this process.) Hence we consider the apo structure as our final state, i.e., the one upon which dynamics occurs subsequent to *t=0*. We model V_I, V_F as multi-dimensional harmonic oscillator PESs that describe fluctuations of all the protein atoms about the equilibrium positions implied by the initial (ligand bound) and final (apo) state structures, respectively. We base the harmonic oscillator PES on a local quadratic expansion of the AMBER PES about the mechanical equilibrium positions attained by the atoms after "quenching" to a local minimum of the full (anharmonic) AMBER PES, starting from the appropriate X-ray crystal structure. In order to apply the form of

dynamic LRT developed in Sections. 4.4 and 4.5 above, we need to select one of the two equilibrium configurations, apo or ligand-bound, and compute the 2nd derivative matrix $\boldsymbol{K}$ via local quadratic expansion on the AMBER 94 force field [97]. In order to apply the mathematical models developed in Sections. 4.4, 4.5 above, we will then adopt the same 2nd derivative matrix for the other configuration. Again, the two relevant PESs will be identical except that one will be shifted relative to the other in coordinate space based on the X-ray structure coordinates (suitably quenched as indicated above). In the present work we have chosen to compute **K** for the *apo* form of FBP.

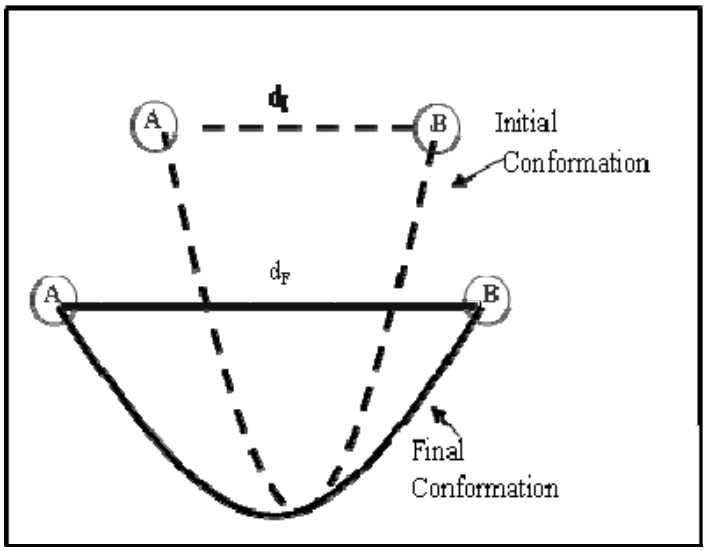

Figure 18. Schematic illustration of a protein conformation change from the Initial (dashed line) to Final (solid line) equilibrium structure ("state"). The change is induced by removal of an applied perturbation force, such as that generated by a ligand molecule introduced into a binding pocket of the protein prior to t=0. In the process of the indicated conformational change the distance between two residues A and B on the protein evolves in time from d_I to d_F.

V_I then has the same general form (characterized by the same 2^{nd} derivative matrix), but expanded about a different mechanical equilibrium configuration, namely the one obtained by quenching to a local equilibrium conformation using the full (anharmonic) AMBER PES starting from the ligand-bound X-ray structure. We shall denote this set of coordinates as $\vec{x}_{eq}$ (measured w.r.t. the corresponding equilibrium position values on V_F). Note that this procedure directly determines the effective linear perturbation potential which would have to be exerted by a ligand bound in the initial state in order to generate the desired final state equilibrium configuration, given the posited quadratic fluctuation potentials assumed for the two PESs. We thereby

sidestep the problem of explicitly calculating the force supplied by the bound Fe^{3+} ligand on the protein, which would make an interesting subject for further work on the applicability of the dynamic LRT approach for FBP and other systems.

Our goal is to carry out Langevin dynamics based on the system potential V_F. Thus, we need to specify a friction constant matrix for motion on this PES. To do this in the simplest possible fashion, we assume that the matrix of friction coefficients is diagonal, with numerical values determined by applying Stokes' formula [41] to the individual atoms. Thus, for Cartesian displacement coordinate i ($1 \le i \le 3N$), $\gamma_{i,i} = 6\pi\eta a_i$ where η is the solvent viscosity, and a_i is the hydrodynamic radius associated with coordinate i (for each atom, all three Cartesian coordinates are characterized by the same hydrodynamic radius); furthermore, $\gamma_{i,j} = 0$ for $i \neq j$. Bearing this in mind, in our calculations we will consider a range of solvent viscosities in order to better understand the connection between this parameter and the relaxation time scales computed for the protein.

The hydrodynamic radii of the atoms in the protein were determined by the relative degree of surface exposure of each atom to the solvent. The NACCESS molecular surface program [125] 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 ascertained, the hydrodynamic radius of each atom was scaled to these fully exposed atomic radii using the square root of the fraction of surface accessible area obtained via the NACCESS program. For hydrogen atoms only, the hydrodynamic radius was set to a maximum value of 0.2 Å if the resulting effective radius was larger than this value.

After constructing the matrix $\mathbf{A}'$ from the second potential derivative matrix and the friction matrix described above, the propagator matrix $\mathbf{U}'(t)$ can be synthesized at any time from the eigenvalues and eigenvectors of $\mathbf{A}'$ using standard procedures for matrix exponentiation. (Equivalently, the entire computation can be carried out in mass-weighted coordinates: explicit details of this implementation are given in Ref. [126].) The propagator matrix can then be utilized to obtain $< x_j(t) >_I$, representative values of which are presented in the results section below.

For purposes of analysis (again, see below), it is useful to work with a friction matrix which can be inverted. Because the entire friction matrix scales linearly with the solvent viscosity, we set the latter temporarily to 1 cP, and thus focus on the intrinsic geometry of the protein (in particular, the SASA of each atom). Atoms which are completely shielded from water solvent by other atoms in the protein, and hence have zero SASA and a friction constant of 0, are re-assigned a non-zero friction constant whose value is 1000 times smaller than the smallest non-zero element of the original $\eta = 1$ cP friction matrix. This proportionality is maintained for arbitrary values of the solvent viscosity simply by re-scaling the entire friction matrix accordingly. We checked that this modification (i.e., replacing zero on-diagonal elements of the friction matrix with tiny non-zero values) had no effect on the results for $< x_j(t) >_I$ obtained from Equation 4.23 for any value of the viscosity considered in the numerical results presented below.

4.6.1 Results

We selected several amino acid pairs and monitored their motion after the perturbation potential was turned off at *t=0* (when the initial phase space density function $\exp(-\beta H_I)/Z_I$ began to evolve on the PES V_F). We first considered the case of zero friction (solvent viscosity $\eta = 0$), i.e., we carried out Hamiltonian dynamics on the protein system in vacuum. Since there are $3N = 14376$ degrees of freedom in the protein, one might expect that the motion of a single atomic coordinate (or the displacement between atomic coordinates on two chosen atoms), being a linear combination of many vibrational normal modes of the system which are characterized by incommensurate frequencies, would dephase to zero at long times. However, we found that this did not happen, i.e., the displacement coordinates continued to oscillate in time ("ring") *ad infinitum*.

In order to ascertain why, we examined the degree of participation of various normal modes in the motion of the oscillator system. If we denote the jth unit-normalized eigenvector of the mass-weighted hessian matrix as $\vec{\nu}_j$ and its corresponding eigenvalue as ω_j^2 (cf. Equation 4.17 in section 4.4.2), then for initial conditions corresponding to $\dot{\vec{x}}(0) = 0$ (no initial velocity), the exact time propagation of the system is given by: $\vec{y}(t) = \sum_{j=1}^{3N} b_j \cos(\omega_j t)\vec{\nu}_j$. Here $\vec{y}$ is the vector of mass-weighted coordinates defined by $\vec{y}(t) = \mathbf{m}^{1/2}\vec{x}(t)$ where $\mathbf{m}$ is the (diagonal) mass matrix. The superposition coefficients b_j, which prescribe the degree to which mode j participates in the motion of the oscillator system for a given set of initial displacements, can easily obtained using the orthonormality of the Hessian eigenvectors, namely $b_j = \vec{\nu}_j \cdot \vec{y}(0)$. For a PES based on a local quadratic Taylor's series expansion about mechanical equilibrium of an

isolated protein molecule, one always finds 6 zero modes corresponding to zero frequency, i.e., 3 modes corresponding to center of mass translation and 3 modes corresponding to overall rotations. Motion along these modes does not influence the relative internal protein motion, and thus they can be disgarded from the present analysis. Focusing on the remaining 3N-6 vibrational modes (all characterized by $\omega_j^2 > 0$), we first identified the mode with the maximum value of $|b_j|$. Then, each mode for which $|b_j|$ was greater than 5% percent of the maximal value, for a particular *j*, was considered "significant". For FBP, we obtained a participation number of ca. 90, i.e., relatively few out of >14,000 normal modes in the overall superposition. Hence, within the harmonic oscillator model adopted here an additional feature, namely, frictional damping, needs to be added to generate irreversible relaxation to the desired final state configuration..

When frictional damping is included, the relaxation dynamics is computed via Equation 4.22. Note that because of the exactness of the LRT formula Equation 4.25 for the system under consideration here, evaluation of the equilibrium time correlation functions on the r.h.s. of Equation 4.25 (as would be done in more general implementations of dynamic LRT) would lead to the same prediction for the time dependence of the mean displacements $< x_j(t) >_I$. The time evolution of the distance between selected pairs of residues now proceeds in a qualitatively reasonable fashion, as illustrated in Figure 19 for solvent viscosity $\eta = 1$ cP.

Three residue pairs, depicted in Figure 19a, were selected for detailed study. All three pairs have the characteristic that each member of the pair is on the exterior (solvent exposed) surface of the protein. Presumably, each of these residues could be tagged by chemically attaching a FRET fluorophore, since such an attachment would not be sterically hindered. The separation distance for the C-alpha atoms of the residue pair E167-K237, which is depicted in

red in Figure 19a, and for which the computed distance vs. time curve is shown in Figure 19b, can be taken as a measure of the domain opening motion since the residues are in two different domains and the distance between them changes significantly in going from ligand-bound to apo state. The computed distance vs. time curve for the C-alpha atoms of the residue pair G33-T110 (orange pair in Figure 19a) is displayed in Figure 19c. Again, these two residues are in two different domains, and their separation distance monitors the opening motion around the ligand binding site. The distance in the closed conformation is around 36.5 Å while the same distance is around 40.5 Å in the open conformation. The magnitude of the change in distance is almost the same as for the residue pair in Figure 19b, although the absolute distance between two residues is much smaller. The final residue pair N184-D270 is indicated in yellow in Figure 19a. The time evolution of the distance between the C-alpha atoms of these two residues is displayed in Figure 19d. This distance changes in going from the initial to the final state configuration by ca. 1.5 Å, which is much smaller than the distance changes observed in Figure 19b and Figure 19c. Furthermore, the distance between this pair actually *de*creases in going from ligand bound to the ligand free state. The detailed time-evolution of each inter-AA distance investigated, which, again, could be probed by FRET or ESR measurements, is different. Nevertheless, the overall time scale of all three AA pair distances lies in the ~200-300 picosecond range. This time scale is consistent with previous computational studies of Langevin normal modes in proteins [106, 127], which reported time scales of collective motions in the 10ps-1ns range, although to our knowledge ours is the first study in which the time dependence of inter-residue distances in a protein undergoing conformational transformation as a result of ligand dissociation has been directly calculated via harmonic oscillator level Langevin dynamics.

To the best of our knowledge real-time relaxation of FBP upon dissociation of the Fe^{3+} ligand from its binding pocket has not yet been monitored at the single molecule level via experiment. Below we discuss experimental measurements and relevant all-atom MD studies of binding cleft conformational changes in other proteins of roughly similar size below. Before doing so, it behooves us to consider a second numerical example, namely, T4 lysozyme (T4L), which we analyzed in the same manner as FBP.

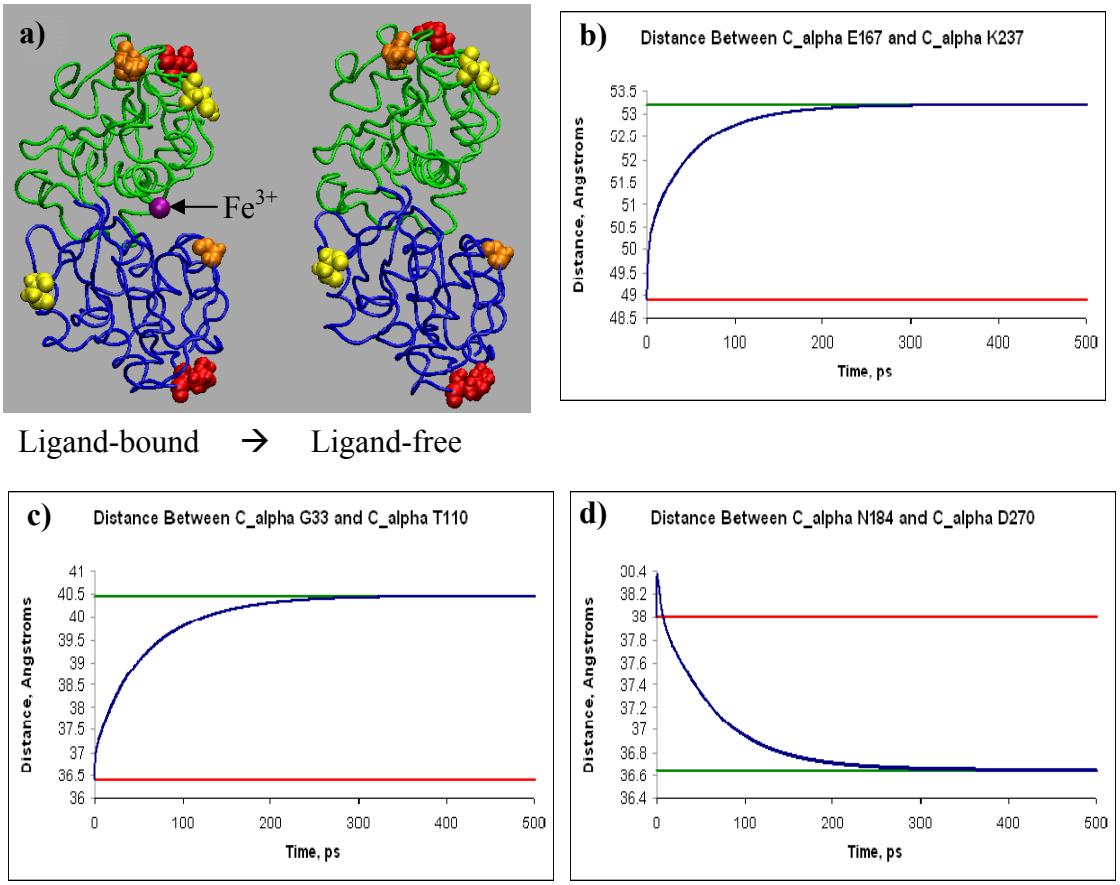

Figure 19. (FBP) Panel a) depicts 3 pairs of residues whose separation distance vs. time is monitored via Langevin dynamics model described in detail in the text. The two domains which close and open around ligand binding site is colored via blue and green (Domains are determined via DYNDOM program [128]).Temporal relaxation of the selected distance pairs is shown in panels b-d. In each panel, green line is the inter-residue separation distance in the ligand free state. Red line is the analogous distance in the ligand-bound state. Blue line shows the propagation via the Langevin Eq, with a solvent viscosity of 1 cP. Specifically: b) Red pair of C alpha atoms indicated in a) c) Orange pair of C alpha atoms indicated in a) d) Yellow pair of C alpha atoms indicated in a).

4.7 NUMERICAL EXAMPLE 2: AN ELEMENTARY MODEL OF CONFORMATIONAL CHANGE IN T4L INDUCED BY SUBSTRATE DISSOCIATION FROM A BINDING POCKET.

The enzyme T4L is a 164-residue globular protein with two domains which binds polysaccharide molecules prior to cleaving them. When substrate is covalently bound to the cleft in between two domains, the substrate-bound enzyme is locked in a state in which two domains have closed around the substrate with respect to the unbound state. An X-ray crystal structure of the WT protein has been obtained. In addition, there are a number of extant X-ray structures for T4L molecules that have been selectively mutated in such a way the cleft of the crystallized structure is considerably more open than in the WT structure. These structures are hypothesized to be similar in geometric detail to that of wild type T4L substrate in the *absence* of bound ligand. Thus we will use one of these mutated T4L X-ray structures as our model of the ligand-free form of the protein and the WT X-ray structure as an approximate configuration of the ligand-bound enzyme (cf. Figure 20a, in which the ligand-bound form is the one obtained after energy minimization and ligand-free form of the protein mutant is X-ray structure).

The crystallographic WT state (PDB entry 2LZM) was energy minimized in vacuum. We used decreasing harmonic restraints on the backbone atoms using a mixed steepest descent and conjugate gradient minimization protocol, until an RMS gradient on the order of 10^{-5} kcal/(mol·Å^2) was reached. No cutoff distance for non-bonded interactions was enforced. All the minimizations were carried out with the AMBER 94 force field

The choice of which surface to perform the Hessian evaluation on was dictated by our desire to focus on the protein without any mutations. We chose to explicitly calculate the force constant matrix for the ligand-bound form (WT structure), and then use this **K** matrix to describe

harmonic fluctuations about the apo state equilibrium configuration as well. (This necessitated the minimization of the WT X-ray structure under the AMBER force field, as described above.) Again, the two PESs surfaces used in our multidimensional harmonic oscillator model are identical except that one is shifted relative to the other in coordinate space based on the X-ray structure coordinates. In the case of T4L the equilibrium configuration of the closed structure is given by the end product of the minimization of the relevant X-ray structure (using the protocol described above), while that of the open structure is taken for simplicity to be the X-ray crystal structure coordinates of the mutant T4L selected to represent the WT protein structure. In particular, the PDB structure 150L fourth conformer is used for the final (apo, i.e. open) structure by computationally mutating the single M6I mutation back to the appropriate WT residue (M). The hinge bending angle in this structure is ca. 35 degrees greater than in the energy-minimized WT structure (cf Figure 20a) (as compared to a change of 30.5 degrees for this angle when the X-ray structure of the WT protein is employed). Note: The two structures are first superimposed by standard fitting techniques [98] and residues 163 and 164 are excluded from the analysis, since their coordinates are absent in the mutated X-ray structure.

4.7.1 Results

As in the FBP case, we first considered the case of zero friction (solvent viscosity $\eta = 0$), i.e., we carried out Hamiltonian dynamics on the protein system in vacuum and monitor the motion of several amino acids. The motion of a single atomic coordinate (or the displacement between atomic coordinates of two chosen atoms) did not dephase to zero at long times. Thus we checked the degree of participation of various normal modes in the motion of the oscillator system as we

did in FBP case: again, ca. 90 normal modes out of overall 3N=7830 participate significantly in the cleft opening process in our T4L model.

In order to generate irreversible relaxation to the desired final state configuration, frictional damping is included. The friction matrix was calculated via the same protocol as in FBP case: the NACCESS program was used to calculate the surface exposure of each atom by using the coordinates of the apo(open) structure.

With frictional damping included, the relaxation dynamics was calculated using Equation 4.22. The time evolution of the distance between selected pairs of residues is illustrated in Figure 20b-d for solvent viscosity $\eta = 1$ cP. When choosing the residue pairs to study, the two criteria, namely (i) being exposed to solvent and (ii) not being next to bulky side chains (and hence being a viable candidate for tagging in a possible FRET experiment), are considered as was the case in our FBP study. Figure 20b shows the relaxation dynamics of the distance between the C-alpha atoms of the E22-R137 pair. These two residues are in very close contact when the structure is in its ligand-bound state, and there is ca. 11 Å increase in distance going from closed to open state of the protein. This distance can be taken as a measure of opening of the two domains (cf. Figure 20a for an illustration of the domain opening motion). The residue pair R52-A160, for which the corresponding relaxation dynamics of the C-alpha atoms of these residues is indicated in Figure 20c, behaves somewhat differently. In going from ligand-bound to ligand-free state, the distance between these two residues decreases. The time evolution of the distance between the C-alpha atoms of the final residue pair studied, namely, S38-K135, is displayed in Figure 20d. It behaves like the residue pair in Figure 20b, except that the absolute distance between the two residues in Figure 20d is much larger than for those in Figure 20b, since the two residues "tagged" in Figure 20d lie on the outermost surface of the protein away from the ligand binding

cleft. Nevertheless, the change in the distance between this residue pair can be taken as a measure of the degree of the conformational change of the protein, i.e., the degree of domain opening. In spite of the different locations of the chosen pairs, their time scales of relaxation are quite similar. Comparing to the relaxation times extracted for FBP (cf. Figure 19), the relaxation is almost 3 times faster for T4L. Indeed, the relaxation times predicted for T4L by our harmonic oscillator Langevin model are similar to suggested time scales (several hundred picoseconds) extracted from previous MD simulations of the same protein by de Groot et al.[129] for related (though not identical) cleft opening and closing events. It is also intuitively reasonable, since the moving domains of T4L are smaller than those of FBP, and thus not only easier to move, but easier to bring together as the binding cleft closes.

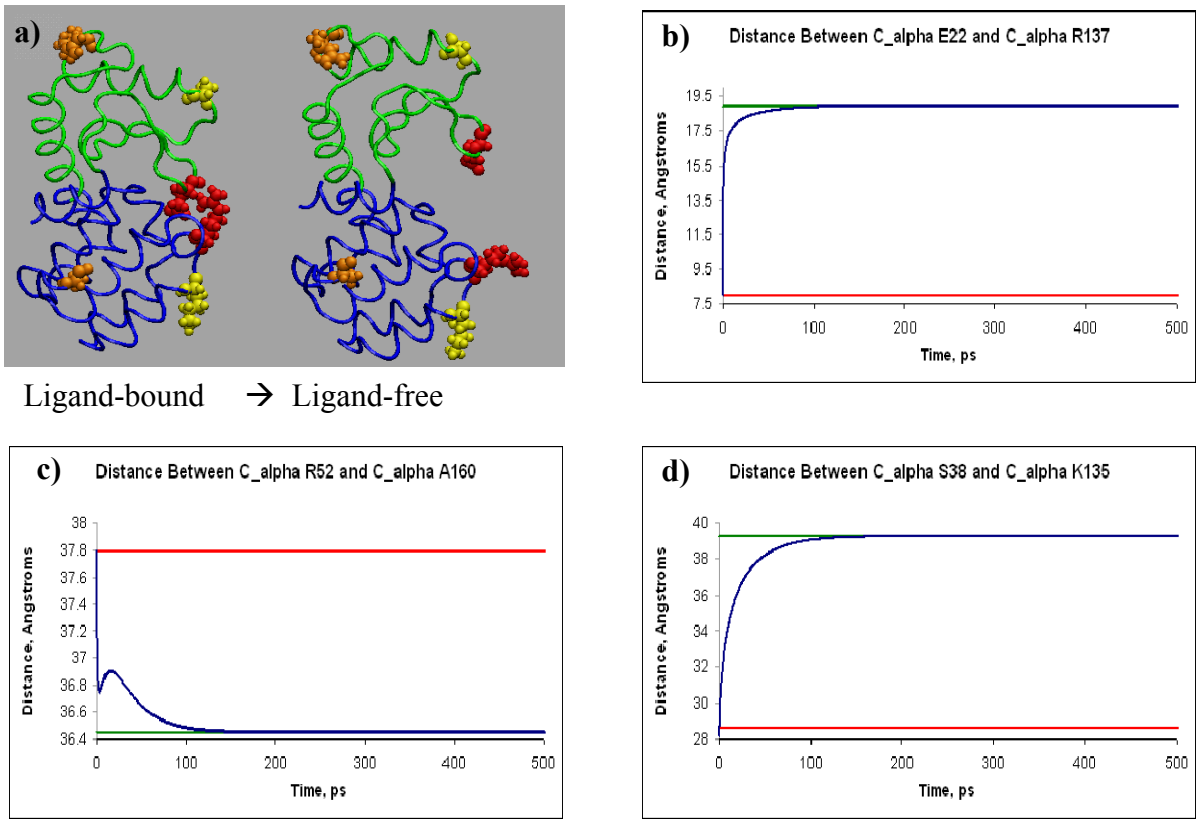

Figure 20. (T4L) Panel a) depicts 3 pairs of residues whose separation distance vs. time is monitored via Langevin dynamics model described in detail in the text. The two domains which close and open around ligand binding site are colored via blue and green (Domains are determined via DYNDOM program[128]). Temporal relaxation of the selected distance pairs is shown in panels b-d. In each panel, green line is the inter-residue separation distance in the ligand free state. Red line is the analogous distance in the ligand-bound state. Blue line shows the propagation via the Langevin Eq, with a solvent viscosity of 1 cP. Specifically: b) Red pair of C alpha atoms indicated in a). c) Orange pair of C alpha atoms indicated in a). d) Yellow pair of C alpha atoms indicated in a).

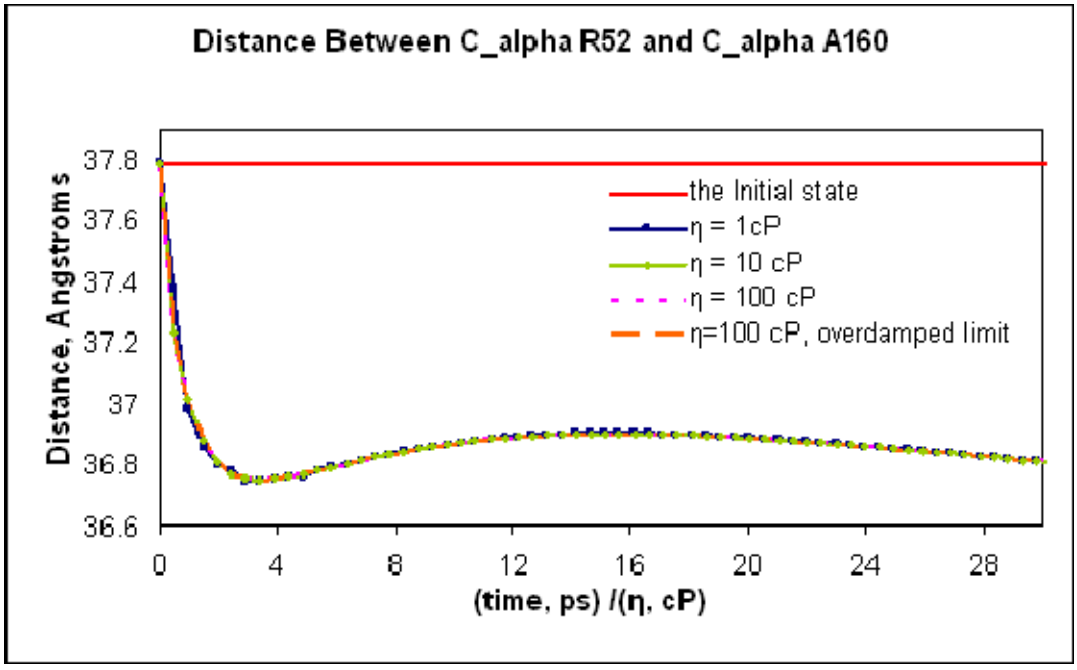

Figure 21. (T4L) Time evolutions of the separation distance of C_alpha_R52 and C_alpha_A160 propagated according to Langevin Equation in solvents characterized by 3 different viscosities. Red line is the inter-residue distance in the ligand-bound state. The abscissa plots time normalized by system viscosity. For orange dashed line the calculation is carried out with overdamped relaxation equations in 3N-dim. configuration space. Details are given in the text.

To provide insight into the relationship between input parameters (e.g., solvent viscosity and details of the protein Hessian matrix corresponding to V_F) and computed relaxation times we performed a range of numerical studies with our model. In particular, we varied the solvent viscosity from 1 cP to >>1 cP. A sample of comparative results is presented in Figure 21 focusing on the distance between the R52-A160 pair of T4L that was studied for the viscosity $\eta = 1cP$ in Figure 20c. Note that the AA-pair distance vs. time curve changes only by a dilatation of the time coordinate (i.e., doubling the viscosity doubles the decay time) as the solvent viscosity is further increased, all other parameters being held constant. (The low viscosity case, i.e. $\eta = 1cP$, shows slight deviation from this behavior, which is, however, difficult to discern to the resolution of Figure 21.)

The behavior noted above suggests that a high friction analysis of the Langevin Equation may prove fruitful. In this limit, the acceleration ("p-dot") term in the Langevin Equation can be set to zero, and the time-dependent expectation value of the displacement coordinates (cf. Equation 4.20) is determined by the equation $< \vec{x}(t) >_I = \exp(-\boldsymbol{\gamma}^{-1}\boldsymbol{K}t)\vec{x}_{eq}$. According to this overdamped relaxation equation, if only the solvent viscosity changes, i.e., $\boldsymbol{\gamma} = \eta\boldsymbol{\gamma}_0$, where $\boldsymbol{\gamma}_0$ is independent of η, then the eigenvectors of the $D \times D$ matrix $\boldsymbol{\gamma}^{-1}\boldsymbol{K}$ (here $D = 3N$) are *in*dependent of the viscosity coefficient η, while all D eigenvalues scale directly proportional to η^{-1}. This implies that for a given initial displacement $\vec{x}_{eq}$, the scaled time evolution $< \vec{x}(\tau = t/\eta) >_I$ is a universal curve, i.e., obtained for all values of η. It also allows us to easily predict how the relaxation time scales with η in the problem under study here (all other parameters being fixed), or, alternatively, how the relaxation time scales with the curvature of the PES, i.e., $\boldsymbol{K}$ (all other parameters being fixed). Figure 21 shows that the overdamped regime

has clearly set in when the solvent viscosity parameter reaches the value 100 cP (the time evolution calculated via overdamped relaxation equation is denoted as "overdamped limit" in the figure legend), and thus the simple overdamped limit expression for $< \vec{x}(t) >_I$ above can be employed to deduce the expected scaling of relaxation times with increasing effective viscosity (see Discussion Section below).

To the best of our knowledge real-time relaxation of either FBP upon dissociation of its Fe3+ ligand from its binding pocket or T4L upon dissociation of its polysaccharide ligand from its cleft has not been experimentally monitored at the single molecule level. Some evidence that the time scales deduced from the present analysis are relevant to ligand unbinding processes in cleft opening motions of proteins of the rough size of FBP and T4L comes from measurements on 3 periplasmic binding proteins, namely, the L-arabinose binding protein, the L-histidine binding protein and the D-maltose binding protein. For these proteins, cleft closure time scales on the order of nanoseconds have been experimentally inferred. The MD simulations of de Groot et al. mentioned above also suggest that significant cleft motion can occur on a time scale of several hundred picoseconds. Of course, we cannot claim that all ligand-binding and unbinding processes are well-described by the simple relaxational mechanisms implied by the multi-dimensional harmonic oscillator invoked herein. In particular, experiments on some cleft opening/closing proteins suggest much longer time scales for the analogous processes. For example, for the S1S2 domain of GluR2 [130, 131], the best experimental estimates of the time scale of the induced transformation from liganded to un-linganded conformation are on the order of milliseconds. Presumably, these long times reflect an underlying thermally activated barrier crossing process with significant (multiple $k_B T$) barrier(s) separating ligand bound and ligand un-bound conformations. Such a mechanism is probably not well described by either linear

response theory (since large-scale conformational distortions are entailed) or by harmonic oscillator models for the underlying PESs (due the essential role played by the intrinsically anharmonic barriers that separate initial and final states).

Nevertheless, it is interesting to note that the basic model studied here, i.e., diffusional relaxation on effective harmonic oscillator PESs, can in fact accommodate multi-nanosecond (and longer) time scales given appropriate values of the PES parameters and friction constants. Furthermore, more sophisticated modeling of the PES characteristics and the effective friction constants governing the essential conformational motion involved in the relaxation process may generate numerical values for these parameters that would imply multi-nanosecond time scales for the underlying protein relaxation process according to our basic model. These issues are considered in detail in the Discussion section which follows.

4.8 DISCUSSION AND CONCLUSIONS

4.8.1 Discussion

There are two readily apparent mechanisms by which the basic Langevin harmonic oscillator model utilized in our computations can be retained and the time scale of the relaxation extended by several orders of magnitude. First, there is the protein PES itself. We employed a local 2nd order Taylor Series about the minimum of an appropriate reference protein configuration. Such an expansion is designed to account for small amplitude vibrations about equilibrium, but is not guaranteed to be accurate for large scale amplitudes of atomic motion. Although anharmonic terms are surely non-negligible over the course of the relatively large amplitude motions under

investigation here, it may be possible to subsume these into a roughly quadratic *free energy* surface. Full computation of this surface is beyond the scope of the present work, but it is plausible that the resultant effective mean quadratic restoring forces would be less stiff (corresponding to lower effective vibrational frequencies), because steric clashes encountered in the course of a typical normal mode vibration would be relaxed via Boltzmann sampling of the other protein coordinates. A recent analysis along these lines by Bertaccini et al. [28] on the 1655 residue pentameric Glycine Receptor protein showed that when the atoms were moved slowly along the directions implied by the lowest vibrational normal mode (which was calculated via a local 2^{nd} order Taylor Series as we have done in this work for FBP and T4L), and energetically relaxed using the underlying anharmonic protein PES to relieve the development of high energy contacts, the curvature of the effective potential energy vs. coordinate displacement curve decreased by a factor of ca. 3600, corresponding to a reduction in vibrational frequency from the prediction of a the local 2^{nd} order Taylor series by a factor of ca. 60. Thus, this analysis hints that the relevant PESs for FBP and T4L conformational motion may have a considerably smaller curvature along critical coordinates than the values adopted in the calculations presented here. If we imagine scaling the 2^{nd} derivative matrix $\boldsymbol{K}$ by an overall factor of $10^3 - 10^4$ to account for these effects in a crude fashion, this would increase the time scale of the 100-300 ps relaxation processes observed in our computations (cf. Figure 19 and Figure 20, for FBP and T4L, respectively) to ca. $10^{-7} - 10^{-6}$ seconds, i.e., approaching the microsecond time scale. And indeed, there is another effect which could also increase the relaxation time scale further, namely, intrinsic roughness of the energy surface.

The notion that free energy surfaces relevant to large scale conformation changes (encountered, for example, in protein folding [132] or ligand binding/unbinding[133]) may

include a "rough" corrugation pattern added to a smooth functional dependence on dynamically important protein coordinates or other order parameters has been supported by experimental evidence for over 20 years[113]. Analysis of models based on this premise has shown that roughness of the energy landscape can lead to a significant increase in the effective friction constant governing Brownian motion on a smooth underlying energy surface [134]. For Gaussian distributed random roughness with a characteristic energy scale ε in the energy landscape , the scaling of friction constant with the characteristic energy scale of the surface roughness has been postulated to go as $\exp([\varepsilon / k_B T]^2)$ [134]: in this case the modest value of $\varepsilon \approx 2.6 k_B T$ could generate a factor of 1000 increase in the effective friction constant relevant to our FBP and T4L computations, and hence, assuming the corresponding dilatation of time implied by the overdamped Langevin Equation analysis sketched at the end of Section 4.7, the same scaling factor in the protein conformational relaxation time scale.

The salient point of the above discussion is that two separate factors, namely (i) a reduction in the curvature of the effective harmonic oscillator PES due to relaxation of energetically unfavorable contacts in the course of the conformational dynamics and (ii) increase in the effective system friction (viscosity) due to intrinsic roughness of the effective free energy surface, could each easily lead to a factor of, say, 1000 slowdown in the relaxation dynamics.

4.8.2 Conclusions

In this paper, we have exploited the dynamical version of linear response theory (LRT) to assist in the prediction of the response of a macromolecular system (e.g., a protein) to a change in the system potential. Such a change will be induced, for example, when a ligand molecule bound to

a well-defined binding pocket within a protein molecule dissociates from the binding pocket. Due to the resultant removal of the energetic interactions between ligand molecule and protein, a conformational change in the latter occurs[23]. Dynamic LRT enables one to calculate the relaxation of the protein from ligand-bound to ligand-free ("apo") states of the protein in terms of equilibrium time correlation functions associated with the apo-state Hamiltonian. These equilibrium time correlation functions can be evaluated by standard MD techniques. In particular, the effects of water solvent on the process can be included in the analysis in a manner which is straightforward in principle (by including explicit solvent water molecules in the MD simulation).

Given the time-consuming nature of all-atom MD simulations in explicit solvent, as well as the restriction of LRT (implied by its name) to small perturbations in the system force field and hence small distortions of the protein molecule, we developed in this work the details of dynamic LRT for a multidimensional harmonic oscillator Hamiltonian, of the type extensively used to describe protein fluctuations about an equilibrium conformation [89, 90, 135]. The analytical advantages of invoking a harmonic oscillator Hamiltonian as compared to a more realistic anharmonic Hamiltonian are considerable. First, all desired quantities can be calculated from manipulations of finite dimensional matrices (linear dimension at most 2D, where D is the number of degrees of Cartesian coordinates in the protein; typically, D=3N, where N is the number of protein atoms), in tandem with judicious evaluation of Gaussian integrals. Second, LRT is exact for any perturbation potential (hence protein conformational change) that is linear in the system coordinates, large or small. An obvious *dis*advantage of a harmonic oscillator model is its degree of idealization (i.e., neglect of anharmonic term in the PES), which may be considerable. One immediate concern is the treatment of solvent, which is responsible for

critical dissipative effects upon the protein motion. Fortunately, it is possible to include solvent effects in the harmonic oscillator model in a phenomenological but nevertheless physically appealing manner by introducing friction and a random force term, i.e., by considering a Langevin model of the harmonic oscillator dynamics of the system (protein molecule). We developed the details of this model, too, and showed that dynamic LRT is also exact for any size perturbation force which is coordinate independent when frictional (solvent) damping is included in this fashion.

We then applied the Langevin harmonic oscillator LRT formulation to numerically study relaxation of two protein systems, FBP and T4L, from ligand-bound to apo structures. For both systems, we utilized the AMBER force field to calculate the Hessian matrix, i.e., matrix of 2^{nd} derivatives of the full protein potential at the equilibrium configuration of one of these structures. In this way we constructed the required harmonic oscillator model of the protein system. Friction coefficients were estimated using the venerable Stokes model, which inputs the solvent viscosity. A main goal of the study was to ascertain the *time scales* of the relaxation process, focusing on the time-dependence of the distance between selected residue pairs. For a reasonable choice of solvent viscosity (1 cP), we found relaxation times in 100-300 ps range for these two proteins. Experimental determination these time scales is difficult, and not presently available for these proteins. However, there is some evidence from both experimental measurements on proteins of similar size and from all-atom MD simulations that this time scale is not completely unrealistic. Of course, there surely exist processes characterized by free energy surfaces which, even after averaging over local roughness (as discussed above), are quintessentially anharmonic: for example, systems where large-amplitude conformational change requires an activated process-type barrier crossing [136, 137] For such systems, both Linear Response Theory and an

effective harmonic oscillator free energy surface model are suspect. Clearly, a fully satisfactory explanation for the relaxation times of ligand dissociation induced conformational changes of large proteins will require careful further study, including both accurate free energy computations and careful extraction from the resultant energy landscapes of the dynamical response to ligand dissociation.

4.9 ACKNOWLEDGMENTS

This work was supported in part by NSF grant CHE-0750332. We are grateful to Dr. R. Christie and Prof. M.G. Kurnikova for helpful discussions and to Prof. Kurnikova for a critical reading of the manuscript. Numerical computations were carried out at the Univ. of Pittsburgh's Center for Molecular and Materials Modeling.

APPENDIX A

THE UNDERLYING CONSTRAINT FORCE IN RTB-DYNAMICS AND ITS IMPLICATIONS

At *t=0*, we compute the components of the initial (mass-weighted) velocity vector $\vec{v}_0$ along each of the n_r relevant displacement coordinate vectors. Let us also compute the remainder velocity $\vec{v}_\perp^{(0)}$ using Equation 2.6 from the text. This describes initial motion in the subspace of irrelevant displacements. So, we introduce a constraint force along the direction of $\vec{v}_\perp^{(0)}$. Note that in the absence of any forces acting in the space of motions mutually perpendicular to all the relevant coordinates, after time δt these displacement coordinates would take on the value $\vec{\eta}_\perp(\delta t) = \vec{v}_\perp^{(0)}\delta t$. What we want, instead, is to find $\vec{\eta}_\perp(\delta t) = \vec{0}$. To achieve this we need to add a (constant, time-independent) constraint force $\vec{F}_C$ of the form,

$$\vec{F}_C = -\vec{F}_\perp + \vec{g}\,.$$

Here $\vec{F}_\perp$ is the component of the systematic force (derived from the overall potential function) on the non-RTB internal coordinates. (N.B.: this does not have to explicitly calculated, as is made clear below.) Since the net force on the non-RTB coordinates is $\vec{F}_\perp + \vec{F}_C$, the constraint force cancels out the systematic force due to the physical molecular force field, and adds an extra

"piece", namely $\vec{g}$, whose purpose is simply to counter-balance the inertial motion imparted by $\vec{v}_{\perp}^{\,(0)}$. That is,

$$\vec{\eta}_{\perp}(\delta t) = \vec{v}_{\perp}^{(0)}\delta t + \tfrac{1}{2}\vec{g}\delta t^2 .$$

By choosing $\vec{g} = -2\vec{v}_{\perp}^{(0)}/\delta t$, we achieve $\vec{\eta}_{\perp}(\delta t) = \vec{0}$, as desired. The consequence of this constraint force for the output velocity of the irrelevant internal coordinates is

$$\vec{v}_{\perp}(\delta t) = \vec{v}_{\perp}^{(0)} + [\tfrac{-2\vec{v}_{\perp}^{(0)}}{\delta t}]\delta t = -\vec{v}_{\perp}^{(0)} .$$

The final velocity of each irrelevant z-coordinate is seen to be simply the negative of its value at *t=0*.

APPENDIX B

PROOF OF ENERGY CONSERVATION IN RTB-DYNAMICS

RTB rigid-body dynamics conserves the total system energy in the limit of small time step, with the accuracy of the velocity-Verlet algorithm. For each propagation time-step we can construct a modified time-independent potential energy function $\tilde{V}$ which corresponds to the desired force on the system, including the constraint force needed to freeze all irrelevant coordinates (cf. App. A). Thus, if $K(t)$ is the kinetic energy of the system then an exact solution of Newton's Equations will have the property that

$$K(0)+\tilde{V}(0)=K(\delta t)+\tilde{V}(\delta t)\,. \tag{B1}$$

The velocity Verlet algorithm yields an approximate solution of Newton's Equations, but one which is very accurate for small δt; hence for small δt, Equation (B1) holds to a very good approximation. Finally, the appropriate $\tilde{V}$ turns out to have the property that $\tilde{V}(0)=V(0)$ and $\tilde{V}(\delta t)=V(\delta t)$. Thus, it follows from Equation (B1) that

$$K(0)+V(0)=K(\delta t)+V(\delta t)\,, \tag{B2}$$

which is the desired conservation of energy condition.

We want next to exhibit the pertinent potential energy function $\tilde{V} = V + V_C$, where V_C is the additional term that must be added to the system potential to generate the appropriate constraint force. To do this, recall a few details from the discussion above. There are $3N - n_r$ irrelevant coordinates; let us label these here via $\kappa = n_r + 1, ..., 3N$. The initial mass-weighted "perpendicular velocity" $\vec{v}_\perp^{(0)}$ can be written as a linear combination of the irrelevant basis vectors, namely,

$$\vec{v}_\perp^{(0)} = \sum_{\kappa=n_r+1}^{3N} v_0^{(\kappa)} \hat{u}^{(\kappa)} \ ,$$

with $v_0^{(\kappa)} \equiv \vec{v}_0 \cdot \hat{u}^{(\kappa)}$. Similarly, the $\vec{g}$ component of the constraint force has a similar decomposition,

$$\vec{g} = \sum_{\kappa=n_r+1}^{3N} g^{(\kappa)} \hat{u}^{(\kappa)} \ .$$

We showed above that $\vec{g} = -2\vec{v}_\perp^{(0)} / \delta t$, and thus we can identify $g^{(\kappa)} = -2 v_0^{(\kappa)} / \delta t$. The complete constraint force is then given as

$$\vec{F}_C = \sum_{\kappa=n_r+1}^{3N} [-F_S^{(\kappa)} - 2v_0^{(\kappa)} / \delta t] \hat{u}^{(\kappa)} \ ,$$

where $F_S^{(\kappa)}$ is the component of systematic force along $\hat{u}^{(\kappa)}$ at $t = 0$. It is worth emphasizing that $[-F_S^{(\kappa)} - 2v_0^{(\kappa)} / \delta t]$ is a constant on the interval $0 < t < \delta t$. Thus, $\vec{F}_C$ is the negative of the gradient of the following scalar function:

$$V_C = \sum_{\kappa=1}^{3N-n_r} [F_S^{(\kappa)} + 2v_0^{(\kappa)} / \delta t] \eta^{(\kappa)} \ .$$

Note that this time-independent function of coordinates has the critical properties claimed above. First, it generates the desired constraint force, which is designed to produce unchanged positions

of the irrelevant coordinates at the end of the time interval compared to their initial positions; specifically, $\eta^{(\kappa)}(0) = 0 = \eta^{(\kappa)}(\delta t)$, $\kappa = n_r + 1, \ldots, 3N$. Consequently, $V_C(0) = 0 = V_C(\delta t)$. This establishes conservation of energy over the interval (cf. Equation B2).

APPENDIX C

COUPLING TO MULTI-DIMENSIONAL BATH

We wish to show how the dynamic LRT formulae involving a harmonic oscillator model of system (protein) coordinates coupled to a dissipative environment via the Langevin Equation which were developed in Section 4. 5 can be obtained by appropriately coupling the system Hamiltonian to a large set of auxiliary harmonic oscillator coordinates (a "bath") and subsequently considering the purely Hamiltonian motion (and statistical mechanics) of the augmented system-bath super-system. The basic idea can be conveyed with minimal effort and notational clutter for the case of a one-dimensional system described by coordinate X and momentum P coupled to a multi-dimensional bath (which mimics frictional effects on the system). Thus, consider a final state Hamiltonian of the form:

$$H_F = H_X + H_{bath}, \tag{C1}$$

where

$$H_X(X,P) = P^2/2 + V(X),$$

with a simple harmonic oscillator system potential $V(X) = \frac{1}{2}\Omega^2 X^2$, and:

$$H_{bath}(\vec{x},\vec{p};X)=\sum_j\left\{p_j^2/2+\omega_j^2\left[x_j-\lambda_j X/\omega_j^2\right]^2/2\right\},$$

where the sum is over N_b bath modes, the frequency of mode j (described by coordinate x_j and momentum p_j) being ω_j. Note that we have set all masses to unity, with no loss of generality. Of greater significance, note that H_{bath} depends parametrically on the system coordinate X with coefficient λ_j determining the coupling strength between the system coordinate and bath coordinate (mode) x_j.

Now add a linear perturbation which couples directly only to the system coordinate, namely $\delta V=-fX$, and define, as prescribed in the main text, $H_I=H_F-fX$. Thus, explicitly:

$$H_I=H_X^{'}+H_{bath}\ , \tag{C2}$$

with

$$H_X^{'}(X,P)=P^2/2+\frac{1}{2}\Omega^2X^2-fX$$

$$=P^2/2+\frac{1}{2}\Omega^2(X-f/\Omega^2)^2-f^2/2\Omega^2.$$

H_F in Equation C1 is an N_b+1 -dim. harmonic oscillator Hamiltonian, and H_I in Equation C2 is the same as H_F except with an added perturbation that is linear in the coordinates. Thus H_I, H_F are of the form considered in Section 4.4, and therefore we know immediately that:

$$<X(t)>_I=\beta f<X(t)X>_F\ , \tag{C3}$$

where, as usual, the dynamics on both sides of Equation C3 is generated according to Hamiltonian H_F, while the initial phase space density relevant to the left hand side of this

Equation is (properly normalized) $\exp(-\beta H_I)$ and the corresponding initial phase space density relevant to the right hand side is $\exp(-\beta H_F)$.

Defining Z_X as the thermal partition function associated with the isolated system Hamiltonian, and Z_{bath} as the partition function for the bath in the absence coupling to the system coordinate, i.e.,

$$Z_X = \int dXdP \exp[-\beta H_X(X,P)] \quad ; \quad Z_{bath} = \int d\vec{x}d\vec{p} \exp[-\beta H_{bath}(\vec{x},\vec{p};0)],$$

and noting that the complete system-bath partition function is given by the product of these two factors, we can write the l.h.s. of Equation C3 as,

$$< X(t) >_I = \frac{1}{Z_X} \int dX_0 dP_0 \exp[-\beta H_X'(X_0,P_0)] \bar{X}(X_0,P_0,t), \tag{C4}$$

where

$$\bar{X}(X_0,P_0,t) = \frac{1}{Z_{bath}} \int d\vec{x}_0 d\vec{p}_0 \exp[-\beta H_{bath}(\vec{x}_0,\vec{p}_0;X_0)] X(X_0,P_0,\vec{x}_0,\vec{p}_0,t), \tag{C5}$$

with $X(X_0,P_0,\vec{x}_0,\vec{p}_0,t)$ being the time evolution on H_F of the system coordinate out of the initial phase point (X_0,P_0), which is then *averaged* over initial positions and momenta of the bath according to the thermal phase space distribution of the bath for the given initial position of the system X_0 to obtain $\bar{X}(X_0,P_0,t)$. The equilibrium correlation function $< X(t)X >_F$ can also be written in terms of $\bar{X}(X_0,P_0,t)$, namely,

$$< X(t)X >_F = \frac{1}{Z_X} \int dX_0 dP_0 \exp[-\beta H_X(X_0,P_0)] \bar{X}(X_0,P_0,t) X_0. \tag{C6}$$

Now, the temporal function $X(X_0,P_0,\vec{x}_0,\vec{p}_0,t)$ is obtained by solving Hamilton's Equations for the full N_b+1 -dim. system, given the initial values of all N_b+1 positions and N_b+1 momenta.

Because of the bilinear coupling between X and each of the x_j, each bath mode is linearly driven by the system coordinate trajectory. This allows the bath degrees of freedom to be "integrated out", leaving a perturbed equation of motion for the system coordinate [138]. Specifically:

$$d^2X/dt^2 = -\frac{\partial}{\partial X}V(X) - \int_0^t du\Gamma(t-u)\dot{X}(u) + R(t), \tag{C7}$$

where

$$\Gamma(t) = \sum_j \frac{\lambda_j^2}{\omega_j^2}\cos(\omega_j t), \tag{C8}$$

and

$$R(t) = \sum_j \lambda_j \left[(q_{0,j} - \lambda_j Q_0/\omega_j^2)\cos(\omega_j t) + \frac{p_{0j}}{\omega_j}\sin(\omega_j t) \right]. \tag{C9}$$

Note that $R(t)$ is a Fourier series whose coefficients are linearly dependent on the initial phase space variables of the bath modes. Thus when the bath average $\int d\vec{x}_0 d\vec{p}_0 \exp[-\beta H_{bath}(\vec{x}_0, \vec{p}_0; X_0)]$ is performed, $R(t)$ can be described as a Gaussian noise term. Its defining properties are that $< R(t) >= 0$ and that $< R(s)R(t) >= k_B T\Gamma(t-s)$. In other words, the Fluctuation-Dissipation Theorem [123, 124] is satisfied, and we can say that $\bar{X}(X_0, P_0, t)$ is the trajectory of a 1-dim. system evolved out of initial phase point (X_0, P_0) according to the Generalized Langevin Equation C7 corresponding to the memory friction kernel $\Gamma(t)$ prescribed in Equation C8. Since the derivation holds for arbitrary memory friction kernel, it holds for the special case of very rapidly decaying memory kernel (this is simply a matter of choosing the appropriate values of the λ_j), i.e., such that the GLE in Equation C7 reduces to the ordinary Langevin Equation featured in Section 4.5 of the text with a prescribed time-local friction

coefficient. The proof presented here for a system consisting of a single coordinate can easily be generalized to a multi-coordinate system characterized by any desired multi-dimensional harmonic oscillator potential and friction constant matrix.

BIBLIOGRAPHY

1. Durand, P., G. Trinquier, and Y.H. Sanejouand, *New Approach for Determining Low-Frequency Normal-Modes in Macromolecules.* Biopolymers, 1994. **34**(6): p. 759.
2. Tama, F., F.X. Gadea, O. Marques, and Y.H. Sanejouand, *Building-block approach for determining low-frequency normal modes of macromolecules.* Proteins-Structure Function and Genetics, 2000. **41**(1): p. 1.
3. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nature Structural Biology, 2002. **9**: p. 646.
4. Becker, O.M., A.D. MacKerell, B. Roux, and M. Watanabe, eds. *Computational Biochemistry and Biophysics*. 2001, Marcel Dekker: New York.
5. Hille, B., *Ion Channels of Excitable Membranes*. 3rd ed. 2001, Sunderland: Sinaur Associates,Inc.
6. Barchi, R.L., *Ion channel mutations affecting muscle and brain.* Current Opinion in Neurology, 1998. **11**: p. 461.
7. Li, M. and H.A. Lester, *Ion channel diseases of the central nervous system.* CNS Drug Reviews, 2001. **2001**(7(2)): p. 214.
8. Paterson, D. and A. Nordberg, *Neuronal nicotinic receptors in the human brain.* Progress in Neurobiology, 2000. **61**(1): p. 75.
9. Arseniev, A.S., I.L. Barsukov, V.F. Bystrov, A.L. Lomize, and A.O. Yu, *1HNMR study of gramicidin A transmembrane ion channel. Head-to-head right-handed, single-stranded helices.* FEBS Lett.. 1985. **186**: p. 168.
10. Doyle, D.A., C.J. Morais, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait, and R. MacKinnon, *The structure of the potassium channel: Molecular basis of K+ conduction and selectivity.* Science, 1998. **280**: p. 69.
11. Song, L., H. C., S. Shustak, H. Cheley, H. Bayley, and J.E. Gouaux, *Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore.* Science, 1996. **274**(5294): p. 1859.
12. Tang, P., P.K. Mandal, and Y. Xu, *NMR structures of the second transmembrane domain of the human glycine receptor alpha(1) subunit: Model of pore architecture and channel gating.* Biophysical Journal, 2002. **83**(1): p. 252.
13. Barry, P.H. and J.W. Lynch, *Ligand-gated channels.* IEEE Transactions on Nanobioscience, 2005. **4**(1): p. 70.
14. Cascio, M., *Structure and function of the glycine receptor and related nicotinicoid receptors.* Journal of Biological Chemistry, 2004. **279**(19): p. 19383.
15. Brejc, K., W.J. van Dijk, R.V. Klaassen, M. Schuurmans, J. van der Oost, A.B. Smit, and T.K. Sixma, *Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors.* Nature, 2001. **411**(6835): p. 269.

16. Miyazawa, A., Y. Fujiyoshi, and N. Unwin, *Structure and gating mechanism of the acetylcholine receptor pore.* Nature, 2003. **423**(6943): p. 949.
17. Cascio, M., *Structure and function of the glycine receptor and related nicotinicoid receptors.* Journal of Biological Chemistry, 2004. **279**(19): p. 19383.
18. Ma, D.J., Z.W. Liu, L. Li, P. Tang, and Y. Xu, *Structure and dynamics of the second and third transmembrane domains of human glycine receptor.* Biochemistry, 2005. **44**(24): p. 8790.
19. Unwin, N., *Refined structure of the nicotinic acetylcholine receptor at 4 angstrom resolution.* Journal of Molecular Biology, 2005. **346**(4): p. 967.
20. Bocquet, N., H. Nury, M. Baaden, C. Le Poupon, J.P. Changeux, M. Delarue, and P.J. Corringer, *X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation.* Nature, 2009. **457**(7225): p. 111.
21. Hilf, R.J.C. and R. Dutzler, *Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel.* Nature, 2009. **457**(7225): p. 115.
22. Hilf, R.J.C. and R. Dutzler, *X-ray structure of a prokaryotic pentameric ligand-gated ion channel.* Nature, 2008. **452**(7185): p. 375.
23. Auerbach, A., *Life at the top: the transition state of AChR gating.* Science STKE, 2003. **188**: p. re11.
24. Auerbach, A., *Gating of acetylcholine receptor channels: brownian motion across a broad transition state.* Proc. Natl. Acad. Sci. U S A, 2005. **102**(5): p. 1408.
25. Purohit, P., A. Mitra, and A. Auerbach, *A stepwise mechanism for acetylcholine receptor channel gating.* Nature, 2007. **446**(7138): p. 930.
26. Law, R.J., R.H. Henchman, and J.A. McCammon, *A gating mechanism proposed from a simulation of a human alpha 7 nicotinic acetylcholine receptor.* Proc. Natl. Acad. Sci. USA, 2005. **102**(19): p. 6813.
27. Cheng, X.L., I. Ivanov, H.L. Wang, S.M. Sine, and J.A. McCammon, *Nanosecond-timescale conformational dynamics of the human alpha 7 nicotinic acetylcholine receptor.* Biophysical Journal, 2007. **93**: p. 2622.
28. Bertaccini, E.J., J.R. Trudell, and E. Lindahl, *Normal-mode analysis of the glycine alpha1 receptor by three separate methods.* Journal of Chemical Information and Modeling, 2007. **47**(4): p. 1572.
29. Freddolino, P.L., F. Liu, M. Gruebele, and K. Schulten, *Ten-microsecond molecular dynamics simulation of a fast-folding WW domain.* Biophysical Journal, 2008. **94**(10): p. L75.
30. Cheng, X.L., B.Z. Lu, B. Grant, R.J. Law, and J.A. McCammon, *Channel opening motion of alpha 7 nicotinic acetylcholine receptor as suggested by normal mode analysis.* Journal of Molecular Biology, 2006. **355**(2): p. 310.
31. Taly, A., M. Delarue, T. Grutter, M. Nilges, N. Le Novere, P.J. Corringer, and J.P. Changeux, *Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism.* Biophysical Journal, 2005. **88**(6): p. 3954.
32. Wilson, G.G. and A. Karlin, *The location of the gate in the acetylcholine receptor channel.* Neuron, 1998. **20**: p. 1269.
33. Wilson, G.G. and A. Karlin, *Acetylcholine receptor channel structure in the resting, open, and desensitized states probed with the substituted-cysteine accessibility method.* Proc. Natl. Acad. Sci. USA, 2001. **98**: p. 1241.

34. Delarue, M. and Y.H. Sanejouand, *Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the Elastic Network Model.* Journal of Molecular Biology, 2002. **320**(5): p. 1011.
35. Chun, H.M., C.E. Padilla, D.N. Chin, M. Watanabe, V.I. Karlov, H.E. Alper, K. Soosaar, K.B. Blair, O.M. Becker, L.S.D. Caves, R. Nagle, D.N. Haney, and B.L. Farmer, *MBO(N)D: A multibody method for long-time molecular dynamics simulations.* Journal of Computational Chemistry, 2000. **21**(3): p. 159.
36. Andersen, H.C., *RATTLE: A Velocity Version of the SHAKE Algorithm for Molecular Dynamics.* Journal of Computational Physcis, 1983. **52**: p. 24.
37. van Gunsteren, W.F. and M. Karplus, *Effect of constraints on the dynamics of macromolecules.* . Macromolecules, 1982. **15**: p. 1528.
38. Schlitter, J., M. Engels, P. Kruger, E. Jacoby, and A. Wollmer, *Targeted Molecular Dynamics Simulation of Conformational Change - Application to the T-R Transition in Insulin.* Molecular Simulation, 1993. **10**(2-6): p. 291.
39. Isralewitz, B., M. Gao, and K. Schulten, *Steered molecular dynamics and mechanical functions of proteins.* Current Opinion in Structural Biology, 2001. **11**: p. 224.
40. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: Survey of methods for simulating the activity of proteins.* Chemical Reviews, 2006. **106**: p. 1589.
41. Chandrasekhar, S., *Stochastic Problems in Physics and Astronomy.* Reviews of Modern Physics, 1943. **15**: p. 1
42. Leach, A., *Molecular Modeling, Principles and Applications*. 2nd ed. 2001: Pearson Education Limited.
43. Chandler, D., *An introduction to Modern Statistical Mechnaics*. 1987, New York: Oxford Univ. Press.
44. Li, G. and Q. Cui, *A coarse-grained normal mode approach for macromolecules: An efficient implementation and application to Ca2+-ATPase.* Biophysical Journal, 2002. **83**(5): p. 2457.
45. Cullum, J. and R.A. Willioughby, *Lanczos Algorithms for Large Symetric Eigenvalue Computations*. 1985, Boston: Birkhauser.
46. Hayward, S., A. Kitao, and N. Go, *Harmonic and Anharmonic Aspects in the Dynamics of Bpti - a Normal-Mode Analysis and Principal Component Analysis.* Protein Science, 1994. **3**(6): p. 936.
47. Amadei, A., A.B.M. Linssen, and H.J.C. Berendsen, *Essential Dynamics of Proteins.* Proteins-Structure Function and Genetics, 1993. **17**(4): p. 412.
48. Vanaalten, D.M.F., A. Amadei, A.B.M. Linssen, V.G.H. Eijsink, G. Vriend, and H.J.C. Berendsen, *The Essential Dynamics of Thermolysin - Confirmation of the Hinge-Bending Motion and Comparison of Simulations in Vacuum and Water.* Proteins-Structure Function and Genetics, 1995. **22**(1): p. 45.
49. Balsera, M.A., W. Wriggers, Y. Oono, and K. Schulten, *Principal component analysis and long time protein dynamics.* Journal of Physical Chemistry, 1996. **100**(7): p. 2567.
50. Li, G. and Q. Cui, *Analysis of functional motions in Brownian molecular machines with an efficient block normal mode approach: Myosin-II and Ca2+-ATPase.* Biophysical Journal, 2004. **86**(2): p. 743.
51. Tirion, M.M., *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis.* Physical Review Letters, 1996. **77**(9): p. 1905.

52. Atilgan, A.R., S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, and I. Bahar, *Anisotropy of fluctuation dynamics of proteins with an elastic network model.* Biophysical Journal, 2001. **80**(1): p. 505.
53. Bahar, I., A.R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential.* Folding & Design, 1997. **2**(3): p. 173.
54. M. Delarue and S. Y.-H, *Simplified Normal Mode Analysis of Conformations Transitions in DNA-dependent Polymerases: the Elastic Network Model.* Journal of Molecular Biology, 2002. **320**: p. 1011.
55. Tama, F., O. Miyashita, and C.L. Brooks, *Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM.* Journal of Structural Biology, 2004. **147**(3): p. 315.
56. Tama, F. and Y.H. Sanejouand, *Conformational change of proteins arising from normal mode calculations.* Protein Engineering, 2001. **14**(1): p. 1.
57. Space, B., H. Rabitz, and A. Askar, *Long-Time Scale Molecular-Dynamics Subspace Integration Method Applied to Anharmonic Crystals and Glasses.* Journal of Chemical Physics, 1993. **99**(11): p. 9070.
58. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis.* Proc. Natl. Acad. Sci. U S A, 1958. **44**: p. 98.
59. Ikeguchi, M., J. Ueno, M. Sato, and A. Kidera, *Protein structural change upon ligand binding: Linear response theory.* Physical review letters, 2005. **94**(7).
60. Hansen, J.P. and I.R. McDonald, *Theory of Simple Liquids*. 1986: Academic Press.
61. Weiss, S., *Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy.* Nature Structural Biology, 2000. **7**(9): p. 724.
62. Jun, S., J.S. Becker, M. Yonkunas, R. Coalson, and S. Saxena, *Unfolding of alanine-based peptides using electron spin resonance distance measurements.* Biochemistry, 2006. **45**(38): p. 11666.
63. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins.* Nature, 2007. **450**: p. 964.
64. Gerstein, M. and W. Krebs, *A database of macromolecular motions.* Nucleic Acids Research, 1998. **26**(18): p. 4280.
65. Bruns, C.M., D.S. Anderson, K.G. Vaughan, P.A. Williams, A.J. Nowalk, D.E. McRee, and T.A. Mietzner, *Crystallographic and biochemical analyses of the metal-free Haemophilus influenzae Fe3+-binding protein.* Biochemistry, 2001. **40**(51): p. 15631.
66. Bruns, C.M., A.J. Nowalk, A.S. Arvai, M.A. McTigue, K.G. Vaughan, T.A. Mietzner, and D.E. McRee, *Structure of Haemophilus influenzae Fe+3-binding protein reveals convergent evolution within a superfamily.* Nature Structural Biology, 1997. **4**(11): p. 919.
67. Faber, H.R. and B.W. Matthews, *A Mutant T4 Lysozyme Displays 5 Different Crystal Conformations.* Nature, 1990. **348**(6298): p. 263.
68. Weaver, L.H. and B.W. Matthews, *Structure of Bacteriophage-T4 Lysozyme Refined at 1.7 A Resolution.* Journal of Molecular Biology, 1987. **193**(1): p. 189.
69. Kuroki, R., L.H. Weaver, and B.W. Matthews, *A Covalent Enzyme-Substrate Intermediate with Saccharide Distortion in a Mutant T4 Lysozyme.* Science, 1993. **262**(5142): p. 2030.

70. Sansom, M.S.P., I.H. Shrivastava, K.M. Ranatunga, and G.R. Smith, *Simulations of ion channels - watching ions and water move.* Trends in Biochemical Sciences, 2000. **25**(8): p. 368.
71. Shrivastava, I.H. and M.S.P. Sansom, *Simulations of ion permeation through a potassium channel: Molecular dynamics of KcsA in a phospholipid bilayer.* Biophysical Journal, 2000. **78**(2): p. 557.
72. Crozier, P.S., R.L. Rowley, N.B. Holladay, D. Henderson, and D.D. Busath, *Molecular dynamics simulation of continuous current flow through a model biological membrane channel.* Physical Review Letters, 2001. **86**(11): p. 2467.
73. Aksimentiev, A. and K. Schulten, *Imaging alpha-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map.* Biophysical Journal, 2005. **88**(6): p. 3745.
74. Cheng, M. H. and R.D. Coalson, *An accurate and efficient empirical approach for calculating the dielectric self-energy and ion-ion pair potential in continuum models of biological ion channels.* Journal of Physical Chemistry B, 2005. **109**(1): p. 488.
75. Graf, P., M.G. Kurnikova, R.D. Coalson, and A. Nitzan, *Comparison of dynamic lattice Monte Carlo simulations and the dielectric self-energy Poisson-Nernst-Planck continuum theory for model ion channels.* Journal of Physical Chemistry B, 2004. **108**(6): p. 2006.
76. Chung, S.H., T.W. Allen, and S. Kuyucak, *Conducting-state properties of the KcsA potassium channel from molecular and Brownian dynamics simulations.* Biophysical Journal, 2002. **82**(2): p. 628.
77. Graf, P., A. Nitzan, M.G. Kurnikova, and R.D. Coalson, *A dynamic lattice Monte Carlo model of ion transport in inhomogeneous dielectric environments: Method and implementation.* Journal of Physical Chemistry B, 2000. **104**(51): p. 12324.
78. Im, W., S. Seefeld, and B. Roux, *A Grand Canonical Monte Carlo-Brownian dynamics algorithm for simulating ion channels.* Biophysical Journal, 2000. **79**(2): p. 788.
79. Noskov, S.Y., W. Im, and B. Roux, *Ion permeation through the alpha-hemolysin channel: Theoretical studies based on Brownian dynamics and Poisson-Nernst-Plank electrodiffusion theory.* Biophysical Journal, 2004. **87**(4): p. 2299.
80. Chung, S.H., T.W. Allen, and S. Kuyucak, *Modeling diverse range of potassium channels with brownian dynamics.* Biophysical Journal, 2002. **83**(1): p. 263.
81. Burykin, A., C.N. Schutz, J. Villa, and A. Warshel, *Simulations of ion current in realistic models of ion channels: The KcsA potassium channel.* Proteins-Structure Function and Genetics, 2002. **47**(3): p. 265.
82. Grewer, C., *Investigation of the alpha(1)-glycine receptor channel-opening kinetics in the submillisecond time domain.* Biophysical Journal, 1999. **77**(2): p. 727.
83. Lewis, T.M., P.R. Schofield, and A.M.L. McClellan, *Kinetic determinants of agonist action at the recombinant human glycine receptor.* Journal of Physiology-London, 2003. **549**(2): p. 361.
84. Matsubara, N., A.P. Billington, and G.P. Hess, *How Fast Does an Acetylcholine-Receptor Channel Open - Laser-Pulse Photolysis of an Inactive Precursor of Carbamoylcholine in the Microsecond Time Region with Bc3h1 Cells.* Biochemistry, 1992. **31**(24): p. 5507.
85. Unwin, N., *Acetylcholine-Receptor Channel Imaged in the Open State.* Nature, 1995. **373**(6509): p. 37.

86. Ikeguchi, M., *Partial rigid-body dynamics in NPT, NPAT and NP gamma T ensembles for proteins and membranes.* Journal of Computational Chemistry, 2004. **25**(4): p. 529.
87. Cardenas, A.E. and R. Elber, *Atomically detailed Simulations of helix formation with the stochastic difference equation.* Biophysical Journal, 2003. **85**(5): p. 2919.
88. Wilson, E.B., D. J.C, and C. P.C., *Molecular Vibrations*. 1955, New York: McGraw-Hill.
89. Brooks, B.R. and M. Karplus, *Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor.* Proc. Natl. Acad. Sci. U S A, 1983. **80**: p. 6571.
90. Brooks, B.R., D. Janezic, and M. Karplus, *Harmonic-Analysis of Large Systems.* Journal of Computational Chemistry, 1995. **16**(12): p. 1522.
91. Cui, Q., G.H. Li, J.P. Ma, and M. Karplus, *A normal mode analysis of structural plasticity in the biomolecular motor F-1-ATPase.* Journal of Molecular Biology, 2004. **340**(2): p. 345.
92. Lay, D.C., *Linear Algebra and Its Applications*. 3rd ed. 2003: Addison-Wesley.
93. Swope, W.C., H.C. Andersen, P.H. Berens, and K.R. Wilson, *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules:Application to small water clusters.* Journal of Chemical Physics., 1982. **76**: p. 637.
94. Landau, L.D., *Mekhanika. English Mechanics / by L.D. Landau and E.M. Lifshitz ; translated from the Russian by J. B. Sykes and J. S. Bell*. 3rd ed. 1976, Oxford , New York: Pergamon Press.
95. Mamonov, A., R.D. Coalson, A. Nitzan, and M. Kurnikova, *The role of the dielectric barrier in narrow biological channels: A novel composite approach to modeling single-channel currents.* Biophysical Journal, 2003. **84**: p. 3646.
96. Arseniev, A.S., A.L. Lomize, and V.Y. Orekhov, *Refinement of the spatial structure of the gramicidin A ion channel.* Bioorg Khim, 1992. **18**: p. 182.
97. Case, D.A., D.A. Pearlman, J.W. Caldwell, T.E.C. III, J. Wang, W.S. Ross, C.L.Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R.J. Radmer, Y. Duan, J. Pitera, I. Massova, G.L. Seibel, U.C. Singh, P.K. Weiner, and P.A. Kollman, *AMBER 7*. 2002: University of California, San Francisco.
98. Humphrey, W., A. Dalke, and K. Schulten, *VMD - Visual Molecular Dynamics.* Journal of Molecular Graphics & Modelling, 1996. **14**: p. 33.
99. Ermak, D.L. and J.A. McCammon, *Brownian dynamics with hydrodynamic interactions.* Journal of Chemical Physics, 1978. **69**: p. 1352.
100. Thorpe, M.F., M. Lei, A.J. Rader, D.J. Jacobs, and L.A. Kuhn, *Protein flexibility and dynamics using constraint theory.* Journal of Molecular Graphics & Modelling, 2001. **19**(1): p. 60.
101. Keskin, O., S.R. Durell, I. Bahar, R.L. Jernigan, and D.G. Covell, *Relating molecular flexibility to function: A case study of tubulin.* Biophysical Journal, 2002. **83**(2): p. 663.
102. Taly, A., M. Delarue, T. Grutter, M. Nilges, N. Le Novere, P.J. Corringer, and J.P. Changeux, *Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism.* Biophysical Journal, 2005. **88**(6): p. 3954.
103. Valadie, H., J.J. Lacapcre, Y.H. Sanejouand, and C. Etchebest, *Dynamical properties of the MscL of Escherichia coli: A normal mode analysis.* Journal of Molecular Biology, 2003. **332**(3): p. 657.

104. Essiz, S. and R.D. Coalson, *A rigid-body Newtonian propagation scheme based on instantaneous decomposition into rotation and translation blocks.* Journal of Chemical Physics, 2006. **124**: p. 144116.
105. Lamm, G. and A. Szabo, *Langevin modes of macromolecules.* Journal of Chemical Physics, 1986. **12**: p. 7344.
106. Kottalam, J. and D.A. Case, *Langevin Modes of Macromolecules: Applications to Crambin and DNA Hexamers.* Biopolymers, 1990. **29** p. 1409.
107. Pastor, R.W. and M. Karplus, *Parametrization of the friction constant for stochastic simulations of polymers.* Journal of Physical Chemistry, 1988. **92**: p. 2636
108. Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. 1992, Cambridge: Cambridge University Press.
109. Smith, B.T., J.M. Boyle, J.J. Dongarra, B.S. Garbow, Y. Ikebe, V.C. Klema, and C.B. Moler, *Matrix Eigensystem Routines*. 1976, New York: Springer-Verlag.
110. Wilson, E.B., J.C. Decius, and P.C. Cross, *Molecular Vibrations*. 1955, New York: McGraw-Hill.
111. Fraczkiewicz, R. and W. Braun, *Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules.* Journal of Computational Chemistry, 1998. **19**(3): p. 319.
112. Kitao, A., R. Hirata, and N. Go., *The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum.* Chemical Physics, 1991. **158**: p. 447.
113. Ansari, A., J. Berendzen, S.F. Bowne, H. Frauenfelder, I.E. Iben, T.B. Sauke, E. Shyamsunder, and R.D. Young, *Protein states and proteinquakes.* Proc. Natl. Acad. Sci. USA, 1985. **82**(15): p. 5000.
114. Hardcastle, I.R., *Inhibitors of the MDM2-p53 interaction as anticancer drugs.* Drugs of the Future, 2007. **32**(10): p. 883.
115. Noble, M.E.M., J.A. Endicott, and L.N. Johnson, *Protein kinase inhibitors: Insights into drug design from structure.* Science, 2004. **303**(5665): p. 1800.
116. Wlodawer, A. and J. Vondrasek, *Inhibitors of HIV-1 protease: A major success of structure-assisted drug design.* Annual Review of Biophysics and Biomolecular Structure, 1998. **27**: p. 249.
117. Wolf, F.A.D. and G.M. Brett, *Ligand-Binding Proteins: Their Potential for Application in Systems for Controlled Delivery and Uptake of Ligands.* Pharmacological Reviews , 2000. **52** p. 207.
118. Amiri, S., M.S.P. Sansom, and P.C. Biggin, *Molecular dynamics studies of AChBP with nicotine and carbamylcholine: the role of water in the binding pocket.* Protein Engineering Design & Selection, 2007. **20**(7): p. 353.
119. Gao, F., N. Bren, T.P. Burghardt, S. Hansen, R.H. Henchman, P. Taylor, J.A. McCammon, and S.M. Sine, *Agonist-mediated conformational changes in acetylcholine-binding protein revealed by simulation and intrinsic tryptophan fluorescence.* Journal of Biological Chemistry, 2005. **280**(9): p. 8443.
120. Henchman, R.H., H.L. Wang, S.M. Sine, P. Taylor, and J.A. McCammon, *Ligand-induced conformational change in the alpha 7 nicotinic receptor ligand binding domain.* Biophysical Journal, 2005. **88**(4): p. 2564.

121. Speranskiy, K., M. Cascio, and M. Kurnikova, *Homology Modeling and molecular dynamics simulations of the glycine receptor ligand binding domain.* Proteins-Structure Function and Bioinformatics, 2007. **67**(4): p. 950.
122. Speranskiy, K. and M. Kurnikova, *On the binding determinants of the glutamate agonist with the glutamate receptor ligand binding domain.* Biochemistry, 2005. **44**(34): p. 11508.
123. Kampen, N.G.V., *Stochastic Processes in Physics and Chemistry*. 3rd ed. 2001, Amsterdam: North Holland.
124. Zwanzig, R., *Nonequilibrium Statistical Mechanics* 1st ed. 2001, New York: Oxford University Press.
125. Hubbard, S.J. and J.M. Thornton, *NACCESS, Department of Biochemistry and Molecular Biology, University College,*. 1993: London.
126. Essiz, S.G. and R.D. Coalson, *Langevin dynamics of molecules with internal rigid fragments in the harmonic regime.* Journal of Chemical Physics, 2007. **127**: p. 104109
127. Ansari, A., *Langevin modes analysis of myoglobin.* Journal of Chemical Physics, 1999. **110**(3): p. 1774.
128. Hayward, S. and H.J.C. Berendsen, *Systematic Analysis of Domain Motions in Proteins from Conformational Change.* Proteins, Structure, Function and Genetics, 1998. **30**: p. 144
129. Groot, B.L.d., S. Hayward, D.M.F.v. Aalten, A. Amadei, and H.J.C. Berendsen, *Domain motions in bacteriophage T4 lysozyme: A comparison between molecular dynamics and crystallographic data* Proteins, Structure, Function and Genetics, 1998. **31**: p. 116.
130. McFeeters, R.L. and R.E. Oswald, *Structural mobility of the extracellular ligand-binding core of an ionotropic glutamate receptor. Analysis of NMR relaxation dynamics.* Biochemistry, 2002. **41**(33): p. 10472.
131. Ahmed, A.H., A.P. Loh, D.E. Jane, and R.E. Oswald, *Dynamics of the S1S2 glutamate binding domain of GluR2 measured using F-19 NMR spectroscopy.* Journal of Biological Chemistry, 2007. **282**(17): p. 12773.
132. Wang, J., J. Onuchic, and P. Wolynes, *Statistics of kinetic pathways on biased rough energy landscapes with applications to protein folding.* Physical Review Letters, 1996. **76**(25): p. 4861.
133. Rico, F. and V.T. Moy, *Energy landscape roughness of the streptavidin-biotin interaction.* J Mol Recognit, 2007. **20**(6): p. 495.
134. Zwanzig, R., *Diffusion in a rough potential.* Proc. Natl. Acad. Sci. USA, 1988. **85**: p. 2029.
135. Brooks, C.L., M. Karplus, and B.M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics, Volume 71*. Advances in Chemical Physics. Vol. 71. 1988, New York: John Wiley & Sons, Inc.
136. Miloshevsky, G.V. and P.C. Jordan, *The open state gating mechanism of gramicidin A requires relative opposed monomer rotation and simultaneous lateral displacement.* Structure, 2006. **14**(8): p. 1241.
137. Miloshevsky, G.V. and P.C. Jordan, *Open-state conformation of the KcsA K+ channel: Monte Carlo normal modes following Simulations.* Structure, 2007. **15**(12): p. 1654.
138. Zwanzig, R., *Nonlinear Generalized Langevin Equations.* Journal of Statistical Physics, 1973. **9**: p. 215.