SANDIA REPORT
SAND 2006-7085
Unlimited Release
Printed November 2006

Substructured Multibody Molecular Dynamics

Paul S. Crozier, Andrei I. Drăgănescu, Gary S. Grest, Ahmed E. Ismail, Richard B. Lehoucq, Rudranarayan M. Mukherjee, Steven J. Plimpton, Mark J. Stevens, and Thomas B. Woolf

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND 2006-7085
Unlimited Release
Printed November 2006

Substructured Multibody Molecular Dynamics

Paul S. Crozier
Multiscale Computational Materials Methods

Andrei I. Drăgănescu and Richard B. Lehoucq
Computational Math/Algorithms

Gary S. Grest
Surface & Interface Sciences

Ahmed E. Ismail
Perf. Assmt. & Dec. Anal.

Steven J. Plimpton
Computational Biology

Mark J. Stevens
Biomolecular Analys & Imaging

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

Rudranarayan M. Mukherjee
Department of M.A.N.E.
Rensselaer Polytechnic Institute
Troy, New York 12180

Thomas B. Woolf
Johns Hopkins University
School of Medicine
Baltimore, Maryland 21205

Abstract

We have enhanced our parallel molecular dynamics (MD) simulation software LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator, lammps.sandia.gov) to include many new features for accelerated simulation including articulated rigid body dynamics via coupling to the Rensselaer Polytechnic Institute code POEMS (Parallelizable Open-source Efficient Multibody Software). We use new features of the LAMMPS software package to investigate rhodopsin photoisomerization, and water model surface tension and capillary waves at the vapor-liquid interface. Finally, we motivate the recipes of MD for practitioners and researchers in numerical analysis and computational mechanics.

Contents

Chapter 1: Substructured molecular dynamics simulations using multibody dynamics algorithms through LAMMPS-POEMS coupling

Rudranarayan M. Mukherjee
Department of M.A.N.E., Rensselaer Polytechnic Institute, Troy, New York 12180

Paul S. Crozier
Sandia National Laboratories, Albuquerque, New Mexico 87185

Abstract

In this paper we outline a method of reduced order modeling of biomolecular systems as sub-structured multi-rigid body articulated systems and the integration of molecular dynamics software with multibody dynamics software to facilitate this modeling effort. We use a recursive $O(n)$ method based on Kane's method for generating and solving the equations of motion. The methodology is verified by simulating several biomolecular systems. The method shows good energy conservation in NVE ensembles and preserves the essential dynamics of the system. We have developed an open-source computational tool by combining a classical molecular dynamics software LAMMPS and a multibody dynamics research code called POEMS. This tool is freely available to all researchers and gains on the complementary nature of the two codes by coupling the efficient force calculation algorithms in LAMMPS with the efficient algorithm in POEMS for generating and solving the coupled equations of motion.

1. Introduction

Molecular dynamics (MD) simulations provide the methodology for detailed fine scale modeling on the molecular level. MD in the most traditional sense can be viewed as a process by which one generates atomic trajectories of a system of particles by direct numerical integration of Newton's equations of motion for each particle, with the appropriate initial and boundary conditions. This type of procedure has the advantage that as the integration/simulation progresses, the simulation yields important information not only about the intermediate states and the mechanisms which produced them, but also the rates at which these processes occur. Additionally, not just these states, but the predicted rates serve as valuable means for validating the models. Unfortunately MD simulation using standard atomistic models quickly run into significant challenges for all but the most elementary systems. This is because classical molecular dynamics (MD) propagates the motion of molecular models by solving the equations of motion for all the atoms in the model. In the fully atomistic case, (i.e. the Newton's equations of motion are derived and solved for every atom of the system), we have the most direct application of the physics involved, with the associated implementation being in many regards conceptually the simplest and easiest to apply. Unfortunately, due to the nature of the molecular interactions, specifically the stiffness of the bonds and other interaction, solving these equations

(though straight forward) requires very fine time steps in order to maintain temporal integrator stability (the highest frequency of the systems must be accurately captured). Depending on the temporal integration method used, at least one system wide force determination (to drive the equations of motion) must be performed and this determination is extremely expensive (often the most costly aspect) for large systems. Thus, though conceptually simple and easy to implement, such simplistic brute force methods grind to an effective halt under the burden of their sub-femtosecond ($<10^{15}$ sec.) required time steps and associated expensive force determinations.

Due to these difficulties many approaches have been developed in an attempt accelerate the simulations. Many efforts have focused on overcoming the strict time step limits in MD simulations. If larger stable integration time steps can be taken, then fewer expensive force determinations (which generally dominate the overall cost) are needed. This accelerates the simulation because the CPU time required is roughly proportional to the number of systems level force determinations executed. These efforts have primarily focused on removing (or at least not considering) the high frequency components of the system. It is these highest frequency components which govern the required temporal integration step size [1]. Examples of these approaches include: Constrained Dynamics Through Explicit Constraints [2], Constrained Dynamics Through Implicit Constraints via Generalized Coordinates [3], Reduced Computational Order (Cost) Algorithms [4], Multirate Temporal Integration [5][6], Eigenvector (Modal) Schemes [7], Implicit integration schemes [8], Reduced Cost Force Determination [9] and the use of Multibody Dynamics Algorithms [4] among others.

Each of these approaches offers it own advantages and real disadvantages. Outwardly these methods appear to have little in common other than the shared objective of performing accurate integration of the equations of motion in less time. Each in fact represents a form of model reduction. In this paper we outline a method of reduced order modeling of molecular systems as sub-structured multi-rigid body articulated systems and the integration of molecular dynamics software with multibody dynamics software to facilitate this modeling effort.

2. Modeling Approach

The fundamental idea behind this work is the coarse-graining of select spatial domains in a molecular dynamics simulation into uncoupled and coupled rigid body systems. This process can be viewed as aggregation where a large number of discreet particles such as atoms or molecules are constrained to move as either a single rigid body or a system of articulated rigid bodies connected by kinematic joints. This modeling approach is a valid if the in the spatial domain in consideration, the relative motion of these discreet particles is limited or localized.

2.1 Kinematic Model

The dynamics of a single rigid body as modeled using Newton-Euler equations of motion would include three degrees of translational motion and three degrees of orientational change. The translational degrees of freedom are easily those associated with the three Cartesian coordinates of the center of mass of the body. The time derivatives of these degrees of freedom give rise to the translational velocity and acceleration of the body. Similarly the time derivatives of the

orientational degrees of freedom (or some combination of the same) may give rise to the angular velocity and angular acceleration of the body. However the choice of the orientational degrees of freedom can be tricky and result in numerical difficulties. The easiest way to model the orientation would correspond to the three Euler angles with each angle associated with a rotation about each of the Cartesian directions. Although commonly used, this choice of degrees of freedom can result in numerical singularities arising from dependency in the kinematic equations relating the orientation angles. To overcome this singularity, quaternion or Euler parameters are used in modeling the orientation. Euler parameters are a set of four parameters related to each other through a constraint equation. Use of these parameters result in robust kinematic equations which never suffer from numerical singularities. The time derivative of the Euler parameters is related to the angular velocity of the body. The angular velocity of the body is the time rate of change of orientation of the body and modeled as having a component about each of the Cartesian directions.

There are two different representations of these reduced order models. The first are single rigid bodies where there is no kinematic coupling between the dynamics of the bodies. The interactions between these rigid bodies are modeled explicitly through the use of force-fields such as all-atom or unified atom interactions. The second type is coupled rigid bodies. Two rigid bodies are coupled when they share a common atom. This common atom is treated as a kinematic joint location. In these models, along with the force-field interactions, there exist constraint inertial loads between bodies. This is because the bodies are coupled at the common atom location and a kinematic constraint exists between any two bodies in the coupled system. A set of coupled rigid bodies are hence forth referred to as chains. A system may have any number of chains in it. Different chains interact only through the force field interactions as there is no inertial coupling between different chains.

Another type of reduced order model which again reduces to an articulated chain topology is the axial bond constrained systems. The fully atomistic representation of these systems consists of atoms or beads connected to each other by stiff joints. In the reduced order model, the axial stiff spring is replaced by a constant length massless rigid link. Each link and the next bead it is attached to is treated as a single body with unit mass and negligible inertia. The successive bodies are connected to the base body by joints allowing only rotational degrees of freedom.

In our model, the chains are free floating. The base body is modeled as connected to the inertial reference frame by a six degree of freedom joint allowing relative translational and rotational degrees of freedom. Each joint is modeled using Euler parameters to avoid any singular configurations. Unlike the SHAKE or RATTLE formulations which require additional nonlinear equations to be solved to impose the constraints, our model does not require solving any additional equations for maintaining the constraints. Further while SHAKE and RATTLE impose the constraints iteratively and only to a specified tolerance, our modeling approach enforces the constraints non-iteratively and exactly with no constraint violation. The constraints are imposed implicitly through the use of relative (internal) coordinates which reduces the formulation to a minimum set of ordinary differential equations.

2.2. Generating the rigid body properties

The calculation of the total mass, position of the center of mass and the velocity of the center of mass of any rigid body is a simple from the properties of all the atoms that are aggregated into that rigid body. However calculating the inertia matrix and the angular velocity of the bodies is more involved. The calculation of the inertia matrix of the body involves taking the second moment of the mass of each atom about the center of mass of the rigid body. As this inertia matrix would vary with the motion of the body, the temporally invariant principle moments of inertia are calculated by solving an eigenvalue problem from the calculated inertia matrix. The eigenvalue problem also produces the three principle directions associated with the principle moments. These directions are treated as the basis vectors of the body based reference frame thereby forming the transformation matrix from the body basis to the Newtonian basis. To calculate the angular velocity, the angular momentum of the system of atoms is calculated about the center of mass of the body by summing the first moment of momenta of each atom. It is then converted to the body basis. This now equals the product of the diagonal inertia matrix and the unknown angular velocity, which can now be easily calculated by a scalar division.

3. Algorithm Overview

In this section an overview of the recursive $O(n)$ algorithm is presented. This algorithm uses the projection method as promoted by Kane and others [10] and uses internal coordinates instead of the Cartesian coordinates to formulate and solve the equations of motion. The algorithm begins by generating a topology or connectivity map of the chains in terms of relative coordinates, joints and body fixed reference frames. Each body is associated with its own body fixed dextral set of unit vectors. The joint locations, inertia values and the generalized velocities are expressed with respect to the body fixed reference frames. One end of the chain is chosen as the base body and it is connected to the inertial reference frame by a kinematic joint. Each successive body is connected by two kinematic joints, one to an inward body and the other to the outward body on the chain. The orientation and motion of a body is expressed in terms of the admissible degrees of freedom of the joints. The linear and angular velocities of a body are expressed as invertible linear combinations of the generalized speeds i.e. time derivatives of the relative degrees of freedom. The algorithm works in three recursive sweeps or traversals. The first sweep begins at the base body and moves outwards to the tip while recursively generating the kinematic preliminaries like partial velocities [10], inertias and applied forces from the known states of the system. The second traversal then begins at the tip and recursively moves inwards towards the base body. This is a triangulation traversal and it recursively generates the articulated compound inertias and forces. This traversal is equivalent to successively shifting the inertias and the forces inwards towards the base body. Since the boundary conditions are known at the base body, at the end of the triangulation sweep the equations of motion for the base body can be solved to generate the state derivatives of the base body. Starting at the base body the third traversal works outward while recursively solving for the state derivatives of each successive body. By the time the sweep ends at the tip, the state derivatives are all solved in an efficient $O(n)$ complexity.

3.1. Mathematical Preliminaries

To aid in the subsequent development, consider the notation associated with the description of an arbitrary set of interconnected rigid bodies shown in figure (1). For this system, proximal (parent) body $Pr[k]$ is connected to its child body k through joint-k, via joint points k^- and k^+ which reside in bodies $Pr[k]$ and k, respectively. Similarly, the *distal* (child) bodies of body k are given as members of the set of bodies $Dist[k]$. The position vector $\mathbf{s}^k$ locates joint-k relative to the mass center of body $Pr[k]$, while the position vector $\mathbf{r}^k$ locates the mass center of body k with respect to the outboard end of this same joint. It will also prove convenient to describe the position of child mass center k^* relative to proximal mass center $Pr[k]^*$ by the vector $\vec{\gamma}^k$.

1. Notation Associated with an Arbitrary Set of Interconnected Rigid Bodies

The angular velocity of any body k with respect to the Newtonian reference frame N, and velocity of its associated mass center k^* may always be written in terms of the generalized speeds as

$$\boldsymbol{\omega}^k = \sum_{r=1}^{n} \boldsymbol{\omega}_r^k u_r + \boldsymbol{\omega}_t^k \tag{1}$$

and

$$\mathbf{v}^{k^*} = \sum_{r=1}^{n} \mathbf{v}_r^{k^*} u_r + \mathbf{v}_t^{k^*}. \tag{2}$$

In these expressions $\boldsymbol{\omega}_r^k$ and $\mathbf{v}_r^{k^*}$ are termed the r^{th} *partial angular velocity of body* k and r^{th} *partial velocity of point* k^*, in N, respectively. These quantities may be thought of as basis vectors for the space of admissible system velocities and angular velocities, while the associated generalized speeds are the velocity space measure numbers. Additionally, the terms $\boldsymbol{\omega}_t^k$ and $\mathbf{v}_t^{k^*}$ appearing in equations (1)–(2), are referred to as the *angular velocity remainder term* of body k and *velocity remainder term* of point k^*, in N, respectively. These quantities are most often associated with specified/prescribed motion, and thus are not associated with the time derivative of a system degree of freedom.

When deriving this method it is often convenient to express quantities in a scalar matrix, as opposed to a tensor (vector and dyadic) form. For this purpose an arbitrary vector $\boldsymbol{\vartheta}^k$ will be represented in matrix form as $\underline{\vartheta}^k$, which is associated with the local dextral orthogonal unit vectors $\hat{k}_1$, $\hat{k}_2$, $\hat{k}_3$, fixed in body k. One may then define the *velocity*, *partial velocity*, and *velocity remainder term* matrices as

$$\underline{V}^k = \begin{bmatrix} \underline{\omega}^k \\ \underline{v}^{k^*} \end{bmatrix}, \qquad \underline{P}_r^k = \begin{bmatrix} \underline{\omega}_r^k \\ \underline{v}_r^{k^*} \end{bmatrix}, \quad \text{and} \quad \underline{V}_t^k = \begin{bmatrix} \underline{\omega}_t^k \\ \underline{v}_t^{k^*} \end{bmatrix}. \tag{3}$$

With these matrices so defined, equations (1) and (2) may be expressed as

$$\underline{V}^k = \overline{\underline{V}}^k + \underline{V}_t^k = \sum_{r=1}^{n} \underline{P}_r^k u_r + \underline{V}_t^k. \tag{4}$$

One can similarly represent the generalized acceleration matrix of an arbitrary body k as defined in previous works [11], [12], as

$$\underline{A}^k = \begin{bmatrix} {}^N\underline{\alpha}^k \\ {}^N\underline{a}^{k^*} \end{bmatrix}, \tag{5}$$

where $\underline{A}^k$ may also be divided into two portions. One is $\overline{\underline{A}}^k$, which contains all terms which are explicit in the unknown state derivatives $\dot{\underline{u}}$ and the other is the *acceleration remainder term* $\underline{A}_t^k$, which represents all of the other acceleration terms (and may be calculated directly from the system state), giving

$$\underline{A}^k = \overline{\underline{A}}^k + \underline{A}_t^k. \tag{6}$$

4. $O(n)$ Forward Dynamics Analysis

The basic recursive $O(n)$ algorithm for performing forward dynamics simulation associated with tree-structure systems consists for three principal steps, or "sweeps". These steps are the *Kinematics Sweep*, the *Triangularization Sweep*, and the *Back Substitution Sweep*.

4.1. Recursive Kinematic Relationships

The Kinematic sweep starts at the base body and works outward to the tip while recursively using the kinematic relations to generate the partial velocities, transformation matrices, translational and rotational velocities and state dependent acceleration terms in the body basis. With the generalized velocity, generalized acceleration, and generalized acceleration remainder term matrices as represented above, it is possible to compactly represent the recursive relationships necessary for determining all system kinematic quantities. As has been demonstrated in [11] we have

$$\underline{V}^k = [(\underline{S}^k)^T \overline{\underline{V}}^{Pr[k]} + \underline{P}_k^k \underline{u}_k] + \underline{V}_t^k, \tag{7}$$

and

$$\underline{A}^k = [(\underline{S}^k)^T \overline{\underline{A}}^{Pr[k]} + \underline{P}_k^k \underline{u}_k] + \underline{A}_t^k. \tag{8}$$

The quantity $\underline{S}^k$ appearing in equations (7)–(8) is the basis consistent linear transformation matrix

$$\underline{S}^k = \begin{bmatrix} \underline{C}^k & \underline{C}^k \underline{\gamma}_\times^k \\ \underline{0} & \underline{C}^k \end{bmatrix}_{6\times 6}. \tag{9}$$

Within this expression $\underline{C}^k \equiv^{Pr[k]} \underline{C}^k$ is the direction cosine matrix which relates the body k basis vectors to those fixed in its parent body $Pr[k]$; $\underline{0}$ is a 3×3 zero matrix; and $\underline{\gamma}_\times^k$ is the skew symmetric matrix equivalent to the vector cross product operation $\gamma^k\times$. The *shift matrix* transformation $\underline{S}^k$ converts a system of forces and moments acting through the center of mass of k, to an equivalent force system, acting though a point of k which is instantaneously coincident with the center of mass of $Pr[k]$.

At this time, it is also convenient to define the body k *generalized inertia* $\underline{I}^k$ and the body k *generalized force* $\underline{F}^k$ matrices

$$\underline{I}^k = \begin{bmatrix} \underline{I}^{k/k^*} & \underline{0}(10) \\ \underline{0} & \underline{M}^k \end{bmatrix}_{6\times6}, \tag{11}$$

$$\underline{F}^k = \begin{bmatrix} \underline{T}^k - (\underline{I}^{k/k^*}\underline{\alpha}_t^k + \underline{\omega}_\times^k \underline{I}^{k/k^*}\underline{\omega}^k)(12) \\ \underline{R}^k - \underline{M}^k \underline{a}_t^k \end{bmatrix}_{6\times1}. \tag{13}$$

Within these expressions, $\underline{I}^{k/k^*}$ is the 3×3 central inertia matrix of body k, and $\underline{M}^k$ is the diagonal translational mass matrix of this same body. By comparison $\underline{T}^k$ and $\underline{R}^k$ represent the resultant force system of all moments and forces, respectively, acting on body k through its center of mass k^*.

4.2. Triangularization

The triangularization procedure works recursively inward to compute the articulated inertia mass matrix $\mathbf{I}_3^k$ and the articulated generalized active force $\mathbf{F}_3^k$ expressing as

$$\mathbf{I}_3^k = \mathbf{I}_1^k + \sum_{j\in Dist[k]} \mathbf{T}^j \mathbf{I}_3^j (\mathbf{S}^j)^T, \tag{14}$$

$$\mathbf{F}_3^k = \mathbf{F}_1^k + \sum_{j\in Dist[k]} \mathbf{T}^j \mathbf{F}_3^j, \tag{15}$$

where $Dist[k]$ is the *distal* (children) set associated with body k. The triangularization operator $\mathbf{T}^j$ and the basis consistent shifting operator $\mathbf{S}^j$ used in (14)-(15) are defined as

$$\mathbf{T}^j = (\mathbf{S}^j)^T \left[\mathbf{U} - \frac{1}{M_j} \mathbf{I}_3^j \mathbf{P}_j^j (\mathbf{P}_j^j)^T \right], \tag{16}$$

$$\mathbf{S}^j = \begin{bmatrix} \mathbf{C}^j & 0(17) \\ 0 & \mathbf{C}^j \end{bmatrix} \begin{bmatrix} \mathbf{U} & \gamma_\times^j(18) \\ 0 & \mathbf{U} \end{bmatrix}. \tag{19}$$

In (16) and (19), matrix $\mathbf{C}^j$ is the direction cosine matrix relating to local basis vector of body $Dist[k]$ to k, $\mathbf{U}$ is an identity matrix, and $\gamma_\times^j$ is the matrix representation of the vector cross product. The quantities M_k is also given by

$$M_k = (\mathbf{P}_k^k)^T \, \mathbf{I}_3^k \, \mathbf{P}_k^k \ , \tag{20}$$

with $\mathbf{P}_k^k$ defined as

$$\mathbf{P}_k^k \equiv \begin{bmatrix} {}^N\omega_k^k (21) \\ {}^N\mathbf{v}_k^k \end{bmatrix} . \tag{22}$$

4.3. Back-Substitutions

At the base body, information associated with an entire set of outboard bodies has all been accumulated and is explicitly available such that the equation $M_1 \dot{\mathbf{u}}_1 = \mathbf{P}_1^1 \mathbf{F}_3^1$ can be isolated and yields the solution of $\dot{\mathbf{u}}_1$. The solution for $\dot{\mathbf{u}}_1$ is then substituted into the next equation to solve for $\dot{\mathbf{u}}_2$. Proceeding in this manner, a generalized function expression for the solution of each generalized acceleration $\dot{\mathbf{u}}_k$ is given as follows

$$\dot{\mathbf{u}}_k = \frac{(\mathbf{P}_k^k)^T}{M_k} [\, \mathbf{F}_3^k - \mathbf{I}_3^k \, (\mathbf{S}^k)^T \, \mathbf{A}^{k-1} \,] \, , \tag{23}$$

with $\mathbf{A}^k$ computed from

$$\mathbf{A}^k = (\mathbf{S}^k)^T \, \mathbf{A}^{k-1} + \mathbf{P}_k^k \, \dot{\mathbf{u}}_k \ . \tag{24}$$

5. Time Integration

Velocity-Verlet temporal integration scheme for temporally advancing a dynamics simulations has been extensively used for atomistic simulations. Velocity-Verlet algorithms are symplectic and gives very good energy conservation characteristics. However for reduced order models involving coupled bodies, the performance of the Velocity-Verlet is not as good as with atomistic simulations. This is because the velocity dependent inertial forces such that the gyroscopic and Coriolis forces come into play for the reduced order models. Further, the motion of individual bodies in the articulated system is coupled and the motion of one affects all others resulting in constraint forces acting on the bodies. Also, as compared with an atomistic simulation, the integrands are not the Cartesian accelerations and velocities but the time derivatives of the generalized coordinates. For these reasons, the multibody dynamics equations of motions have been traditionally integrated by high order methods such as the Runge-Kutta4-5 schemes. These schemes have traditionally been highly accurate and give good energy conservations for macroscopic problems like aerospace applications. However when applied to reduced order molecular dynamics simulations, these methods quickly become computationally

expensive as they require four expensive force calculations per integration step. For the systems studied in our work, the Lobatto III a-b partitioned Runge-Kutta integration scheme has been used. It is a second order method which iteratively calculates the velocities at the half step. The iteration quickly converges in one or two steps. The method efficiently accommodates for the velocity dependent inertial forces and requires only one force calculation per integration step.

6. Applications and Results

Discussed in this section are the test cases simulated to verify the validity of this development. The primary objective of simulating these test cases is to be able to reproduce the previously published results, and validate the stability of the simulations. Because of space constraints, the details of the results are not shown and can be found in an upcoming journal article.

6.1. Water Box

The fundamental test case to check the coupling of the two pieces of software is a box of waters consisting of 512 water molecules in NVE ensemble using different time steps of 0.5, 1 and 2 fs for a total simulation length of 100 ps. Each water molecule is treated as a single rigid body and the mass properties are calculated using the method outlined above. All atom force fields are used under periodic boundary conditions. The dynamics of the rigid bodies is simulated using POEMS and compared with the results obtained by imposing holonomic constant bond length constraints using SHAKE. While in SHAKE the Newton-Euler equations of motion are solved directly to generate the dynamics, in POEMS, Kane's [10] equations of motion are used. In either case each rigid body is modeled as having three degrees of translational freedom modeled using Cartesian displacements and three degrees of rotational freedom modeled using body based reference frames and Euler parameters. In SHAKE the temporal integration was carried out using Velocity Verlet while the Lobatto III A-B scheme was used in POEMS. The energy conservation in the simulations is calculated using the ratio of standard deviation in energy to the mean energy value as the comparison metric. The simulations at different time steps showed good energy conservation with smaller time steps giving better energy conservation. To analyze the simulation results we calculated the mean square displacement of the water molecules and the coefficient of diffusion using Einstein's equation. We also compared the thermodynamic properties using the POEMS and SHAKE approach. The results between the two sets of simulations were found to be in very good agreement.

6.2. Alanine Dipeptide

This molecule, CH3CONHCH(CH3)CONH, is a small enough molecule to avoid complexities of structure and yet has interesting dynamic behavior which makes it a prime candidate for initial simulation test to validate the method. The molecule has been sub-structured into two rigid bodies each corresponding to the -CONH unit while the remaining atoms are treated as bond constrained particles. All atom forces were calculated using LAMMPS while solving the multibody dynamics equations of motion of the sub-structured articulated system and the temporal integration were handled by POEMS. The simulation results, including energy

conservation, structural properties and configurational parameters were found to match with those generated using atomistic simulations. Further, because of the coarse-graining, an improvement of an order of magnitude in the integration time step was achieved.

6.3. DNA simulations

In these simulations we modeled the bond constrained dynamics of tethered DNA strands of lengths of 16 and 32 atoms. The fully atomistic interactions consist of FENE bonds and truncated Lennard Jones and Coulomb interactions. The DNA strands are tethered to a membrane which is modeled using a 9-3 Lennard Jones wall potential. In the reduced order model, the FENE bonds are constrained to fixed lengths modeled as massless rigid links. This reduces it to an articulated serial chain system modeled using POEMS. The simulation results were compared and validated against an atomistic simulation and an improvement in the integration time step of up to a factor of 6 was achieved.

6.4. Box of alkanes

We simulated a box of alkanes to validate the performance of the modeling scheme with a united atom potential. The systems under consideration were three boxes containing 216 chains of alkanes each of chain lengths 8,16 and 32 modeled under periodic boundary conditions in a NVE ensemble. The united atom Trepp3 force field was used in these simulations. The axial vibrations of the beads were constrained by modeling the stiff bonds as fixed length massless rigid links. This rendered the model as articulated chains of point masses connected by rigid links and kinematic joints. Three different time-steps of 1fs, 5fs and 10fs were used in the simulations. For these simulations too we used the energy metric discussed above to compare the simulations results with those obtained using SHAKE. The united atom force fields are smoother than the Lennard Jones potential and hence better energy conservation at larger time steps were expected using the articulated rigid body representation. We were able to achieve stable simulations with good energy conservation with increase in time steps by an order of magnitude.

6.5. C-Terminal of Ribosomal

This is a larger problem which provides better understanding of the performance of the methodology for complex biological systems. The C-terminal fragment (1CTF) of the L7/L2 ribosomal protein from E. coli is used for this simulation. This system has been simulated in several references with different substructuring schemes. In this method the system is sub-structured into 31 small rigid bodies with hinges at the ϕ or ψ angles. The simulation was monitored for energy conservation and preservation of essential dynamics.

6.6. C-Terminal of Rubisco

We have built a model of the RuBisCO (Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase) enzyme for simulation using our coupled LAMMPS POEMS simulation software. The fully

atomistic model of the C-terminal of RuBisCO consists of 510 atoms modeled using harmonic bond potentials, CHARMM angle and dihedral potentials and non-bonding CHARMM Lennard-Jones and Coulomb force fields with cut offs at 8 and 10. In the sub-structured model, the system consists of 11 rigid bodies connected together to form an articulated serial chain topology. All atom explicit force calculations are supported by this model. However intra-body interactions between the atoms that make up a rigid body are ignored as these would sum to zero. This is a modest sized problem which is a good example to validate the modeling approach. Using this model, NVE simulations were run for 1 nano-second. Different time steps were used to determine the drift in energy as a function of time-step. Ignoring the intra-body atomistic interactions give an immediate computational saving. Further, by using a rigid body model, an increase in the integration time step by an order of magnitude was observed.

6.7. Rhodopsin

Rhodopsin is a G-protein coupled receptor with a defined tertiary structure. It is a good example to study transduction and a large amount of experimental results are available about its structure and function. We have generated a sub-structured articulated rigid body model of the rhodopsin protein and have simulated the same with the LAMMPS-POEMS coupling. This is a fairly large system with the fully atomistic model consisting of about 5000 atoms. This atomistic model is sub-structured into 26 connected rigid bodies that form an articulated serial chain topology. Similar to the RuBisCO model, the Rhodopsin model was simulated in NVE ensemble for 1 nanosecond at different time steps. The simulations showed good energy conservation at larger time steps. The coarse-graining provided significant computational savings in the calculations of the force interactions as the intra-body interactions were not calculated. Further using this model, we were able to obtain stable simulation with an increase in the time step by an order of magnitude.

7. Software Development

In this section the generation of the open source computational tool is discussed. The two research codes that are fundamental in this work are the LAMMPS: **L**arge-scale **A**tomic/**M**olecular **M**assively **P**arallel Simulator software from Sandia National Laboratories and the multibody dynamics software POEMS: **P**arallelizable **O**pen-source **E**fficient **M**ultibody **S**oftware. LAMMPS is an open source code, with a GPL type license. Under development by the primary author, Steve Plimpton, and others since the mid 1990s, LAMMPS is a general purpose classical molecular dynamics code [13]. The POEMS code [14] is also open source, with a BSD type license. This is a general purpose multibody dynamics research code being developed by Rudranarayan Mukherjee and other members of Prof. Anderson's research group at Rensselaer Polytechnic Institute. The two pieces of software have different functionalities which are complementary in nature. While LAMMPS is a classical molecular dynamics code with emphasis on atomistic simulations, POEMS is a multibody dynamics code with an emphasis on modeling dynamics of reduced order or coarse-grained models. A brief overview of both these pieces of software is presented in the next section.

LAMMPS is a classical molecular dynamics code that models an ensemble of particles in a liquid, solid, or gaseous state. LAMMPS can model atomic, polymeric, biological, metallic, or granular systems using a variety of force fields and boundary conditions. It runs efficiently on single-processor desktop or laptop machines, but is designed for parallel computers. In classical molecular dynamics, inter-particle force calculations are the most expensive part of MD simulations and hence have been carefully optimized and made more efficient as MD codes have matured. Among the many force fields (FF) currently available in LAMMPS are the commonly-used CHARMM [15] and AMBER [16] biomolecular FF. LAMMPS also has all of the capabilities commonly required in biomolecular simulation, including full long-range electrostatics capabilities using Ewald or particle-particle/particle-mesh (PPPM, similar to particle-mesh Ewald), SHAKE bond and angle constraints, rRESPA [17] hierarchical timestepping, and NVE, NVT, and NPT integrators. LAMMPS also has the capability to simulate hybrid bio/non-bio systems through the superimposing of force fields.

Though significant speedup can be gained from efficiently calculating the forces, which is the forte of LAMMPS, further substantial computational gains can be realized if the systems are coarse-grained by enforcing kinematic constraints. This is because imposing kinematic constraints can efficiently eliminate high frequency components thereby allowing larger temporal integration time-step. This results in a multiplying effect on improving simulation speed through combining larger integration step size and reduce force calculation costs. However, generating and solving the equations of motion of reduced order models particularly those which represent coupled multi-body systems can be challenging and unless some efficient formulation is resorted to, the computational cost can be as high as $O(n^3)$ where n is the number of degree of freedom in the system. LAMMPS integrates Newton's equations of motion for collections of atoms, molecules, or macroscopic particles and does not include any efficient formulations for effectively formulating and solving the equations of motions of reduced order models.

This aspect has been addressed in POEMS. POEMS was written as a generic multibody simulation code which can efficiently handle the dynamics aspect of the reduced order models. POEMS is an object-oriented C++ research package for simulating the forward dynamics of multibody systems. Its emphasis was also placed on application to large ($n \gg 1$) systems, i.e. coupled systems involving many generalized coordinates. Majority of the effort in algorithm development has been oriented toward methods that perform well with applied to coupled systems involving many generalized coordinates. This code features libraries of different dynamics formulations for efficiently generating and solving the equations of motion of articulated system as well as different time integration schemes for advancing a simulation temporally. Commonly used kinematic and dynamic identities, organization of multibody topologies, and data structure with matrix manipulations which are generic to most multibody algorithms are built into the software using an object oriented design.

The software in its current form has three algorithms for solving equations of motion of articulated multi-rigid body systems in chain and tree topologies.

- *KaneSolver()* : The $O(n^3)$ complexity solver based on Kane's method [10].
- *OnSolver()* : The $O(n)$ complexity recursive solver based [12].
- *DCASolver()* : The Divide and Conquer method of $O(log(n))$ complexity [18].

It also contains an implementation of a generalized impulse momentum formulation for correct kinematic coarse-graining of the reduced order models. This is a novel feature that enforces the correct initial conditions required to preserve the essential dynamics of the systems. No other research package has a comparable formulation and it is a significant development as it allows smooth transition between models of different resolutions. Because of space constraints this algorithm is not discussed here and presented in another upcoming research paper.

Along with the solution of equations of motion, the software has temporal integration algorithms for temporal simulations. These include the following algorithms.

Runge Kutta 4-5	*Predictor Corrector*
Verlet and Velocity Verlet	*Lobatto Partitioned III a-b Runge Kutta 4-5*

By coupling together these two pieces of software we have created a synergistic simulation tool which is freely available to all researchers working in molecular dynamics. The inherent features of these codes are complementary, with LAMMPS focused more on the efficient generation of force field and potential calculations while POEMS is aimed at efficiently handling the dynamics aspect of the reduced order models. This two fold approach is instrumental in accelerating molecular dynamics simulations and be applicable to a wide variety of problems in biomolecular and materials modeling. POEMS is built into LAMMPS as an external library and is distributed along with LAMMPS.

8. Conclusions

A novel method based on sub-structured coarse grained models of molecular dynamics systems is developed, implemented and validated. This method uses efficient $O(n)$ complexity multibody dynamics algorithms for modeling the forward dynamics of these sub-structured models. A new computational tool is developed and released for public use under open source licensing. This computational tool culminates from coupling together the molecular dynamics code LAMMPS with the multibody dynamics software POEMS.

Acknowledgments

The authors thank Steve Plimpton and Kurt Anderson for their support and help in this endeavor.

References

[1] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer, 2nd edition, 1993.

[2] H. C. Anderson. Rattle: A 'Velocity' version Shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 54:24–34, 1983.

[3] R. A. J. Abagyan and A.K. Mazur. New methodology for computer-aided modeling of biomolecular structure and dynamics. *Journal of Biomolecular Structure*, 6:833–845, 1989.

[4] H. M. Chun, C. E. Padilla, D. N. Chin, M. Watenabe, 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*, 21(3):159–184, 2000.

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

[6] M. Watanabe and M. Karplus. Dynamics of molecules with internal degrees of freedom by multiple time-step methods. *Journal of Chemical Physics*, 99:8063–8074, 1993.

[7] G. Zhang and T. Schlick. LIN: A new algorithm combining implicit integration and normal mode techniques for molecular dynamics. *Journal of Computational Chemistry*.

[8] C. S. Peskin and T. Schlick. Molecular dynamics by the backward Euler's method. *Communications in Pure and Applied Math*, 42:1001–1031, 1989. in press.

[9] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *Journal of Computational Physics*, 135(2):280–292, 1997.

[10] T. R. Kane and D. A. Levinson. *Dynamics: Theory and Application*. Mcgraw-Hill, NY, 1985.

[11] K. S. Anderson. An order-n formulation for motion simulation of general constrained multi-rigid-body systems. *Computers and Structures*, 43(3):565–572, 1992.

[12] K. S. Anderson. An order-n formulation for the motion simulation of general multi-rigid-body tree systems. *Computers and Structures*, 46(3):547–559, 1993.

[13] J. M. Haile. *Molecular Dynamics Simulation : Elementary Methods*. Wiley Interscience, New York, 1992.

[14] K.S. Anderson, R. Mukherjee, J.H. Critchley, J. L. Ziegler, and S.R. Lipton. Poems: Parallelizable open-source efficient multibody software. *Engineering with Computers*, 2005. In Review.

[15] M. Karplus. CHARMM(Chemistry at Harvard Macromolecular Mechanics). http://brooks.scripps.edu/, 2005. Brooks Group Computational BioPhysics and Chemistry.

[16] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. R. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman. AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Computer Physics Communications*, 91:1–41, 1995.

[17] S. J. Plimpton and M. Stevens. Particle–mesh Ewald and r-RESPA for parallel molecular dynamics simulations. In *Proceedings Eighth SIAM Conference on Parallel Processing for Scientific Computing*, Minneapolis, MN, March 1997. DETC2005-85480.

[18] R. Featherstone. A divide-and-conquer articulated body algorithm for parallel $O(\log(n))$ calculation of rigid body dynamics. Part 1: Basic algorithm. *International Journal of Robotics Research*, 18(9):867–875, Sep. 1999.

Chapter 2: How a small change in retinal leads to G-protein activation: initial events suggested by molecular dynamics calculations

Paul S. Crozier and Mark J. Stevens
Sandia National Laboratories, Albuquerque, New Mexico 87185

Thomas B. Woolf
Johns Hopkins University, School of Medicine, Baltimore, Maryland 21205

Abstract

Rhodopsin is the prototypical G-protein coupled receptor, coupling light activation with high efficiency to signaling molecules. The dark-state x-ray structures of the protein provide a starting point for consideration of the relaxation from initial light activation to conformational changes that may lead to signaling. In this study we create an energetically unstable retinal in the light activated state and then use molecular dynamics simulations to examine the types of compensation, relaxation, and conformational changes that occur following the *cis*-*trans* light activation. The results suggest that changes occur throughout the protein, with changes in the orientation of Helices 5 and 6, a closer interaction between Ala 169 on Helix 4 and retinal, and a shift in the Schiff base counterion that also reflects changes in sidechain interactions with the retinal. Taken together, the simulation is suggestive of the types of changes that lead from local conformational change to light-activated signaling in this prototypical system.

Introduction

Rhodopsin is an excellent system for understanding the details of G-protein coupled receptors (GPCR) due to the large amount of experimental information related to both structure and function (for recent reviews, see Refs. [1; 2; 3; 4; 5; 6]). It is also the first GPCR with a measured tertiary structure[7] and is thus an excellent candidate for yielding insight into the molecular details of GPCR function. Explicit, all-atom molecular simulation can provide a view into the choreographic details of the structure-function relationship. A full understanding of these details is difficult, however, due to the large separation in time-scales between the photocycle of rhodopsin and current computational limits in computer simulation of biomolecules. In particular, the full photocycle occurs on the millisecond time scale,[8] while the state-of-the-art in computer simulation of large proteins is tens of nanoseconds.

Bovine rhodopsin has served as a model system for the understanding of transduction for many years.[8] In particular, studies of bovine rods have led to initial understanding of G-protein coupled systems, to the first GPCR that was sequenced,[9] and to understanding of the connections between particular residues and rhodopsin function.[3; 10] For example, the role of Glu 113 as the counterion,[11] the critical role of certain residues in transduction,[12; 13] and initial

suggestions for spectral tuning[14; 15] all began with rhodopsin. An upcoming frontier is understanding the connections between the photocycle's underlying conformational changes that lead to activation and signaling, and the structures of the G-protein itself [16; 17].
A key aspect of understanding rhodopsin and other GPCRs is the dynamic motion of the ligand activation.[18; 19] In rhodopsin, the *cis-trans* photoisomerization of retinal is the activation mechanism. The structural and energetic consequences of retinal's isomerization are of great interest. Molecular dynamics (MD) simulations offer a means of obtaining atomic-scale dynamics of such systems. A main limitation has been the short time scales attainable in all-atom MD simulations. However, we have performed a 150-ns simulation — long enough to examine important dynamic events along the path from the dark-adapted to the light-adapted state. The transition to the LUMI intermediate state takes about 150 ns. Thus, we are within range of an early rhodopsin intermediate and can compare to corresponding experimental data. By constraining the C11-C12 dihedral angle of retinal in the simulation, we force the isomerization. In the subsequent dynamics, the constraint is turned off. We obtain a single trajectory of the consequences of the isomerization on the structure and energetics of rhodopsin. Within this comparison, we recognize that the simulation yields only one pathway to the state at 150 ns, not the statistical ensemble that actually exists. However, some aspects of the dynamics are highly probable and will occur for most trajectories. These aspects can safely be expected in the dynamics of our simulation.

The details of the simulation are given in the methods section. We then describe the results of the simulation analysis concerning the structural dynamics and the energetics of our model rhodopsin system that are observed after the forced photoisomerization, including changes in retinal's dihedral angles, a narrowing of the distance between Ala 169 and retinal's ionone ring, helix kink and tilt angle transitions, a switch in the protonated Schiff base counterion, and changes in retinal's interaction with its environment, including disengagement from Helix 6. These events are connected to other large-scale transitions in the post-photoisomerized state of rhodopsin and lead towards coupling with transducin.

Simulation method

The present work is a continuation of our earlier simulation of the dark-adapted state of rhodopsin.[20] In this earlier work, MD simulations were performed using an all-atom representation. The lipid and water environment were explicitly treated. The CHARMM force field was used (version 22 for protein and version 27 for lipids, both released in August of 1999)[21; 22] which includes parameters defined for retinal.[23] All calculations started from the first X-ray structure of rhodopsin (1F88).[7] To be consistent with our previous work, we did not use the new rhodopsin dark-adapted structures[24; 25; 26; 27] as our starting point. Analysis of our earlier dark-adapted rhodopsin simulation has suggested relatively small changes from the original structure. The total system size (41,623 atoms) consisted of protein, 99 DOPC lipids, 100 milli-molar salt concentration (14 sodium, 16 chloride), palmitylated lipids attached to Cys 322 and Cys 323, and 7441 TIP3 waters.

The LAMMPS[28] (lammps.sandia.gov) molecular simulation package was used to produce a 150-ns trajectory. The initial state is the final state of the 40-ns simulation of the dark-adapted rhodopsin.[20] Periodic images were used in all directions. The simulation was performed at

constant membrane surface area of 55Å x 77Å. The direction perpendicular to the lipid bilayer was controlled at a constant pressure of 1 atm. The temperature was controlled using the Nose-Hoover thermostat at 307 K. All bonds to H atoms were constrained using the SHAKE algorithm and the equations of motion were integrated using the velocity verlet algorithm with a 2 fs timestep. Electrostatic interactions were treated using the particle-particle particle-mesh (P^3M) method.

In the simulation, the *cis-trans* photoisomerization occurs by constraining the C11-C12 dihedral angle. An MD simulation of 200 fs starts with the C11-C12 dihedral angle constrained in the *cis* state and ends at the *trans* state. Thereafter, the 150-ns simulation does not constrain the dihedral angle.

Results

C11-C12 dihedral transition

We begin by discussing the post-isomerization dynamics of retinal itself. Figure 1 (a) shows the dihedral angle of the C11-C12 dihedral in retinal as a function of time, including both the 150 ns after isomerization and the 40 ns before the isomerization (from the earlier simulation of the dark-adapted state).[20] The time before isomerization is represented as negative in this plot and subsequent plots. After isomerization the dihedral angle does not substantially change during the 150-ns simulation. Thus, the C11-C12 dihedral transition is stable. After isomerization the self-energy rises by about 7 kcal/mol (Figure 1 (b)), which is primarily due to the C11-C12 dihedral transition. In the vicinity of t = 20 ns there are fluctuations that bring the energy below the average *cis* state energy. By about 30 ns, the self-energy drops sharply and permanently below the average *cis* state self-energy and subsequently slowly decreases to about 4 kcal/mol below the *cis* state self-energy. This drop in energy is primarily due to the relaxation of dihedrals other than the C11-C12 dihedral. The C9-C10 dihedral angle transitions from 160° to the lower energy 180° state. Also, prior to t = 30 ns, the C6-C7 dihedral exhibits oscillatory behavior between 60° and 150°, but settles near 60° after t = 30 ns, thus preventing the beta ionone ring from being coplanar with the retinal chain. In the first 30 ns after isomerization, the dihedral angles C12-C13 and C8-C9 have large fluctuations, although about the average value of the dark adapted state. These large fluctuations end at about t = 30 ns, when the retinal self energy changes, and the average dihedral angle and the fluctuations return to the values of the dark adapted state. The effect of the dihedral motion involving C9 is to move the C19 methyl group to the same side of retinal as the C20 methyl group.

The net structural changes in retinal can be examined by calculating a pseudo-dihedral angle. The N16 and C7 atoms with their respective H atoms are at opposite ends of the retinal chain. Hence, the H16-N16-C7-H7 dihedral angle is a measure of the linearity of the retinal chain (up to the ionone ring). In the dark-adapted state, the average value is 45°, which indicates the bend and twist of the chain, whereas after photoisomerization, a sudden shift occurs, and the dihedral then fluctuates between 60° and 180° for 30 ns. At t = 30 ns the dihedral angle changes to 180° and stays there for the remainder of the simulation. In this final conformation, the chain is linear and coplanar. The main relaxation after isomerization is the transition to this final flattened all-*trans* conformation.

We have also examined retinal's bond orientations for comparison with available deuterium NMR structural data for retinal in the dark-adapted state of rhodopsin.[29] Bond orientation vectors with respect to the bilayer normal were computed for retinal's methyl groups bonded to C_5, C_9, and C_{13} for both the dark-adapted state and post-photoisomerization. Considering the large instantaneous fluctuations in bond orientation and the limited statistical sampling achieved within the 40-ns dark-adapted simulation, results for the dark-adapted state are in good agreement with the NMR measurements (Table I). The C_5 and C_9 bond vector orientations undergo substantial transitions post-photoisomerization. This observed transition in the simulation could be compared with future NMR measurements of retinal's methyl bond vectors post-photoisomerization.

Retinal – Ala 169 gap narrows

The isomerization of retinal starts a sequence of structural transformations that ultimately results in the light-adapted structure. For comparison to our 150 ns simulation, structural transformations that have occurred in the transition to the LUMI state are the most relevant. The isomerization directly moves the ionone ring. Consequently, the residues neighboring the ionone ring should change after isomerization. We have examined the distance between the ionone ring and Ala 169 as a function of the simulation time. Crosslinking experiments[30] find that Ala 169 and the ionone ring can be crosslinked for rhodopsin in the LUMI state. Figure 2 shows that in the simulation, the separation distance decreases after isomerization with the final separation at about 9Å. The crosslinking experiments imply a shorter separation, which could be achieved via rotation of Helix 4 (since Ala 169 is on the side of Helix 4 opposite to retinal), or by further movement of the ionone ring. One would expect that there is a set of conformations for the LUMI state, and only an (unknown) fraction of them allow the ionone:Ala 169 crosslinking. Our simulation does show that the separation distance is reduced significantly, and that if Helix 4 were to rotate, or if the ionone ring moved further, the crosslinking could occur. In the following sections, we discuss the structural transformations of the helices seen in the simulations. The large-scale motions of helices are naturally slower than the individual motions of residues or of retinal, which limits the helical motion that can be seen in ns-time-scale simulations. However, the 150 ns simulation is sufficiently long to observe transitions in some of the helices.

Helix tilt and kink angle transitions

An important characteristic of membrane proteins is the tilt of the transmembrane helices relative to the lipid bilayer. The tilt is an important mechanism by which the helix can match the bilayer thickness and the corresponding hydrophobic/hydrophilic regions. Kinks in the helices play a similar geometric role. Within the membrane protein there are internal interactions that also influence the structure of the transmembrane helices. The isomerization of retinal alters the interaction between retinal and the helices, which can lead to significant changes in the geometry of the helices. In this manner, the molecular-scale isomerization event can yield subsequent larger-scale transmembrane helix structural changes. Because these are transmembrane helices, the action of the isomerization is propagated across the membrane. Both the cytoplasmic ends of the transmembrane helices and the loops, which interact with the G-protein, are affected. Thus, we have a basic outline of the mechanism by which rhodopsin interacts with the G-protein.
We examined the degree of tilt and kink of each helix as a function of simulation time. Since a kink can split the transmembrane helix into more than one part, with varying tilt angles, we calculate the tilt of each part of the helix. These parts have been defined in terms of the kink centers (see Ref. [20] and the figure captions for details of the definitions). Figure 3 shows tilt

angles as a function of time for Helix segments 5b, 5c, 6b and 7a. These tilt angles show significant changes during the simulation beyond the short time-scale fluctuations of about ± 5°. At about t = 32 ns, the tilt angles for Helices 5b and 5c start to shift considerably. For 5b, the tilt angle rises and continues to slowly rise until the end of the simulation. Segment 5c reaches a steady state by about t = 70 ns and fluctuates by about an average value of 30°. This structural change in tilt angle is also reflected in the change in kink angle for 5b-5c (Fig. 4). There is a sudden drop in the kink angle at t = 32 ns. Ultimately, this kink angle decreases from 25° in the dark adapted state to about 5° at t = 150 ns. This small kink angle reflects the fact that the tilt angles of 5b and 5c become almost identical after the transition. Thus, the main effect is almost complete removal of the kink in Helix 5.

The retinal ionone ring is close to Helix 5. In fact, in the dark state the ionone ring is in contact with Met 207 on Helix 5. The kink in Helix 5 occurs at His 211. In the dark state, Met 207 is between His 211 and retinal. After isomerization, the ionone ring and Met 207 switch sides. His 211 then can and does come in contact with retinal. Simulations of the single Helix 5 in the membrane have the same tilt and kink angles as the helix does in the dark adapted state.[20] This implies that the interactions of Helix 5 with the rest of the rhodopsin protein do not affect the kink angle. The fact that isomerization results in a change in the kink angle implies that the structural changes resulting from isomerization cause the kink angle to almost disappear. For this to happen, some residues or retinal must be influencing the kink. Since His 211 comes into contact with retinal after isomerization, and since His 211 is the hinge point of the kink, the obvious candidate for the kink removal is the interaction of His 211 with the ionone ring. In Figure 5 the interaction energy between His 211 and the ionone ring is given. Before isomerization, the energy switches between two states with energies of –6 and –10 kcal/mol, respectively. After isomerization, the energy drops to –14 kcal/mol. This confirms that the His 211 interaction with the ionone ring increases in strength. In the first 30 ns, this interaction brings His 211 and ionone into position. Near t = 30 ns retinal completes the transition to a straight and mostly planar structure. At about the same time there is the sharp drop in the Helix 5 kink angle (Figure 4). The straightening of retinal and the continued strong interaction between His 211 and the ionone ring pulls on the kink's hinge and reduces the kink angle.

The tilt and kink angles of Helix 6 exhibit a correlated dynamics. Within 70 ns of the forced isomerization, the tilt angle of Helix part 6b decreases from its dark state value of 35° to an average of about 17° (Fig. 3). The kink angle also decreases for the same 70 ns and thereafter oscillates about an average value of 20°. In this case, the tilt angle of section b is the major part of the dynamics; the kink angle change is a result of just 6b's tilt dynamics. While the ionone ring in the dark state is in contact with Ala 269 of Helix 6, this contact does not appear to be the driver for the change in the tilt of Helix section 6b. Helix section 6b moves away from the ionone ring and there is not a strong interaction between the ionone ring and Ala 269 that would force the whole segment to have such a tilt (cf. Figure 5). It is thus more likely that the connection to Helix 7 through the short loop E3 exerts the pull on Helix 6 that alters the tilt of segment b.

Following retinal's *cis-trans* isomerization, interactions between retinal's beta ionone ring and the nearby aromatic side chains of Helix 6 diminish. The distance between them grows due to the movement of retinal itself. Figure 5 shows that the interaction energies between retinal's beta ionone ring and Phe 261, Trp 265 and Tyr 268 on Helix 6 are well below kT after isomerization.

This shows the decoupling of Helix 6 from retinal that will allow subsequent large-scale Helix 6 movement,[31] which is coupled to C3 motion and transducin coupling.

Since Helix 7 includes Lys 296, it is not surprising that its tilt angle is quickly influenced by the retinal isomerization. Figure 3 shows that after isomerization, the tilt angle of Helix section 7a increases slightly from about 7° to 15°. At about t = 70 ns (where the changes in the Helix 6 tilt angle stop), the tilt angle abruptly returns to the dark state value. Correlation between the dynamics of Helices 6 and 7 likewise is not surprising given the strong interactions between their sidechains.

PSB counterion switch

From the helix tilt and kink angle data, it is clear that some event occurs near t = 70 ns that impacts the helix dynamics and structure. The event, a switch in the dominant counterion of the retinal protonated Schiff base (PSB), is itself interesting. In rhodopsin's dark-adapted state, Glu 113 acts as the counterion to the PSB. Some experimental evidence has suggested that Glu 181 is protonated in the dark-adapted state, but transfers its proton to Glu 113 via Ser 186 and then replaces Glu 113 as the PSB counterion.[32] Other work suggests that both Glu 113 and Glu 181 are unprotonated and that both act as the PSB counterion, with Glu 181 dominating in the Meta I state.[64] We have simulated both Glu 113 and Glu 181 in their unprotonated states, and we find that structural changes occur that could lead to the PSB counterion switching from Glu 113 to Glu 181. The simulation shows that the salt bridge between the retinal PSB and Glu 113 breaks near t = 70 ns after photoisomerization and does not form again (see Figure 6 and Figure 7). Figure 6 also shows that at t = 146 ns a much stronger (~20 kcal/mol) interaction between Glu 181 and the retinal PSB is briefly established. This occurs because of a dihedral transition in Glu 181 and in Lys 296 simultaneously occurring to shorten the separation between the two residues.

Important details concerning the breaking of the salt bridge are indicated in the plot of the interaction energies of retinal with its environment (Figure 8). In these calculations we take 'retinal' to include Lys 296. It does not make sense to separate the PSB, in particular. After isomerization the energies remain at the same average value as before isomerization until just before t = 70 ns. At this point, the interaction with the solvent gets much stronger, while the interaction with the rest of the protein weakens by about 35 kcal/mol. We have already noted that the salt bridge between retinal and Glu 113 breaks at t = 70 ns. This loss of the strong binding energy results in the weakening of the interaction with the protein. The large strengthening in retinal:solvent interaction energy implies that some water molecules have moved into the protein close to retinal. This indeed is true as images of the region for times near t = 70 ns show (see Figure 9).

The changing hydrogen bond network involving retinal, including nearby water molecules, is shown in Figure 9, which shows images at t = 64, 65, 66, and 67 ns. At t = 64 ns the Glu 113 salt bridge with the N of Lys 296 is intact. In addition, there is a hydrogen bond between NH of Cys 187 and the same O of Glu 113 that is part of the salt bridge. At t = 65 ns, the NH of Cys 187 switches to the other O of Glu 113 to form a hydrogen bond. At t = 66 ns, a water molecule is visible behind the Glu 113 and Cys 187. At t = 67 ns the water breaks the salt bridge by moving between the N of Lys 296 and the O of Glu 113. For the rest of the simulation, a water molecule is between Glu 113 and Lys 296.

C3 loop

Cytoplasmic loop 3 (C3) is an important loop in the interaction between rhodopsin and its G-protein. Without the presence of the G-protein in the simulation and without being in the active light-adapted state, direct understanding of how isomerization ultimately results in structural changes that impact the G-protein is not possible. Instead, we want to characterize how the structural changes in the helices affect the C3 loop structure. This gives a sense of the dynamic connection between isomerization and the C3 loop configuration.

The C3 loop connects Helices 5 and 6, which undergo substantial structural changes in the 150 ns time period as discussed above. These helix structural transformations impact the structure of the connecting loop. Figure 10 shows images of the C3 loop, where the final configuration of our dark-state simulation is compared with the final state of the 150 ns isomerization simulation. Sections of Helices 5 and 6 are also shown. Substantial motion has occurred in the loop, but this is inconclusive since the loop is primarily within solution and therefore has a large configurational space. The more important aspects of the figure are the differences in the helix positions and their effect on the loop. The end of Helix 6 which is attached to the C3 loop has rotated significantly, bringing the ends of the two helices closer together. The bottom image shows that the C3 loop near Helix 5 is adopting a more helical structure.

Large-scale structural changes

Figure 11 shows a comparison of the rhodopsin structure in the dark state and at t = 150 ns. To distinguish the two states, the dark state helices are gray in the top image. Otherwise each helix is colored differently to identify it. The structure of transmembrane Helices 1, 2, 3, 4 and 7 do not change much within the 150 ns simulation time. As described above, Helices 5 and 6 do undergo significant structural changes. The top image of Figure 11 shows retinal in the *cis* conformation, while the bottom image shows the *trans* conformation. The helices that have substantial structural changes are the ones near the ionone ring of retinal, i.e. Helices 5 and 6. However, Helix 4, which is near the ionone ring for *trans* retinal, does not undergo a structural change. The difference being that Helices 5 and 6 are in contact with retinal at t = 0 and immediately feel the effects of the isomerization of retinal. The changes for these two helices are complete by the time retinal has completed its structural transformations. More specifically, they are complete by t = 70 ns, after the Glu 113:retinal salt bridge has broken. In contrast, Helix 4 is not in contact with retinal at t = 0 and is not strongly affected by the retinal transformations.

Discussion

The relatively recent x-ray structure of rhodopsin provides an essential starting point for detailed consideration of how tertiary structure is linked to function for this key protein.[7] Important issues for understanding this linkage concern how the conformational change in the retinal, induced by light, is coupled into large-scale conformational changes that ultimately lead to G-protein signaling.[1] Further and closely related questions concern the effects of mutations, the membrane environment, water molecules and charges.

Molecular dynamics simulations can begin to address some of these issues, but also have important limitations that should be presented along with the advantages. A particularly important limitation is that the present molecular dynamics capability can explore events near to

the starting state that are relatively rapid (up to several tens of ns). Without a sense of the structure of photo-intermediates, the method can suggest how these intermediates could form, but can't predict them with confidence. This is related to the same time-scale issue, since the number of possible intermediates formed from the dark-adapted rhodopsin structure is very large. Ideally it would be the case that several thousand alternative starting conformational changes in the retinal could be performed and the ensemble set of these changes then used to determine a relative free energy surface for the transition. Within that type of framework, the confidence for prediction of intermediate conformational states would be considerably greater. Taking advantage of the considerable computer resources available through Sandia, we are able to explore, with low statistical confidence, a key structural transition within the photocycle. Thus, our simulations are suggestive of the types of changes that might occur after photo-isomerization, but should not be considered predictions of intermediate structures.

It might be suggested that the current calculations should be augmented with detailed quantum calculations to understand the initial femtoseconds of excitation and the revised energy surface for the light activated transition. Multiple research groups are performing quantum chemical calculations of the initial events in photoisomerization.[33; 34; 35; 36; 37; 38] We wanted to address the longer-time scale (ns towards ms) relaxation of the system following the initial light activation. While quantum chemical calculations have advanced significantly in the last few decades, they continue to be limited by the number of heavy atoms in the system and to explore very fast (fs to ps) time scales. For this reason, we accepted the fact that our characterization of the very early events in light activation are by necessity incomplete. The main point is that the excited, light activated state, will decay quickly, and that how the energy from the isomerized retinal leads to the relaxation and eventual signal activation of G-proteins from the rhodopsin protein are the main targets of our calculation.

Recent work has suggested that rhodopsin may exist in a homodimer state in the native membrane.[39; 40] The implications of that finding are not directly addressed in these calculations, but we can speculate somewhat, based on the dynamics results, on what might happen within a homodimer model that differ from the results of the monomer. In particular, Helix 4 is believed to be involved in the dimerization. Our results are consistent with experimental work showing that Ala 169 (in Helix 4) moves closer to the beta-ionone ring during activation. This may lead to effects on coupling between two monomers.[41] While it is less clear how this could impact signaling, it does suggest that activation of one monomer in a dimer pair would be communicated to the other monomer. This could be viewed as a type of allostery, where the activation mechanism of the second monomer might be shifted to a more sensitive state due to the initial activation of the first monomer.

In the remaining discussion, we will start with the early events of excitation and work our way to the longer-time scale events. We will conclude the discussion with more speculative conclusions based on extrapolation of the results.

Initial events in photoactivation: A plausible candidate for the initial changes in retinal with light has been suggested based on a minimal motional change induced by light, but still consistent with a *cis-trans* isomerization.[33; 34; 35; 36; 37; 38] Our initial model covers that change by forcing the main *cis-trans* isomerization and then allowing relaxation from that state within the CHARMM potential function. The response to this change is then followed throughout the

remainder of the simulation. Similar to ideas examining the effects of the initial velocity distribution on the range of motions examined in a trajectory, we do expect that the starting point will have an impact on the relaxation observed in the calculation. At the same time, there are not sufficient computer resources available to any biophysical research group to create an ensemble set of all candidate transitions and their response properties. Therefore, we took the most straightforwardly possible approach and forced the transition to allow the initial change to be consistent with experimental work. We examine the nature of the coupling between the local conformational change and the larger-scale conformational re-arrangements that ultimately lead to G-protein activation. While we fully accept that this candidate starting transition may be inaccurate on the femtosecond time-scale, we want to emphasize that the relaxation from this excited state conformational change can still help us to understand how a local conformational change (in the retinal) can lead to larger-scale conformational change (and ultimately to G-protein activation).

It is interesting to speculate, however, on the types of *cis-trans* isomerizations that are not allowed – either quantum mechanically or from the molecular dynamics viewpoint. In either formulation, the type of large *cis-trans* motion that would occur with a rigid retinal moiety would create very large and unfavorable van der Waals clashes within the protein cavity. Thus, it makes biophysical sense that the type of motion adopted by the retinal chain would be as minimal as possible to allow the sensitivity of the response to be magnified and coupled to larger-scale motion. Phrased in another way, the quantum chemical findings suggest that the rapid-relaxation of the retinal molecule is due to the effects of the protein cavity environment in pre-priming the conformation of the retinal such that the efficiency of *cis-trans* change is very high. Furthermore, once the *trans* state has been reached, there is a very high chance of the energy stored in the initially local conformational change of the retinal being carried over to larger changes in the coupling of G-proteins. Thus, we argue that evolution has created an extremely sensitive molecular instrument for converting the energy of photon activation to protein conformational change and signal activation. That the motional changes in the retinal are small, is significant, because it creates a system where the photon energy can be readily and efficiently absorbed and then the isomerized reaction readily focused on conformational change. We emphasize that the dark-adapted state of rhodopsin was the starting point for the calculations. If structures of the intermediates are determined, it would be possible to perform calculations exploring the whole range of conformational change that is initiated by the *cis-trans* photoconversion. In other words, the prediction of intermediate conformations is not nearly as reliable as the experimental determination of those intermediates. From the experimentally-determined intermediates, simulations would be possible that create a thermodynamic picture of the molecular changes that occur during the photocycle. Without x-ray structures for the intermediates, the nature of the changes are inferred from the current calculations and those changes should be seen as suggestive, rather than definitive predictions.

Several other research groups have also performed molecular dynamics calculations of rhodopsin.[42; 43; 44; 45; 46; 47] Each of these studies has something of value to add to our understanding of the molecular nature of rhodopsin and its activation by light. Each group has made different choices with respect to membrane representation, length of simulation, size of simulation, and treatment of electrostatics. It is too early to tell whether a subset of these models is more right or more wrong than others, due to the problem of sampling that all of the computational groups face. It is possible to read each paper as a possible molecular story for

how the system might behave. We suggest that the full set of these papers is very valuable for bringing new insights into the nature of the molecular behavior, in part because of the different characterization of the system (i.e. pK states, environment conditions, etc.).

Nature of cis-trans potential function: There has been considerable discussion in the literature about the nature of the changes in the retinal during photoactivation. These changes are likely to be minimal in overall conformation, but in the absence of a light-adapted structure, it is difficult to be certain about the scale of the overall change with light activation. In our simulations, the change was forced within 200 fs. We emphasize that the potential energy surface used during the isomerization, as well as for the relaxation, is the ground-state surface defined by the CHARMM potential function. That is, there is no attempt made to define a light-activated surface for the isomerization reaction and then to determine the longer-time relaxation process. Papers that examine the quantum dynamics on a short-time scale and connect to the FTIR spectra should be examined for this type of analysis.[36]

Time-scale for isomerization: Experimental work[48] with femtosecond stimulated Raman spectroscopy has shown that there are significant changes in the retinal spectra during the transition from the dark-adapted state to the stable bathorhodopsin intermediate.[49] The experimental work suggests that the full transition from dark-adapted to bathorhodopsin state is about 1 ps in length and that the efficiency of the process, as well as the capture of the energy for large-scale conformational change is very high. In particular, the results suggest changes in C=C, C-C and C-H regions of the spectra that relax very quickly while also showing environmental effects suggestive of a change in interaction between retinal and the protein environment. The bathorhodopsin state is then reached, the effective starting point for our calculations. The conformation of this state is not fully determined experimentally, but may reflect a series of changes along the retinal structure, supporting a minimal change in the retinal overall shape. This is consistent with detailed quantum chemical calculations as well as findings from lipid bilayer simulations that suggest that correlated dihedral changes can be made that prevent very large conformational changes. As noted above, while we share an interest in determining the details of this bathorhodopsin starting stage, we elected not to explore the quantum details of the excited surface, nor to optimize all possible starting points. Instead, our starting point (the bathorhodopsin state) is consistent with experimental information and is thus plausible as an energy storage state intermediate to the signaling cascade and thus to M-II.

Initial relaxation events: In our simulations there is a time-delay from the initial rapid isomerization to the larger-scale conformational transitions and the relaxation of the retinal conformation. This makes sense, since the excited state surface for the initial conformational change will act as a trigger to create the larger-scale conformational changes that are seen experimentally in later stages of the photo-cycle. In order for those longer-term events to occur, there must be a period of relaxation after the isomerization during which the retinal conformational change is adjusted and adapted to by the protein compartment. In the initial stages of relaxation (roughly 30 ns in our simulation), the protein adjusts to the retinal. This can be seen in the energy changes that occur within the retinal: initially 7 kcal/mol higher in energy than the dark-adapted self-energy and then a relaxation, after about 30 ns to a lower self energy state. At the same time, within that 30 ns window, the coupling of the retinal to other sidechains near to the dark-adapted cavity is shifting and the beta-ionone ring is moving from being near one location to being near another. While the nature of these changes is subtle, the overall effect

is very large. This magnification is due to the hydrogen bonding nature of the protein, resembling a large allosteric network, with the changes in one location being communicated energetically to other locations. The net effect is that within 30 ns, in this particular simulation, the initial relaxation events from the *cis-trans* isomerization have already been communicated to many other parts of the protein.

Changes by end of 30 ns: We find it fascinating that large scale changes occur from the small scale changes initiated by the *cis-trans* isomerization of retinal. The nature of these changes, subject to our sampling limitations imposed by molecular dynamics calculations, suggest a tight coupling within the rhodopsin:retinal system.[50; 51] Thus, the large tilt changes in Helix 6 are well underway by 30 ns, the changes in the G-protein activation region (Helices 5 and 6 and cytoplasmic loop C3) have also started. But, the largest indicator of the nature of these shifts is the changes in sidechain-to-retinal coupling. The almost complete decoupling of the retinal interaction energy between retinal and Phe 261, Trp 265 and Tyr 268 is already indicative of the types of changes that occur later in the cycle.

Shift of beta-ionone ring and Ala 169: Experimental work has suggested that at least part of the photocycle involves a shift in Helix 4 with Ala 169 and its interaction with retinal.[30] The detailed molecular nature of the distance shift is not resolved in the experimental work, but it is suggested in the molecular dynamics simulations. In particular, the shift is initiated by the decoupling of the retinal from the dark-adapted state sidechain interactions. This loosening of dark-state interactions enables a relatively modest shift in conformational space of the retinal and a much larger shift in the packing of the protein around the *trans* retinal ring system. The movement of Ala 169 to increase interaction with the retinal ring and the shift in Helix 4 is thus consistent with a shift in the overall energetics of the retinal/rhodopsin system following the relaxation from the *trans* state.

Counter-ion Switch: Experimental work has suggested that part of the light activation could involve a switch of the Schiff base counter-ion from the Glu 113 to Glu 181.[11; 32; 52; 53; 54; 55; 56; 57; 58; 59; 60; 61; 62; 63] This would involve a change in protonation states during the photocycle[64] and experimental work has confirmed that Glu 113 is the dark-state counter-ion and been suggestive, but not conclusive, that Glu 181 is active during the light activated stages.[65; 66] The simulation results are intriguing in suggesting that this counterion switch mechanism could be present as part of the relaxation mechanism of the protein to the retinal conformational change. We should emphasize that in our calculations there is no change in the charge state of Glu 181 during or after the forced isomerization.[55; 67] Thus, the driving force for the change is not wholly from charge transfer during the photocycle, but could also be related to the coupled relaxation of the full system from the initial events of light activation. In this regard, quantum calculations about the effect of the counterion switch are also suggestive of the types of changes occurring with a counterion switch that could underlie function.[33; 68]

Longer (30 ns to 150 ns) changes: Activation of G-protein coupled receptors will depend on changes in Helices 5 and 6 along with the C3 loop regions.[31; 69; 70; 71; 72; 73; 74; 75] These locations have been implicated, by mutagenesis studies, as key players in the coupling of the activated rhodopsin to G-protein signaling.[76] For this reason it is intriguing to note that the simulation results suggest large-scale changes in all three of these locations. Helix 5 shows an increased degree of bending between upper and lower parts of the helices, and that drives further changes

in the C3 loop regions. At the same time, Helix 6 shows a decrease in the kink angle and a shift to a more perpendicular orientation of the helix axis. These three sets of coupled changes are all a relaxation response to the forced isomerization change. It is not hard to imagine that the nature of these changes could lead to a shift in binding affinity for a G-protein system and thus to activation.[77] At the same time, we emphasize that the simulations do not include the G-protein binding, since the details of this interaction are still not known experimentally, and that the timescale of activation within this system is clearly beyond the timescale of the current calculations. Thus, the nature of these changes should be seen as suggestive of the types of shifts that are likely to occur with activation of the system.

Implications for rhodopsin function: Rhodopsin has evolved to be highly tuned to converting light energy to conformational change and signal amplification. The simulations that we have performed are suggestive of some of the types of coupling and change that underlie this tight connection between structure and function. This underlies much experimental work with mutants suggesting that disruption of the tight conformation of rhodopsin will lead to a loss of function.[12] An example is the Cys 110-Cys 187 pair.[78] In particular, if the retinal *cis-trans* isomerization were not coupled into a protein cavity setting that adjusted to support the transition, then the signal amplification and coupling would not be so efficient. In other words, the nature of the coupled set of changes induced by light activation is amplified by the design of the retinal cavity to induce a set of protein changes that lead to activation. In this sense the rhodopsin system is very much like the switch mechanism suggested by experimental work.[79; 80; 81] The relatively small change introduced by the retinal conformational change is then amplified through the hydrogen bond and sidechain interaction network into a shift in helix orientations, Schiff base partners, and eventual G-protein coupling.

An emerging frontier is then to understand more about the nature of the coupling between the rhodopsin protein and G-protein. Several groups have started to attempt this model, based on the x-ray structures of G-proteins[16; 17; 82; 83; 84; 85; 86; 87] and we believe that the simulation results here may lead to speculation about the nature of the interfacial coupling between these domains.[88; 89; 90; 91; 92; 93; 94] In particular, we suggest that the shift in the surface properties of the rhodopsin protein on activation will be found to create a much less favorable environment for protein:protein interaction. In this regard, the mutation studies and their effects on the surface of dark-adapted and light-activated (inferred) could be interesting[95; 96; 97] as well as studies of the role of the cytoplasmic loop region.[98] Several groups are working to understand more about protein:protein interactions and this could be a model system for understanding how the modulation of protein:protein interactions lead to shifts in function.

Implications for spectral tuning: While the current molecular dynamics calculations do not address the changes in the binding cavity that underlie spectral response, it is interesting to suggest several implications for spectral tuning from the results. In particular, the results strongly underlie the importance of the protein cavity in forming a part of the solvation environment dictating the type of conformational change that will be determined by the light-activated retinal.[15; 99; 100; 101; 102; 103; 104; 105; 106; 107] In that sense, the nature of these sidechain and main chain interactions drive a preference for a particular type of excited state and a particular type of relaxation response. Experimental work that shifts the nature of retinal further supports this type of thinking.[108; 109; 110; 111; 112] It can also be inferred that spectral tuning will lead to a similar type of excited state conformation for retinal and, following that transition, to a similar

set of relaxation events that lead to G-protein activation.[113] It is thus intriguing to consider the nature of the changes possibly induced by different environments. An example is the set of mutants looking at hydroxyl changes.[114] We would suggest that by shifting the energy required for isomerization and forcing a similar pathway for relaxation from the excited state, the retinal/rhodopsin system has been engineered for maximum flexibility by evolution.[14]

Effect of the membrane environment: While the current simulations did not directly compare different models for the membrane with the transition, it is clear from experimental work that rhodopsin is very sensitive to the details of the lipid setting.[50; 115; 116; 117; 118; 119; 120; 121; 122; 123,] We can speculate, from our current and previous[120] work about why this may be important. We suggest that the nature of the coupling from the *cis*-*trans* activated state to the relaxed state will depend on an ability of the helices to re-orient, the sidechains to shift, and the cytoplasmic loops to adjust their relative positions. The lipid environment then provides a resistance to this motion that can be either supportive of the change, making the free energy change for the large-scale motion relatively easy, or providing a setting that makes the relative shift in conformation much more difficult, allowing a change, but only with much greater resistance and thus with a decrease in signal to noise outcome. We suggest that the effects of DHA on rhodopsin are to make the transitions between helix orientations (e.g. the kink changes in Helices 5 and 6) much easier and thus to support the nature of the large scale changes underlying function in this system.[123; 124; 125; 126]

Implications for other GPCR systems: We believe that study of the rhodopsin system leads to insights important for other GPCR systems as well.[3; 127; 128; 129] This is especially important due to the large number of important pharmacological targets within the large GPCR family.[130; 131; 132; 133; 134; 135; 136; 137; 138; 139; 140; 141] In our simulation results, the nature of the coupling between a local conformational change and a large-scale helix and loop change may be similar across the GPCR family. This would imply that the effect of ligand binding to a GPCR is tightly regulated by a coupled set of energetic interactions, in a similar manner to that found within the retinal-rhodopsin system.[142] The intriguing result is that the nature of the second extracellular loop (containing the beta-bulge region) may thus be critical in determining the nature of the coupling between local conformational binding induced changes and large-scale signaling induced changes.[143; 144; 145; 146] We would suggest that this is similar, in some ways, to the nature of spectral tuning in that adjustments to the binding cavity can support the relative efficiency and recognition of certain ligands relative to others.

Conclusions

The high quantum efficiency of rhodopsin is coupled to an ability to capture local changes in retinal behavior and induce conformational changes that lead to signaling.[1; 147; 148; 149; 150; 151] How this happens remains a mystery that will require more computational and experimental work.[4; 152] The molecular details of how the stable bathorhodopsin intermediate (about 1 ps after light absorption and *cis*-*trans* isomerization) leads to conformational changes in the Meta II state were the target of our calculations in this paper. What we find emphasizes the nature of the coupling throughout the protein that we had found in our previous simulation result, which is consistent with the general behavior of other protein systems. In particular, the *cis*-*trans* isomerization leads to a change in the sidechain interactions with retinal, and eventually to large-

scale changes in tilt and conformation of the system. While the details of these changes may not be fully sampled in the current simulations, the results that we report are suggestive of the types of compensation that may occur in the system upon light activation.

The preliminary analysis of the long simulation of rhodopsin after isomerization shows that significant structural changes occur in the 150 ns timeframe. The isomerization results in some of the 7 transmembrane helices undergoing tilt and kink angle changes that are well beyond their fluctuation range in the equilibrium dark state. One of the key structural changes is the motion of the ionone ring of retinal. The ring progressively gets closer to Ala 169. This is consistent with crosslink experiments[30] where a crosslink forms between the ring and Ala 169.
A strong energetic transition is observed to occur at t = 70 ns after isomerization. The majority of the energy of this transition involves breaking of the salt bridge between Glu 113 and the protonated Schiff base. Raman spectroscopy[49] has shown that the protonated Schiff base stretching modes are indicative of very different Schiff base environments. This implies that the transition from bathorhodopsin to the LUMI state involves chromophore relaxation and dramatic changes in the Schiff base region. Our results are in agreement with these experimental observations. Furthermore, besides the structural dynamics, we show the connection to the energetics that drives the changes.

Table I

Retinal bond orientations (θ/deg).

system	C_5	C_9	C_{13}
ground state ^{2}H NMR[a]	70 ± 3	128 ± 3	68 ± 2
ground state simulation[b,d]	75 ± 11	129 ± 11	59 ± 14
post-isomerization simulation[c,d]	137 ± 15	80 ± 35	67 ± 15

[a] NMR data from Salgado et al.[29] Note that the bond orientation θ was not distinguished from its supplement π - θ in the NMR measurements. Errors correspond to inverse curvature matrix of the χ^2 hypersurface utilized for the nonlinear regression fits. [b]40-ns dark-adapted state simulation data from our previous work in Ref. 20. [c]150-ns post-photoisomerization simulation data from this work. [d]Errors correspond to the standard deviation from the average values given.

Figures

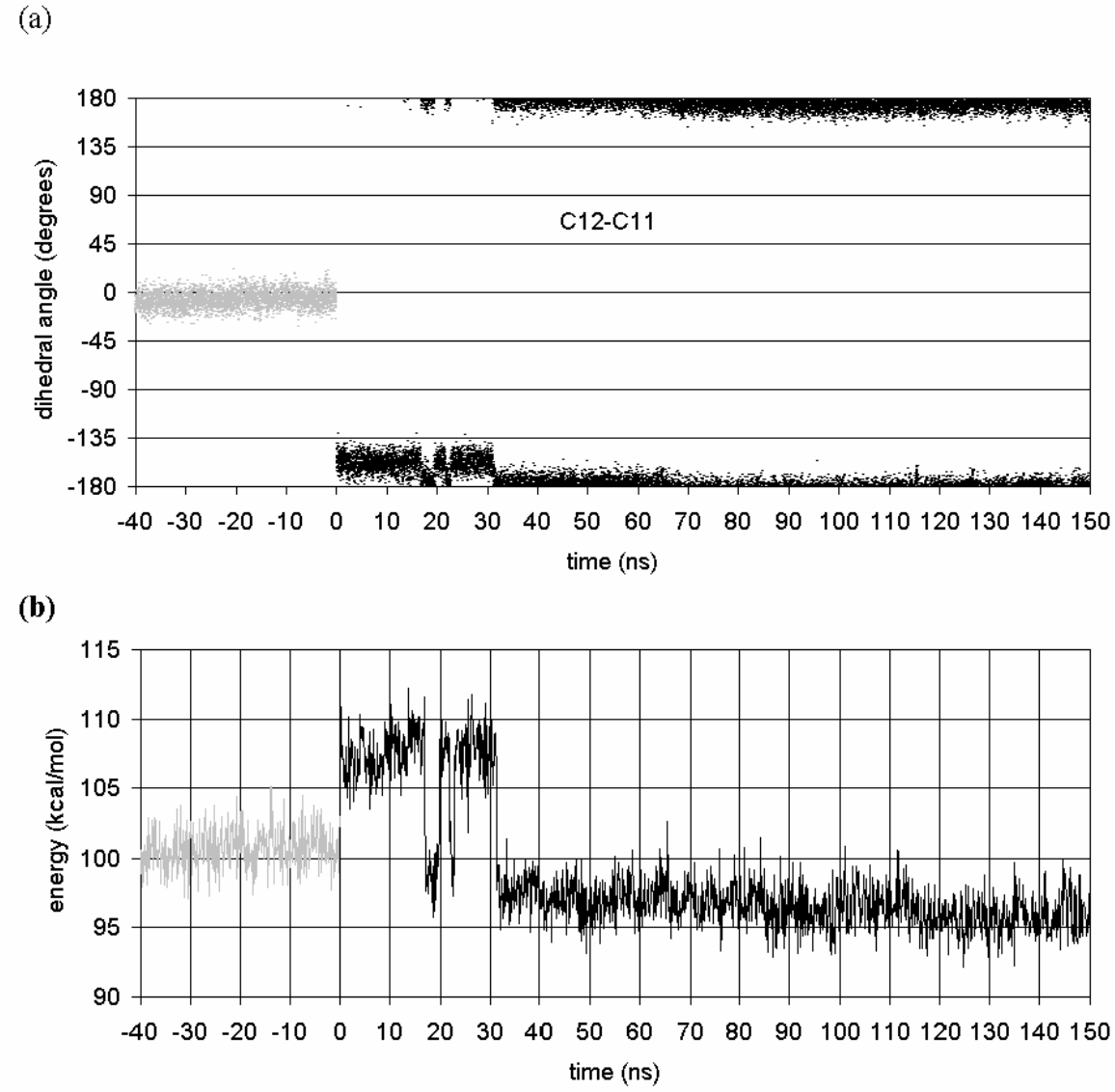

Figure 1. (a) Retinal dihedral angles as a function of time. **(b)** Retinal's self energy as a function of time. The gray line represents the dynamics for the 40 ns run of the dark adapted rhodopsin, *i.e.* retinal in the *cis* state and the C11-C12 dihedral has angle 0. The black line represent the dynamics after the *cis-trans* isomerization, *i.e.* retinal is in the *trans* state with the angle at 180 (=-180).

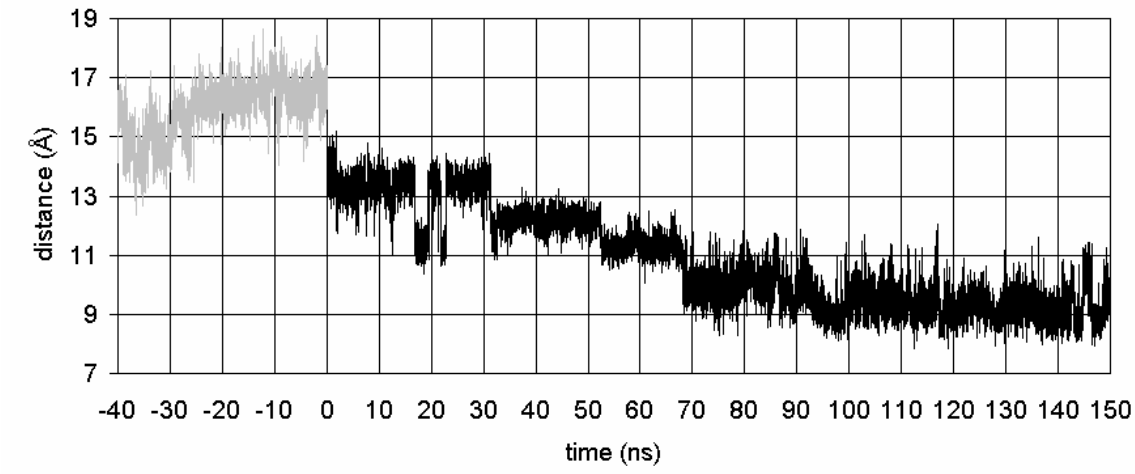

Figure 2. The distance between retinal's ionone ring and Ala 169. The gray line represents the dynamics for the 40 ns run of the dark adapted rhodopsin, *i.e.* retinal in the *cis* state. The black line represents the dynamics after the *cis-trans* isomerization, *i.e.* retinal is in the *trans* state.

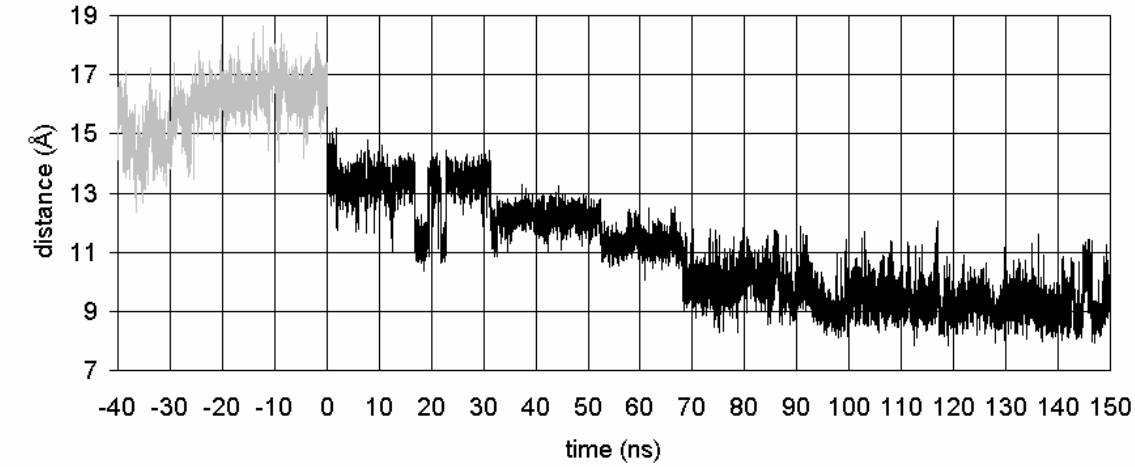

Figure 3. Tilt angles as a function of time for Helices 5b (Phe 203 – His 211), 5c (His 211 – Leu 226), 6b (Pro 267 – Thr 277) and 7a (Ile 286 – Pro 291).

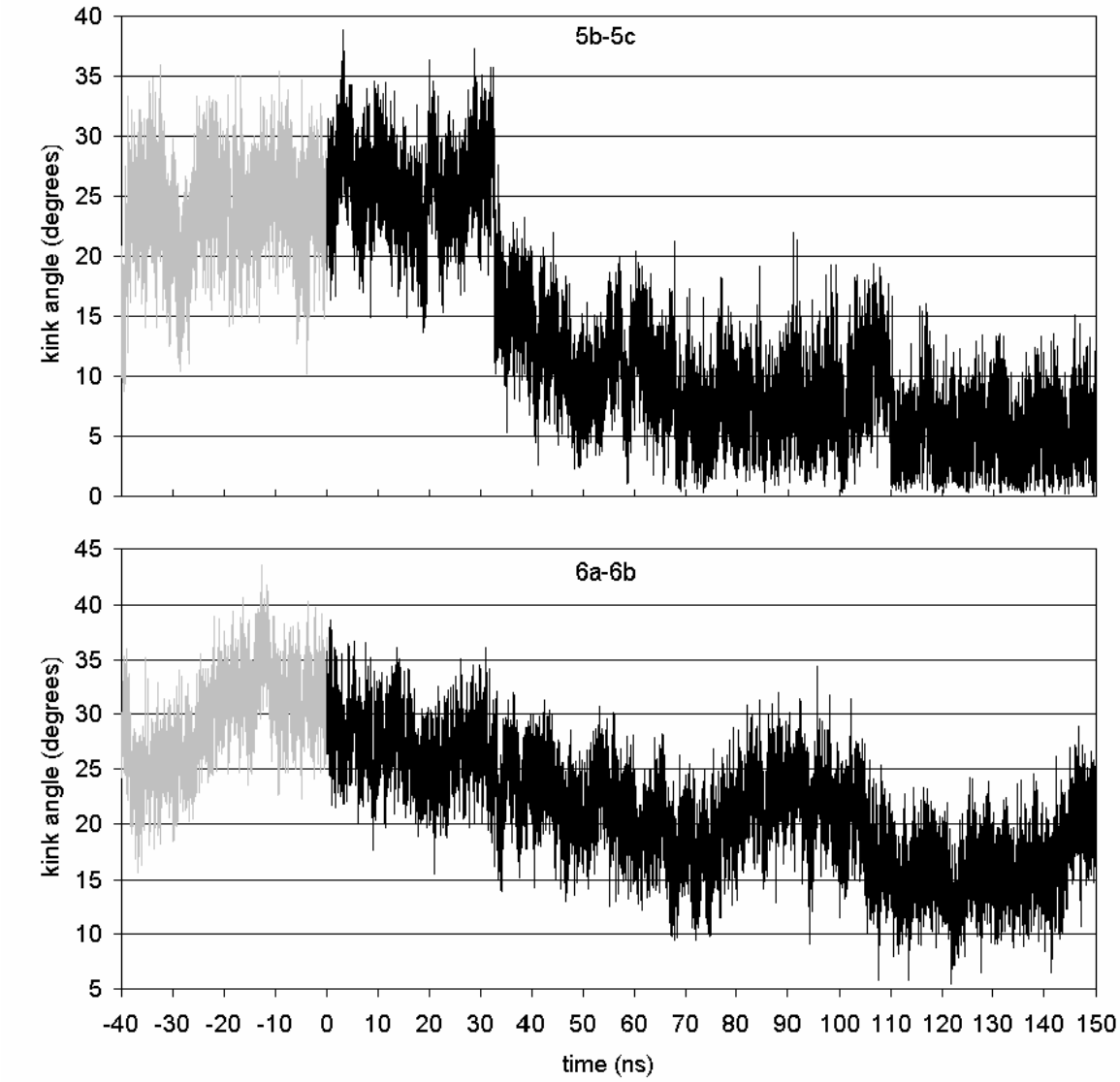

Figure 4. Kink angles as a function of time for Helices 5b-5c (kink at His 211) and 6a-6b (kink at Pro 267).

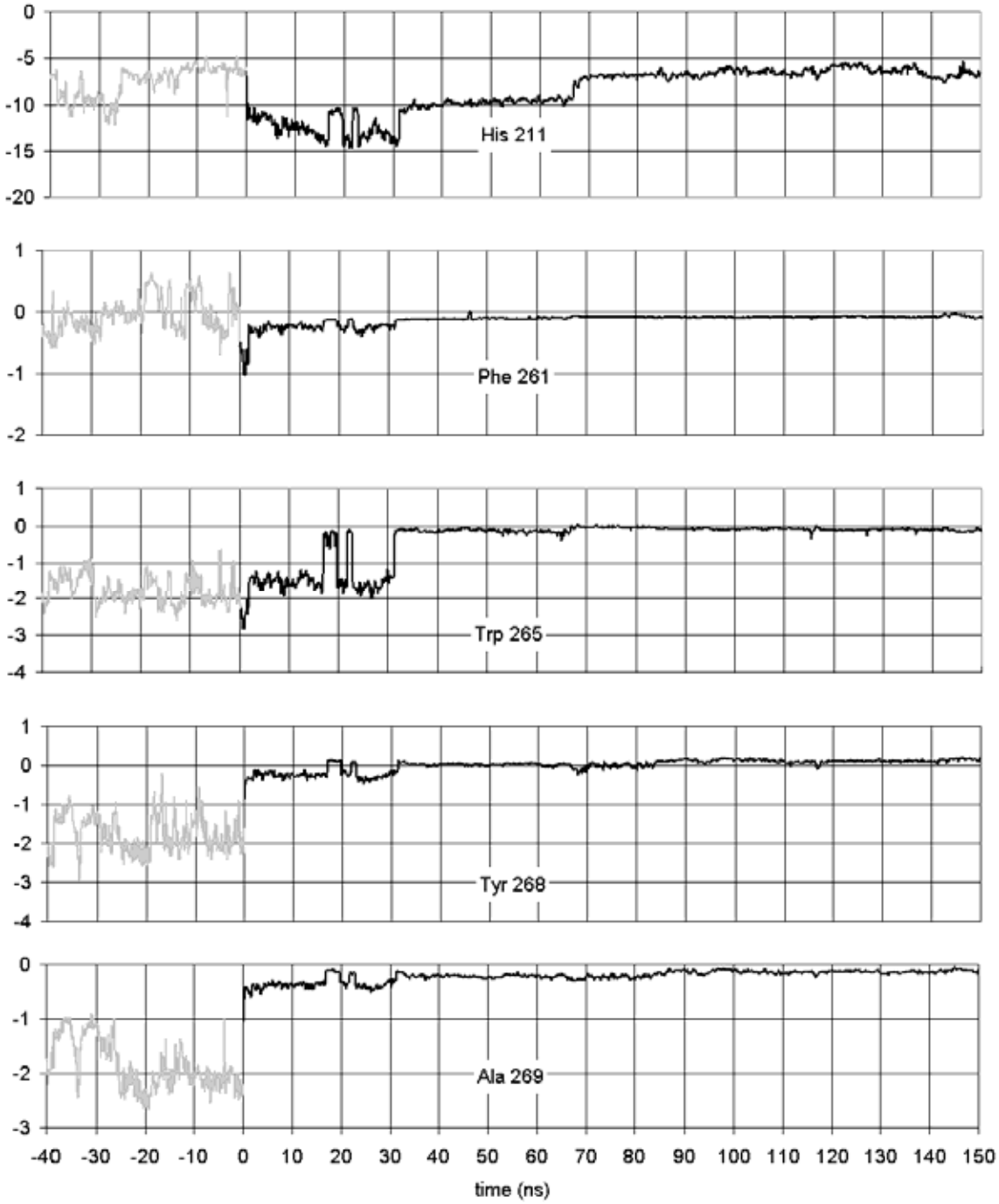

Figure 5. Interaction energies between retinal's beta ionone ring and nearby side chains.

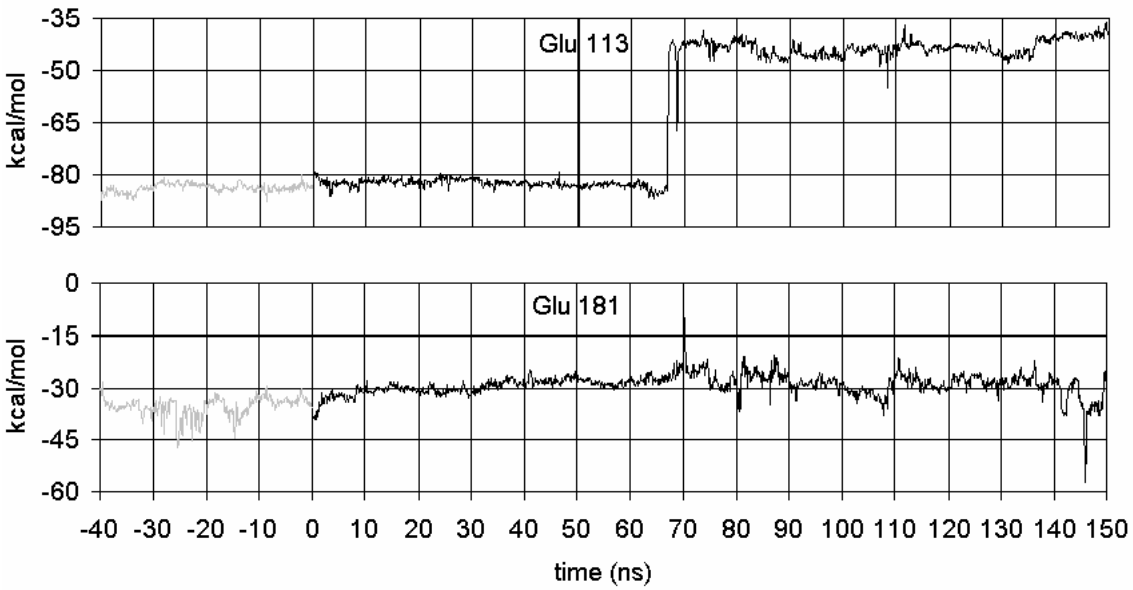

Figure 6. Interaction energies in kcal/mol between the retinal PSB (including Lys 296) and nearby PSB counterion candidates Glu 113 and Glu 181. Glu 113 acts as the counterion in the dark adapted state and through the first 65 ns after forced photoisomerization. Near t = 70 ns, the salt bridge between Glu 113 and retinal breaks. Later, near t = 146 ns, Glu 181 briefly becomes the PSB counterion, hinting at possible subsequent completion of the transition from Glu 113 to Glu 181 acting as the PSB counterion. The gray lines represent the dynamics for the 40 ns run of the dark adapted rhodopsin, *i.e.* retinal in the *cis* state. The black lines represent the dynamics after the *cis*-*trans* isomerization, *i.e.* retinal is in the *trans* state.

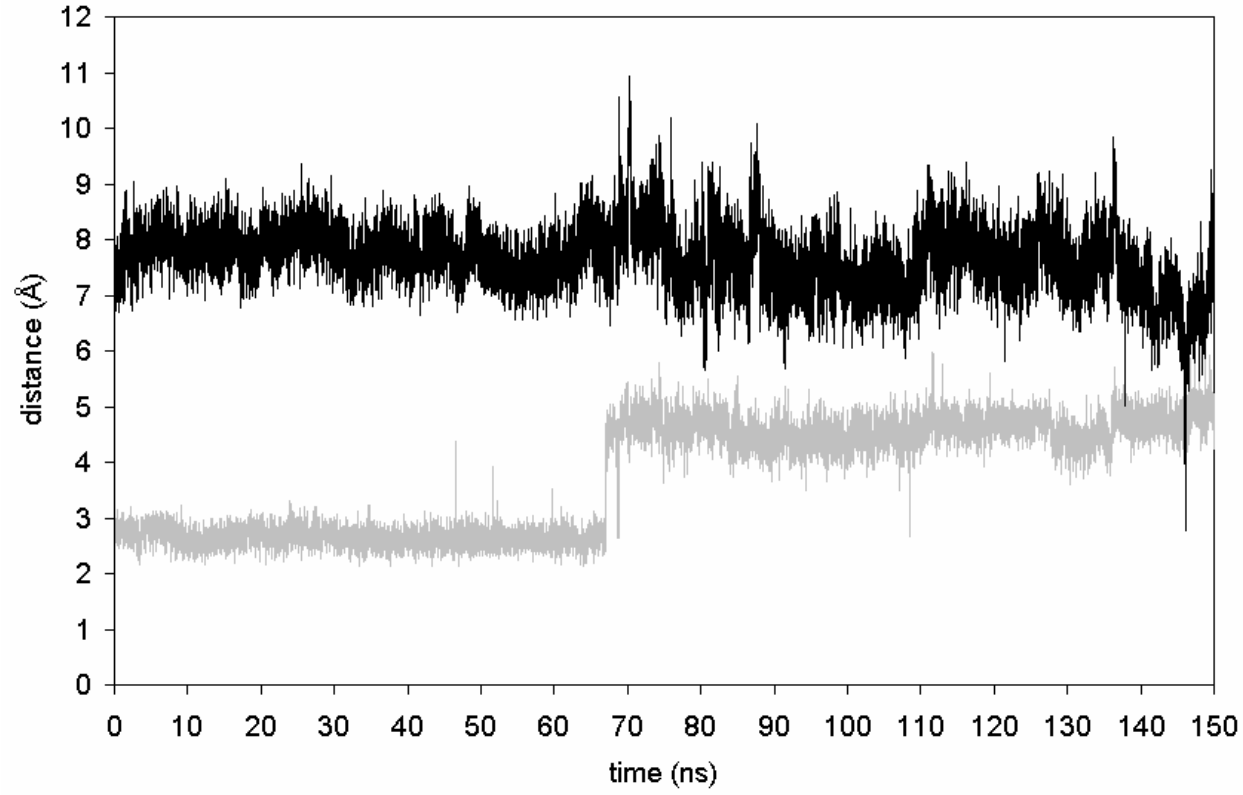

Figure 7. Distance between retinal's H16 atom (the proton on the PSB), and the C_δ atom of potential counterions Glu 113 (gray) and Glu 181 (black). The minimum value of 2.8 Å in the Glu 181 time series corresponds with the maximum value of 6.3 Å in the Glu 113 time series (both at t = 146 ns).

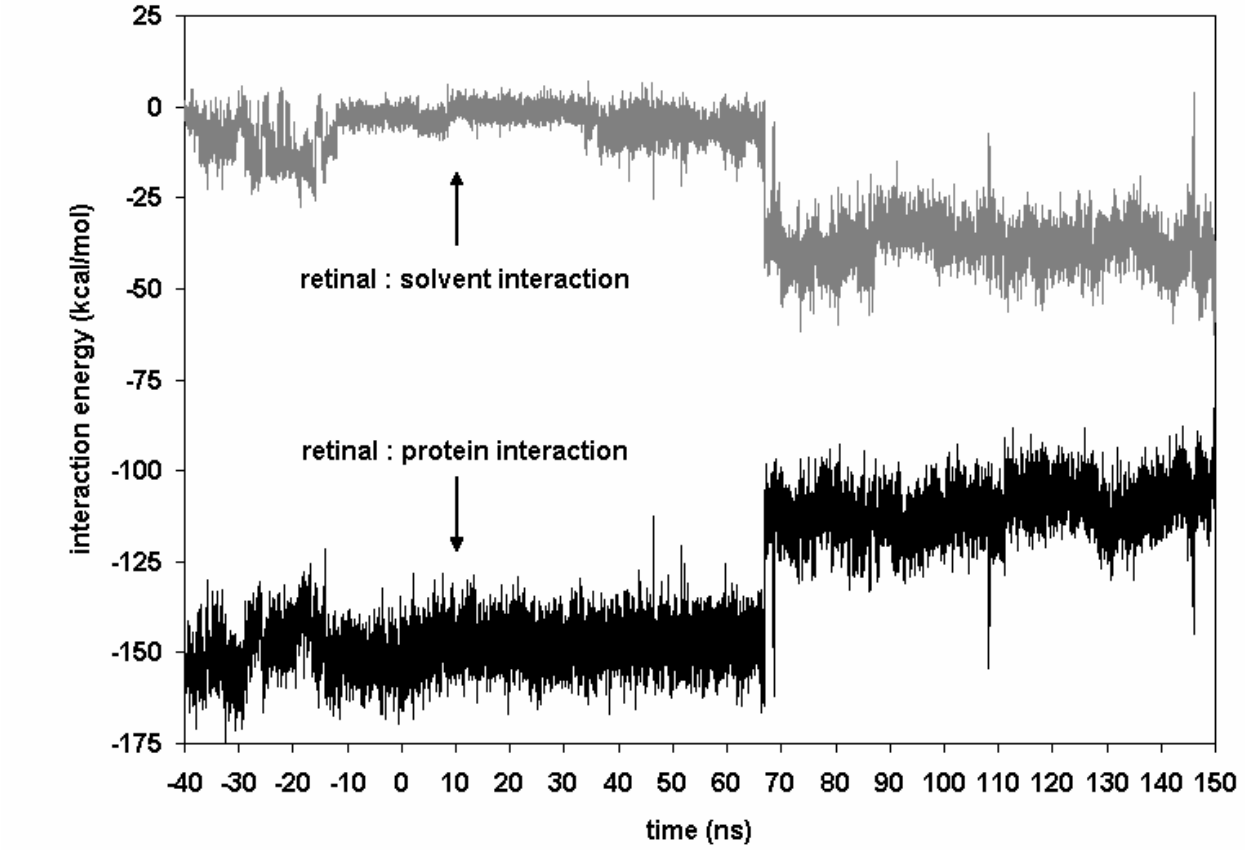

Figure 8. Interaction energy of retinal with the surrounding aqueous solvent, and retinal with the rest of the rhodopsin molecule. Retinal:lipid interaction was negligible. The large shift near 70 ns is due to the breaking of the salt bridge between Glu 113 and the PSB. A water molecule partially compensates for the broken bridge as is evidenced by the lower retinal:solvent interaction energy.

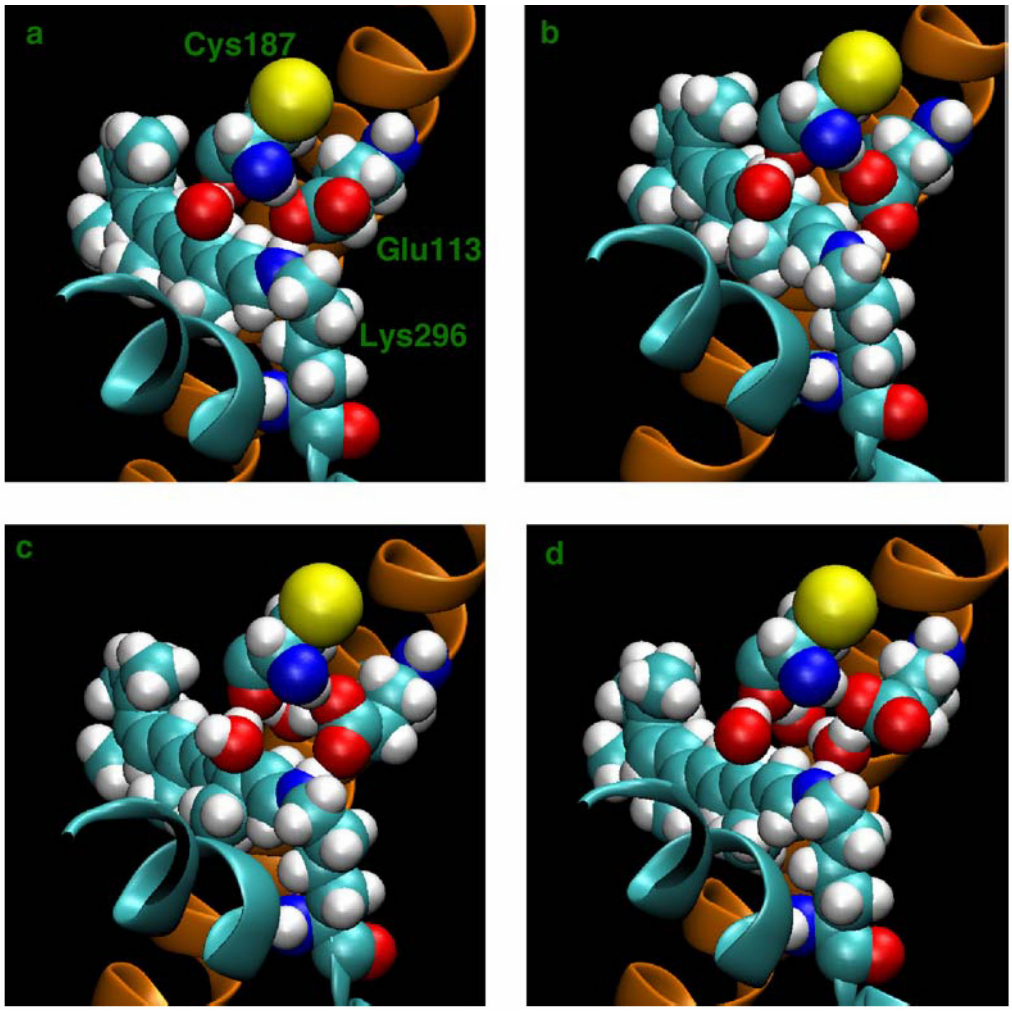

Figure 9. Images[153] of the rhodopsin protein, Lys 296 (retinal) with Glu 113 and Cys 187, and nearby waters at different times a) t = 64 ns, b) t = 65 ns, c) t = 66 ns and d) t = 67 ns. Helices 3 (orange) and 7 (cyan) are also visible in a cartoon representation. Lys 296 is on Helix 7 and Glu 113 is on Helix 3.

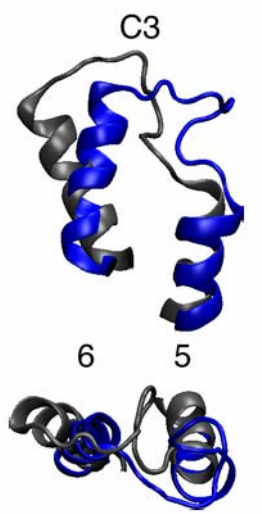

Figure 10. Rhodopsin's C3 loop in dark state (gray) and after 150 ns of simulation (blue).[153] Bottom image is 90° rotation of top image. The label 5 and 6 identify the respective transmembrane helices.

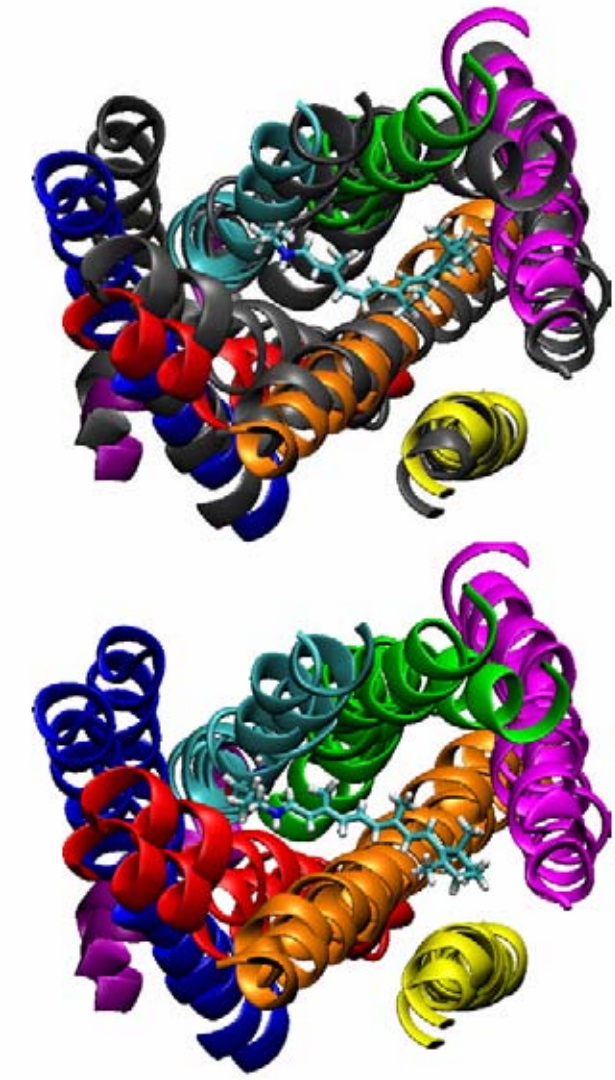

Figure 11. Comparison of helix positions in dark state (gray in top image) and after 150 ns after isomerization (color in top image).[153] Top image shows retinal in *cis* state. The bottom image shows retinal in the *trans* state with the dark and 150 ns state both colored. The helix number and colors are 1 (blue), 2 (red), 3 (orange), 4 (yellow), 5 (magenta), 6 (cyan), 7 (green), and 8 (purple).

References

[1] Sakmar, T. P., Menon, S. T., Marin, E. P. & Awad, E. S. (2002). Rhodopsin: Insights from recent structural studies. *Annual Review of Biophysics and Biomolecular Structure* **31**, 443-484.

[2] Shi, L. & Javitch, J. A. (2002). The binding site of aminergic G protein-coupled receptors: The transmembrane segments and second extracellular loop. *Annual Review of Pharmacology and Toxicology* **42**, 437-467.

[3] Filipek, S., Stenkamp, R. E., Teller, D. C. & Palczewski, K. (2003). G protein-coupled receptor rhodopsin: A prospectus. *Annual Review of Physiology* **65**, 851-879.

[4] Hubbell, W. L., Altenbach, C., Hubbell, C. M. & Khorana, H. G. (2003). Rhodopsin structure, dynamics, and activation: A perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking. *Advances in Protein Chemistry*, **63**, 243-290.

[5] Burns, M. E. & Baylor, D. A. (2001). Activation, deactivation, and adaptation in vertebrate, photoreceptor cells. *Annual Review of Neuroscience* **24**, 779-805.

[6] Lamb, T. D. & Pugh, E. N. (2004). Dark adaptation and the retinoid cycle of vision. *Progress in Retinal and Eye Research* **23**, 307-380.

[7] Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-745.

[8] Sakmar, T. P. (1998). Rhodopsin: A prototypical G protein-coupled receptor. In *Progress in Nucleic Acid Research and Molecular Biology, Vol 59*, Vol. 59, pp. 1-34.

[9] Ferretti, L., Karnik, S. S., Khorana, H. G., Nassal, M. & Oprian, D. D. (1986). Total Synthesis of a Gene for Bovine Rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 599-603.

[10] Fung, B. K. K. & Stryer, L. (1980). Photolyzed Rhodopsin Catalyzes the Exchange of Gtp for Bound Gdp in Retinal Rod Outer Segments. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **77**, 2500-2504.

[11] Sakmar, T. P., Franke, R. R. & Khorana, H. G. (1989). Glutamic Acid-113 Serves as the Retinylidene Schiff-Base Counterion in Bovine Rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 8309-8313.

[12] Franke, R. R., Konig, B., Sakmar, T. P., Khorana, H. G. & Hofmann, K. P. (1990). Rhodopsin Mutants That Bind but Fail to Activate Transducin. *Science* **250**, 123-125.

[13] Fahmy, K., Sakmar, T. P. & Siebert, F. (2000). Structural determinants of active state conformation of rhodopsin: Molecular biophysics approaches. *Methods in enzymology* **315**, 178-196.

[14] Yokoyama, S. (2002). Molecular evolution of color vision in vertebrates. *Gene* **300**, 69-78.

[15] Kuwayama, S., Imai, H., Hirano, T., Terakita, A. & Shichida, Y. (2002). Conserved proline residue at position 189 in cone visual pigments as a determinant of molecular properties different from rhodopsins. *Biochemistry* **41**, 15245-15252.

[16] Noel, J. P., Hamm, H. E. & Sigler, P. B. (1993). The 2.2-Angstrom Crystal-Structure of Transducin-Alpha Complexed with Gtp-Gamma-S. *Nature* **366**, 654-663.

[17] Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. (1994). Structural Determinants for Activation of the Alpha-Subunit of a Heterotrimeric G-Protein. *Nature* **369**, 621-628.

[18] Mathies, R., Freedman, T. B. & Stryer, L. (1977). Resonance Raman Studies of Conformation of Retinal in Rhodopsin and Isorhodopsin. *Journal of Molecular Biology* **109**, 367-372.

[19] Hayward, G., Carlsen, W., Siegman, A. & Stryer, L. (1981). Retinal Chromophore of Rhodopsin Photoisomerizes within Picoseconds. *Science* **211**, 942-944.

[20] Crozier, P. S., Stevens, M. J., Forrest, L. R. & Woolf, T. B. (2003). Molecular dynamics simulation of dark-adapted rhodopsin in an explicit membrane bilayer: Coupling between local retinal and larger scale conformational change. *Journal of Molecular Biology* **333**, 493-514.

[21] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **102**, 3586-3616.

[22] Schlenkrich, M., Brickmann, J., MacKerell, A. D. Jr. and Karplus, M. . (1996). An empirical potential energy function for phospholipids: criteria for parameter optimization and applications. In *Biological membranes: A molecular perspective from computation and experiment* (Merz, K. M. J., and Roux, B., ed.), pp. 31-81. Birkhauser Press, New York, NY.

[23] Nina, M., Smith, J. C. & Roux, B. (1993). Ab-Initio Quantum-Chemical Analysis of Schiff-Base Water Interactions in Bacteriorhodopsin. *Theochem-Journal of Molecular Structure* **105**, 231-245.

[24] Edwards, P. C., Li, J., Burghammer, M., McDowell, J. H., Villa, C., Hargrave, P. A. & Schertler, G. F. X. (2004). Crystals of native and modified bovine rhodopsins and their heavy atom derivatives. *Journal of Molecular Biology* **343**, 1439-1450.

[25] Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. X. (2004). Structure of bovine rhodopsin in a trigonal crystal form. *Journal of Molecular Biology* **343**, 1409-1438.

[26] Okada, T. (2004). X-ray crystallographic studies for ligand-protein interaction changes in rhodopsin. *Biochemical Society Transactions* **32**, 738-741.

[27] Okada, T., Sugihara, M., Bondar, A. N., Elstner, M., Entel, P. & Buss, V. (2004). The retinal conformation and its environment in rhodopsin in light of a new 2.2 angstrom crystal structure. *Journal of Molecular Biology* **342**, 571-583.

[28] Plimpton, S. J. (1995). Fast Parallel Algorithms for Short-Range Molecular Dynamics, *J Comp Phys* **117**, 1-19.

[29] Salgado, G. F. J., Struts, A. V., Tanaka, K., Fujioka, N., Nakanishi, K. & Brown, M. F. (2004). Deuterium NMR structure of retinal in the ground state of rhodopsin. *Biochemistry* **43**, 12819-12828.

[30] Borhan, B., Souto, M. L., Imai, H., Shichida, Y. & Nakanishi, K. (2000). Movement of retinal along the visual transduction path. *Science* **288**, 2209-2212.

[31] Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L. & Khorana, H. G. (1996). Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. *Science* **274**, 768-770.

[32] Yan, E. C. Y., Kazmi, M. A., Ganim, Z., Hou, J. M., Pan, D. H., Chang, B. S. W., Sakmar, T. P. & Mathies, R. A. (2003). Retinal counterion switch in the photoactivation of the G protein-coupled receptor rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9262-9267.

[33] Cembran, A., Bernardi, F., Olivucci, M. & Garavelli, M. (2004). Counterion controlled photoisomerization of retinal chromophore models: a computational investigation. *Journal of the American Chemical Society* **126**, 16018-16037.

[34] Blomgren, F. & Larsson, S. (2002). Using 1,3-butadiene and 1,3,5-hexatriene to model the cis-trans isomerization of retinal, the chromophore in the visual pigment rhodopsin. *International Journal of Quantum Chemistry* **90**, 1536-1546.

[35] Cembran, A., Bernardi, F., Olivucci, M. & Garavelli, M. (2005). The retinal chromophore/chloride ion pair: Structure of the photo isomerization path and interplay of charge transfer and covalent states. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6255-6260.

[36] Buss, V., Weingart, O. & Sugihara, M. (2000). Fast photoisomerization of a rhodopsin model - An ab initio molecular dynamics study. *Angewandte Chemie-International Edition* **39**, 2784-2786.

[37] Blomgren, F. & Larsson, S. (2003). Initial step of the photoprocess leading to vision only requires minimal atom displacements in the retinal molecule. *Chemical Physics Letters* **376**, 704-709.

[38] Blomgren, F. & Larsson, S. (2005). Primary photoprocess in vision: Minimal motion to reach the photo- and bathorhodopsin intermediates. *Journal of Physical Chemistry B* **109**, 9104-9110.

[39] Fotiadis, D., Liang, Y., Filipek, S., Saperstein, D. A., Engel, A. & Palczewski, K. (2003). Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature* **421**, 127-128.

[40] Javitch, J. A. (2004). The ants go marching two by two: Oligomeric structure of G-protein-coupled receptors. *Molecular Pharmacology* **66**, 1077-1082.

[41] Jensen, A. A. & Spalding, T. A. (2004). Allosteric modulation of G-protein coupled receptors. *European Journal of Pharmaceutical Sciences* **21**, 407-420.

[42] Huber, T., Botelho, A. V., Beyer, K. & Brown, M. F. (2004). Membrane model for the G-protein-coupled receptor rhodopsin: Hydrophobic interface and dynamical structure. *Biophysical Journal* **86**, 2078-2100.

[43] Rohrig, U. F., Guidoni, L., Laio, A., Frank, I. & Rothlisberger, U. (2004). A molecular spring for vision. *Journal of the American Chemical Society* **126**, 15328-15329.

[44] Lemaitre, V., Yeagle, P. & Watts, A. (2005). Molecular dynamics simulations of retinal in rhodopsin: From the dark-adapted state towards lumirhodopsin. *Biochemistry* **44**, 12667-12680.

[45] Pitman, M. C., Grossfield, A., Suits, F. & Feller, S. E. (2005). Role of cholesterol and polyunsaturated chains in lipid-protein interactions: Molecular dynamics simulation of rhodopsin in a realistic membrane environment. *Journal of the American Chemical Society* **127**, 4576-4577.

[46] Rohrig, U. F., Guidoni, L. & Rothlisberger, U. (2005). Solvent and protein effects on the structure and dynamics of the rhodopsin chromophore. *Chemphyschem* **6**, 1836-1847.

[47] Schlegel, B., Sippl, W. & Holtje, H. D. (2005). Molecular dynamics simulations of bovine rhodopsin: influence of protonation states and different membrane-mimicking environments. *Journal of Molecular Modeling* **12**, 49-64.

[48] Kukura, P., McCamant, D. W., Yoon, S., Wandschneider, D. B. & Mathies, R. A. (2005). Structural observation of the primary isomerization in vision with femtosecond-stimulated Raman. *Science* **310**, 1006-1009.

[49] Pan, D. H. & Mathies, R. A. (2001). Chromophore structure in lumirhodopsin and metarhodopsin I by time-resolved resonance Raman microchip spectroscopy. *Biochemistry* **40**, 7929-7936.

[50] Brown, M. F. (1994). Modulation of rhodopsin function by properties of the membrane bilayer. *Chemistry and Physics of Lipids* **73**, 159-180.

[51] Janz, J. M. & Farrens, D. L. (2004). Role of the retinal hydrogen bond network in rhodopsin Schiff base stability and hydrolysis. *Journal of Biological Chemistry* **279**, 55886-55894.

[52] Zvyaga, T. A., Min, K. C., Beck, M. & Sakmar, T. P. (1993). Movement of the Retinylidene Schiff-Base Counterion in Rhodopsin by One Helix Turn Reverses the Ph-Dependence of the Metarhodopsin-I to Metarhodopsin-Ii Transition. *Journal of Biological Chemistry* **268**, 4661-4667.

[53] Hufen, J., Sugihara, M. & Buss, V. (2004). How the counterion affects ground- and excited-state properties of the rhodopsin chromophore. *Journal of Physical Chemistry B* **108**, 20419-20426.

[54] Sugihara, M., Buss, V., Entel, P. & Hafner, J. (2004). The nature of the complex counterion of the chromophore in rhodopsin. *Journal of Physical Chemistry B* **108**, 3673-3680.

[55] Kusnetzow, A. K., Dukkipati, A., Babu, K. R., Ramos, L., Knox, B. E. & Birge, R. R. (2004). Vertebrate ultraviolet visual pigments: Protonation of the retinylidene Schiff base and a counterion switch during photoactivation. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 941-946.

[56] Birge, R. R. & Knox, B. E. (2003). Perspectives on the counterion switch-induced photoactivation of the G protein-coupled receptor rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9105-9107.

[57] Jager, S., Lewis, J. W., Zvyaga, T. A., Szundi, I., Sakmar, T. P. & Kliger, D. S. (1997). Time-resolved spectroscopy of the early photolysis intermediates of rhodopsin Schiff base counterion mutants. *Biochemistry* **36**, 1999-2009.

[58] Sakmar, T. P., Franke, R. R. & Khorana, H. G. (1991). The Role of the Retinylidene Schiff-Base Counterion in Rhodopsin in Determining Wavelength Absorbency and Schiff-Base Pka. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 3079-3083.

[59] Fahmy, K., Jager, F., Beck, M., Zvyaga, T. A., Sakmar, T. P. & Siebert, F. (1993). Protonation States of Membrane-Embedded Carboxylic-Acid Groups in Rhodopsin and Metarhodopsin-Ii - a Fourier-Transform Infrared-Spectroscopy Study of Site-Directed Mutants. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 10206-10210.

[60] Arnis, S., Fahmy, K., Hofmann, K. P. & Sakmar, T. P. (1994). A Conserved Carboxylic-Acid Group Mediates Light-Dependent Proton Uptake and Signaling by Rhodopsin. *Journal of Biological Chemistry* **269**, 23879-23881.

[61] Fahmy, K., Siebert, F. & Sakmar, T. P. (1994). A Mutant Rhodopsin Photoproduct with a Protonated Schiff-Base Displays an Active-State Conformation - a Fourier-Transform Infrared-Spectroscopy Study. *Biochemistry* **33**, 13700-13705.

[62] Jager, F., Fahmy, K., Sakmar, T. P. & Siebert, F. (1994). Identification of Glutamic-Acid-113 as the Schiff-Base Proton Acceptor in the Metarhodopsin-Ii Photointermediate of Rhodopsin. *Biochemistry* **33**, 10878-10882.

[63] Kim, J. M., Altenbach, C., Kono, M., Oprian, D. D., Hubbell, W. L. & Khorana, H. G. (2004). Structural origins of constitutive activation in rhodopsin: Role of the K296/E113 salt bridge. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12508-12513.

[64] Ritter, E., Zimmermann, K., Heck, M., Hofmann, K. P. & Bartl, F. J. (2004). Transition of rhodopsin into the active metarhodopsin II state opens a new light-induced pathway linked to Schiff base isomerization. *Journal of Biological Chemistry* **279**, 48102-48111.

[65] Yan, E. C. Y., Kazmi, M. A., De, S., Chang, B. S. W., Seibert, C., Marin, E. P., Mathies, R. A. & Sakmar, T. P. (2002). Function of extracellular loop 2 in rhodopsin: Glutamic acid 181 modulates stability and absorption wavelength of metarhodopsin II. *Biochemistry* **41**, 3620-3627.

[66] Ludeke, S., Beck, R., Yan, E. C. Y., Sakmar, T. P., Siebert, F. & Vogel, R. (2005). The role of Glu181 in the photoactivation of rhodopsin. *Journal of Molecular Biology* **353**, 345-356.

[67] Lewis, J. W., Szundi, I., Fu, W. Y., Sakmar, T. P. & Kliger, D. S. (2000). pH dependence of photolysis intermediates in the photoactivation of rhodopsin mutant E113Q. *Biochemistry* **39**, 599-606.

[68] Schreiber, M., Buss, V. & Sugihara, M. (2003). Exploring the Opsin shift with ab initio methods: Geometry and counterion effects on the electronic spectrum of retinal. *Journal of Chemical Physics* **119**, 12045-12048.

[69] Franke, R. R., Sakmar, T. P., Graham, R. M. & Khorana, H. G. (1992). Structure and Function in Rhodopsin - Studies of the Interaction between the Rhodopsin Cytoplasmic Domain and Transducin. *Journal of Biological Chemistry* **267**, 14767-14774.

[70] Fahmy, K. & Sakmar, T. P. (1993). Regulation of the Rhodopsin Transducin Interaction by a Highly Conserved Carboxylic-Acid Group. *Biochemistry* **32**, 7229-7236.

[71] Min, K. C., Zvyaga, T. A., Cypess, A. M. & Sakmar, T. P. (1993). Characterization of Mutant Rhodopsins Responsible for Autosomal-Dominant Retinitis-Pigmentosa - Mutations on the Cytoplasmic Surface Affect Transducin Activation. *Journal of Biological Chemistry* **268**, 9400-9404.

[72] Han, M., Lin, S. W., Minkova, M., Smith, S. O. & Sakmar, T. P. (1996). Functional interaction of transmembrane helices 3 and 6 in rhodopsin - Replacement of phenylalanine 261 by alanine causes reversion of phenotype of a glycine 121 replacement mutant. *Journal of Biological Chemistry* **271**, 32337-32342.

[73] Han, M., Lin, S. W., Smith, S. O. & Sakmar, T. P. (1996). The effects of amino acid replacements of glycine 121 on transmembrane helix 3 of rhodopsin. *Journal of Biological Chemistry* **271**, 32330-32336.

[74] Sheikh, S. P., Zvyaga, T. A., Lichtarge, O., Sakmar, T. P. & Bourne, H. R. (1996). Rhodopsin activation blocked by metal-ion-binding sites linking transmembrane helices C and F. *Nature* **383**, 347-350.

[75] Abdulaev, N. G. (2003). Building a stage for interhelical play in rhodopsin. *Trends in Biochemical Sciences* **28**, 399-402.

[76] Franke, R. R., Sakmar, T. P., Oprian, D. D. & Khorana, H. G. (1988). A Single Amino-Acid Substitution in Rhodopsin (Lysine-248-]Leucine) Prevents Activation of Transducin. *Journal of Biological Chemistry* **263**, 2119-2122.

[77] Ernst, O. P., Meyer, C. K., Marin, E. P., Henklein, P., Fu, W. Y., Sakmar, T. P. & Hofmann, K. P. (2000). Mutation of the fourth cytoplasmic loop of rhodopsin affects binding of transducin and peptides derived from the carboxyl-terminal sequences of transducin alpha and gamma subunits. *Journal of Biological Chemistry* **275**, 1937-1943.

[78] Karnik, S. S., Sakmar, T. P., Chen, H. B. & Khorana, H. G. (1988). Cysteine Residue-110 and Residue-187 Are Essential for the Formation of Correct Structure in Bovine Rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 8459-8463.

[79] Shieh, T., Han, M., Sakmar, T. P. & Smith, S. O. (1997). The steric trigger in rhodopsin activation. *Journal of Molecular Biology* **269**, 373-384.

[80] Yeagle, P. L. & Albert, A. D. (2003). A conformational trigger for activation of a G protein by a G protein-coupled receptor. *Biochemistry* **42**, 1365-1368.

[81] Spooner, P. J. R., Sharples, J. M., Goodall, S. C., Bovee-Geurts, P. H. M., Verhoeven, M. A., Lugtenburg, J., Pistorius, A. M. A., DeGrip, W. J. & Watts, A. (2004). The ring of the rhodopsin chromophore in a hydrophobic activation switch within the binding pocket. *Journal of Molecular Biology* **343**, 719-730.

[82] Sondek, J., Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. (1994). Gtpase Mechanism of Gproteins from the 1.7-Angstrom Crystal-Structure of Transducin Alpha-Center-Dot-Gdp-Center-Dot-Alf4(-). *Nature* **372**, 276-279.

[83] Lambright, D. G., Sondek, J., Bohm, A., Skiba, N. P., Hamm, H. E. & Sigler, P. B. (1996). The 2.0 angstrom crystal structure of a heterotrimeric G protein. *Nature* **379**, 311-319.

[84] Sondek, J., Bohm, A., Lambright, D. G., Hamm, H. E. & Sigler, P. B. (1996). Crystal structure of a G-protein beta gamma dimer at 2.1 angstrom resolution (vol 379, pg 369, 1996). *Nature* **379**, 847-847.

[85] Sondek, J., Bohm, A., Lambright, D. G., Hamm, H. E. & Sigler, P. B. (1996). Crystal structure of a G(A) protein beta gamma dimer at 2.1 angstrom resolution. *Nature* **379**, 369-374.

[86] Filipek, S., Krzysko, K. A., Fotiadis, D., Liang, Y., Saperstein, D. A., Engel, A. & Palczewski, K. (2004). A concept for G protein activation by G protein-coupled receptor dimers: the transducin/rhodopsin interface. *Photochemical & Photobiological Sciences* **3**, 628-638.

[87] Ling, Y., Ascano, M., Robinson, P. & Gregurick, S. K. (2004). Experimental and computational studies of the desensitization process in the bovine rhodopsin-arrestin complex. *Biophysical Journal* **86**, 2445-2454.

[88] Zvyaga, T. A., Fahmy, K. & Sakmar, T. P. (1994). Characterization of Rhodopsin-Transducin Interaction - a Mutant Rhodopsin Photoproduct with a Protonated Schiff-Base Activates Transducin. *Biochemistry* **33**, 9753-9761.

[89] Ernst, O. P., Hofmann, K. P. & Sakmar, T. P. (1995). Characterization of Rhodopsin Mutants That Bind Transducin but Fail to Induce Gtp Nucleotide Uptake - Classification of Mutant Pigments by Fluorescence, Nucleotide Release, and Flash-Induced Light-Scattering Assays. *Journal of Biological Chemistry* **270**, 10580-10586.

[90] Yang, K., Farrens, D. L., Hubbell, W. L. & Khorana, H. G. (1996). Structure and function in rhodopsin. Single cysteine substitution mutants in the cytoplasmic interhelical E-F loop region show position-specific effects in transducin activation. *Biochemistry* **35**, 12464-12469.

[91] Beck, M., Sakmar, T. P. & Siebert, F. (1998). Spectroscopic evidence for interaction between transmembrane helices 3 and 5 in rhodopsin. *Biochemistry* **37**, 7630-7639.

[92] Han, M., Smith, S. O. & Sakmar, T. P. (1998). Constitutive activation of opsin by mutation of methionine 257 on transmembrane helix 6. *Biochemistry* **37**, 8253-8261.
[93] Gurevich, V. V. & Gurevich, E. V. (2003). The new face of active receptor bound minireview arrestin attracts new partners. *Structure* **11**, 1037-1042.
[94] Smith, W. C., Dinculescu, A., Peterson, J. J. & McDowell, J. H. (2004). The surface of visual arrestin that binds to rhodopsin. *Molecular Vision* **10**, 392-398.
[95] Marin, E. P., Krishna, K. G., Zvyaga, T. A., Isele, J., Siebert, F. & Sakmar, T. P. (2000). The amino terminus of the fourth cytoplasmic loop of rhodopsin modulates rhodopsin-transducin interaction. *Journal of Biological Chemistry* **275**, 1930-1936.
[96] Marin, E. P., Krishna, A. G., Archambault, V., Simuni, E., Fu, W. Y. & Sakmar, T. P. (2001). The function of interdomain interactions in controlling nucleotide exchange rates in transducin. *Journal of Biological Chemistry* **276**, 23873-23880.
[97] Marin, E. P., Krishna, A. G. & Sakmar, T. P. (2002). Disruption of the alpha 5 helix of transducin impairs rhodopsin-catalyzed nucleotide exchange. *Biochemistry* **41**, 6988-6994.
[98] Raman, D., Osawa, S., Gurevich, V. V. & Weiss, E. R. (2003). The interaction with the cytoplasmic loops of rhodopsin plays a crucial role in arrestin activation and binding. *Journal of Neurochemistry* **84**, 1040-1050.
[99] Han, M., Groesbeek, M., Sakmar, T. P. & Smith, S. O. (1997). The C9 methyl group of retinal interacts with glycine-121 in rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13442-13447.
[100] Jager, S., Han, M., Lewis, J. W., Szundi, I., Sakmar, T. P. & Kliger, D. S. (1997). Properties of early photolysis intermediates of rhodopsin are affected by glycine 121 and phenylalanine 261. *Biochemistry* **36**, 11804-11810.
[101] Lin, S. W., Kochendoerfer, G. G., Carroll, H. S., Wang, D., Mathies, R. A. & Sakmar, T. P. (1998). Mechanisms of spectral tuning in blue cone visual pigments - Visible and raman spectroscopy of blue-shifted rhodopsin mutants. *Journal of Biological Chemistry* **273**, 24583-24591.
[102] Kochendoerfer, G. G., Lin, S. W., Sakmar, T. P. & Mathies, R. A. (1999). How color visual pigments are tuned. *Trends in Biochemical Sciences* **24**, 300-305.
[103] Cowing, J. A., Poopalasundaram, S., Wilkie, S. E., Robinson, P. R., Bowmaker, J. K. & Hunt, D. M. (2002). The molecular mechanism for the spectral shifts between vertebrate ultraviolet- and violet-sensitive cone visual pigments. *Biochemical Journal* **367**, 129-135.
[104] Janz, J. M. & Farrens, D. L. (2003). Assessing structural elements that influence Schiff base stability: mutants E113Q and D190N destabilize rhodopsin through different mechanisms. *Vision Research* **43**, 2991-3002.
[105] Lewis, J. W., Szundi, I., Kazmi, M. A., Sakmar, T. P. & Kliger, D. S. (2004). Time-resolved photointermediate changes in rhodopsin glutamic acid 181 mutants. *Biochemistry* **43**, 12614-12621.
[106] Patel, A. B., Crocker, E., Eilers, M., Hirshfeld, A., Sheves, M. & Smith, S. O. (2004). Coupling of retinal isomerization to the activation of rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10048-10053.
[107] Rader, A. J., Anderson, G., Isin, B., Khorana, H. G., Bahar, I. & Klein-Seetharaman, J. (2004). Identification of core amino acids stabilizing rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7246-7251.
[108] Han, M., Groesbeek, M., Smith, S. O. & Sakmar, T. P. (1998). Role of the C-9 methyl group in rhodopsin activation: Characterization of mutant opsins with the artificial chromophore 11-cis-9-demethylretinal. *Biochemistry* **37**, 538-545.

[109] Stecher, H., Prezhdo, O., Das, J., Crouch, R. K. & Palczewski, K. (1999). Isomerization of all-trans-9-and 13-desmethylretinol by retinal pigment epithelial cells. *Biochemistry* **38**, 13542-13550.

[110] Hirano, T., Lim, I. T., Kim, D. M., Zheng, X. G., Yoshihara, K., Oyama, Y., Imai, H., Shichida, Y. & Ishiguro, M. (2002). Constraints of opsin structure on the ligand-binding site: Studies with ring-fused retinals. *Photochemistry and Photobiology* **76**, 606-615.

[111] Das, J., Crouch, R. K., Ma, J. X., Oprian, D. D. & Kono, M. (2004). Role of the 9-methyl group of retinal in cone visual pigments. *Biochemistry* **43**, 5532-5538.

[112] Vogel, R., Siebert, F., Ludeke, S., Hirshfeld, A. & Sheves, M. (2005). Agonists and partial agonists of rhodopsin: Retinals with ring modifications. *Biochemistry* **44**, 11684-11699.

[113] Hunt, D. M., Dulai, K. S., Cowing, J. A., Julliot, C., Mollon, J. D., Bowmaker, J. K., Li, W. H. & Hewett-Emmett, D. (1998). Molecular evolution of trichromacy in primates. *Vision Research* **38**, 3299-3306.

[114] Chan, T., Lee, M. & Sakmar, T. P. (1992). Introduction of Hydroxyl-Bearing Amino-Acids Causes Bathochromic Spectral Shifts in Rhodopsin - Amino-Acid Substitutions Responsible for Red-Green Color Pigment Spectral Tuning. *Journal of Biological Chemistry* **267**, 9478-9480.

[115] Beck, M., Siebert, F. & Sakmar, T. P. (1998). Evidence for the specific interaction of a lipid molecule with rhodopsin which is altered in the transition to the active state metarhodopsin II. *Febs Letters* **436**, 304-308.

[116] Isele, J., Sakmar, T. P. & Siebert, F. (2000). Rhodopsin activation affects the environment of specific neighboring phospholipids: An FTIR spectroscopic study. *Biophysical Journal* **79**, 3063-3071.

[117] Litman, B. J., Niu, S. L., Polozova, A. & Mitchell, D. C. (2001). The role of docosahexaenoic acid containing phospholipids in modulating G protein-coupled signaling pathways - Visual transduction. *Journal of Molecular Neuroscience* **16**, 237-242.

[118] Mitchell, D. C., Niu, S. L. & Litman, B. J. (2001). Optimization of receptor-G protein coupling by bilayer lipid composition I - Kinetics of rhodopsin-transducin binding. *Journal of Biological Chemistry* **276**, 42801-42806.

[119] Niu, S. L., Mitchell, D. C. & Litman, B. J. (2001). Optimization of receptor-G protein coupling by bilayer lipid composition II - Formation of metarhodopsin II-transducin complex. *Journal of Biological Chemistry* **276**, 42807-42811.

[120] Feller, S. E., Gawrisch, K. & Woolf, T. B. (2003). Rhodopsin exhibits a preference for solvation by polyunsaturated docosohexaenoic acid. *Journal of the American Chemical Society* **125**, 4434-4435.

[121] Mitchell, D. C., Niu, S. L. & Litman, B. J. (2003). DHA-rich phospholipids optimize G-protein-coupled signaling. *Journal of Pediatrics* **143**, S80-S86.

[122] Mitchell, D. C., Niu, S. L. & Litman, B. J. (2003). Enhancement of G protein-coupled signaling by DHA phospholipids. *Lipids* **38**, 437-443.

[123] Grossfield A., F., Feller, S.E., and Pitman, M.C. (2006). A role for the direct interactions in the modulation of rhodopsin by ω-3 polyunsaturated lipids. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4888-4893.

[124] Feller, S. E., Gawrisch, K. & MacKerell, A. D. (2002). Polyunsaturated fatty acids in lipid bilayers: Intrinsic and environmental contributions to their unique physical properties. *Journal of the American Chemical Society* **124**, 318-326.

[125] Gerken, U., Jelezko, F., Gotze, B., Branschadel, M., Tietz, C., Ghosh, R. & Wrachtrup, J. (2003). Membrane environment reduces the accessible conformational space available to an integral membrane protein. *Journal of Physical Chemistry B* **107**, 338-343.

[126] Carrillo-Tripp, M. & Feller, S. E. (2005). Evidence for a mechanism by which omega-3 polyunsaturated lipids may affect membrane protein function. *Biochemistry* **44**, 10164-10169.

[127] Sakmar, T. P. (2002). Structure of rhodopsin and the superfamily of seven-helical receptors: the same and not the same. *Current Opinion in Cell Biology* **14**, 189-195.

[128] Archer, E., Maigret, B., Escrieut, C., Pradayrol, L. & Fourmy, D. (2003). Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors? *Trends in Pharmacological Sciences* **24**, 36-40.

[129] Mirzadegan, T., Benko, G., Filipek, S. & Palczewski, K. (2003). Sequence analyses of G-protein-coupled receptors: Similarities to rhodopsin. *Biochemistry* **42**, 2759-2767.

[130] Ballesteros, J. A., Shi, L. & Javitch, J. A. (2001). Structural mimicry in G protein-coupled receptors: Implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors. *Molecular Pharmacology* **60**, 1-19.

[131] Becker, O. M., Shacham, S., Marantz, Y. & Noiman, S. (2003). Modeling the 3D structure of GPCRs: Advances and application to drug discovery. *Current Opinion in Drug Discovery & Development* **6**, 353-361.

[132] Bissantz, C., Bernard, P., Hibert, M. & Rognan, D. (2003). Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins-Structure Function and Genetics* **50**, 5-25.

[133] Cavasotto, C. N., Orry, A. J. W. & Abagyan, R. A. (2003). Structure-based identification of binding sites, native ligands and potential inhibitors for G-protein coupled receptors. *Proteins-Structure Function and Genetics* **51**, 423-433.

[134] Filipek, S., Teller, D. C., Palczewski, K. & Stenkamp, R. (2003). The crystallographic model of rhodopsin and its use in studies of other G protein-coupled receptors. *Annual Review of Biophysics and Biomolecular Structure* **32**, 375-397.

[135] Mosberg, H. I. & Fowler, C. B. (2003). Do GPCR models derived by homology from the rhodopsin xray structure correctly predict helix irregularities? *Biopolymers* **71**, 387-387.

[136] Bartfai, T., Benovic, J. L., Bockaert, J., Bond, R. A., Bouvier, M., Christopoulos, A., Civelli, O., Devi, L. A., George, S. R., Inui, A., Kobilka, B., Leurs, R., Neubig, R., Pin, J. P., Quirion, R., Roques, B. P., Sakmar, T. P., Seifert, R., Stenkamp, R. E. & Strange, P. G. (2004). The state of GPCR research in 2004. *Nature Reviews Drug Discovery* **3**, 574-626.

[137] Becker, O. M., Marantz, Y., Shacham, S., Inbal, B., Heifetz, A., Kalid, O., Bar-Haim, S., Warshaviak, D., Fichman, M. & Noiman, S. (2004). G protein-coupled receptors: In silico drug discovery in 3D. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 11304-11309.

[138] Hillisch, A., Pineda, L. F. & Hilgenfeld, R. (2004). Utility of homology models in the drug discovery process. *Drug Discovery Today* **9**, 659-669.

[139] Oliveira, L., Hulsen, T., Hulsik, D. L., Paiva, A. C. M. & Vriend, G. (2004). Heavier-than-air flying machines are impossible. *Febs Letters* **564**, 269-273.

[140] Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P. & Bowie, J. U. (2004). The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 959-963.

[141] Bosch, L., Iarriccio, L. & Garriga, P. (2005). New prospects for drug discovery from structural studies of rhodopsin. *Current Pharmaceutical Design* **11**, 2243-2256.
[142] Bissantz, C. (2003). Conformational changes of G protein-coupled receptors during their activation by agonist binding. *Journal of Receptors and Signal Transduction* **23**, 123-153.
[143] Mehler, E. L., Periole, X., Hassan, S. A. & Weinstein, H. (2002). Key issues in the computational simulation of GPCR function: representation of loop domains. *Journal of Computer-Aided Molecular Design* **16**, 841-853.
[144] Forrest, L. R. & Woolf, T. B. (2003). Discrimination of native loop conformations in membrane proteins: Decoy library design and evaluation of effective energy scoring functions. *Proteins-Structure Function and Genetics* **52**, 492-509.
[145] Lawson, Z. & Wheatley, M. (2004). The third extracellular loop of G-protein-coupled receptors: more than just a linker between two important transmembrane helices. *Biochemical Society Transactions* **32**, 1048-1050.
[146] Nikiforovich, G. V. & Marshall, G. R. (2005). Modeling flexible loops in the dark-adapted and activated states of rhodopsin, a prototypical G-protein-coupled receptor. *Biophysical Journal* **89**, 3780-3789.
[147] Okada, T., Ernst, O. P., Palczewski, K. & Hofmann, K. P. (2001). Activation of rhodopsin: new insights from structural and biochemical studies. *Trends in Biochemical Sciences* **26**, 318-324.
[148] Okada, T. & Palczewski, K. (2001). Crystal structure of rhodopsin: implications for vision and beyond. *Current Opinion in Structural Biology* **11**, 420-426.
[149] Teller, D. C., Okada, T., Behnke, C. A., Palczewski, K. & Stenkamp, R. E. (2001). Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). *Biochemistry* **40**, 7761-7772.
[150] Stenkamp, R. E., Filipek, S., Driessen, C., Teller, D. C. & Palczewski, K. (2002). Crystal structure of rhodopsin: a template for cone visual pigments and other G protein-coupled receptors. *Biochimica Et Biophysica Acta-Biomembranes* **1565**, 168-182.
[151] Fishkin, N., Berova, N. & Nakanishi, K. (2004). Primary events in dim light vision: A chemical and spectroscopic approach toward understanding protein/chromophore interactions in rhodopsin. *Chemical Record* **4**, 120-135.
[152] Liu, R. S. H. & Colmenares, L. U. (2003). The molecular basis for the high photosensitivity of rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14639-14644.
[153] Humphrey, W., Dalke, A. and Schulten K. (1996). VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **14**, 33-38.

Chapter 3: Capillary waves at the liquid-vapor interface and the surface tension of water models

Ahmed E. Ismail, Gary S. Grest, and Mark J. Stevens
Sandia National Laboratories, Albuquerque, New Mexico 87185

Abstract

Capillary waves occurring at the liquid-vapor interface of water are studied using molecular dynamics simulations. In addition, the surface tension, determined thermodynamically from the difference in the normal and tangential pressure at the liquid-vapor interface, is compared for a number of standard three- and four-point water models. We study four three-point models (SPC/E, TIP3P, TIP3P-CHARMM, and TIP3P-Ew) and two four-point models (TIP4P and TIP4P-Ew). All of the models examined underestimate the surface tension; the TIP4P-Ew model comes closest to reproducing the experimental data. The surface tension can also be determined from the amplitude of capillary waves at the liquid-vapor interface by varying the surface area of the interface. The surface tensions determined from the amplitude of the logarithmic divergence of the capillary interfacial width and from the traditional thermodynamic method agree only if the density profile is fitted to an error function instead of a hyperbolic tangent function.

I. Introduction

The ability to derive accurate property predictions for the liquid-vapor interface is a key test for an atomistic force field. Because of the frequent occurrence of water in systems of chemical and biological interest, interfacial property prediction is especially vital for force fields of water. The most important of these properties is surface tension, an intensive quantity that measures the differential surface work required to increase the interfacial area. Accurate models of the surface tension of water are essential for conducting large-scale simulations of the wetting and spreading of water droplets at surfaces.

An interface between two distinct thermodynamic phases can be characterized by a local gradient of an order parameter whose mean value changes between phases, such as the boundary between a liquid and its own vapor below the critical temperature T_c. For simple fluids, thermodynamic arguments predict that the interfacial width Δ depends only on temperature and the interaction energies within each phase and across the interface. However, the presence of the interface breaks the translational invariance of the system, inducing Goldstone fluctuations or "capillary waves" at the interface [1, 2]. Previous studies of capillary waves involving water have tended to focus on liquid-liquid interfaces or on model fluids [3, 4], and have generally examined relatively small systems of less than $10,000$ molecules; the present study represents the first study of capillary-wave behavior at the liquid-vapor interface of water.

For two-dimensional interfaces, these non-critical fluctuations give rise to a logarithmic increase in the interfacial width Δ with increasing $L_{\square}$, the length of the interface. Most previous simulations [5, 6] of the liquid-vapor interface in three dimensions did not investigate the dependence of Δ on the size of the interface. The purpose of this paper is to present atomistic molecular dynamics (MD) simulations of the liquid-vapor interface of water. In particular, we obtain the surface tension γ in two different ways: from the difference in pressure parallel $p_{\square}$ and perpendicular $p_{\perp}$ to the interface (γ_p), and from the dependence of Δ on $L_{\square}$ (γ_w). We confirm the previous result that γ_w depends on the functional form chosen to fit the order parameter (density profile) through the interface [7]. In particular, fitting the order parameter to an error function gives results for γ_w which are in strong agreement with γ_p. However, fitting our data to a hyperbolic tangent function, a functional form derived from mean-field arguments [2], gives results for γ_w which are systematically smaller than γ_p and further away from experimental results.

There are currently a large number of different atomic models for water. Guillot provides an extensive list of models developed through 2001 [8]; several additional models have been introduced since then [9-12]. The simplest of the commonly-used atomic models, the SPC model [13], is a rigid three-point model with fixed charges; the most complex model, the POL5 model [14], is a polarizable five-point model. Most of the commonly used models are three- and four-

point models. In three-point models, such as SPC/E [15] and TIP3P [16], the electric charges are assigned directly to the hydrogen and oxygen atoms; four-point models, such as the TIP4P [16] and Watanabe-Klein [17] models, locate the negative charge at a massless point a fixed distance away from the oxygen atom. Five-point models, such as TIP5P [18], and the early Bernal-Fowler [19] and ST2 [20] models, represent the negative charge of the oxygen using a pair of massless charges to capture the quadrupolar behavior of water. Polarizable models, including the SPC/FQ and TIP4P/FQ models [21], allow the magnitude of the point charges to be treated as variables which can fluctuate according to the local environment.

The proliferation of models has been motivated largely by the need to reproduce various physical and thermodynamic properties, such as the bulk density, the oxygen-oxygen radial distribution function, the heat of vaporization, and the diffusion coefficient. However, some models, such as the recent TIP3P-Ew [10] and TIP4P-Ew models [11], are reparameterizations of existing models designed to account for changes in the treatment of long-ranged electrostatic interactions. Most of the available water models adequately represent at least some of the thermodynamic properties of water; for a comprehensive review, see Jorgensen *et al.* [12]. Kuo *et al.* have shown that the changes introduced between, for example, the TIP4P and TIP4P/FQ models have little influence on properties such as the bulk liquid density or the mean distance between oxygen atoms either in bulk or at the interface [22]. However, the simulation behavior of models with nearly identical parameters can be markedly different: Mark and Nilsson have noted significant variation in physical and thermodynamic properties such as the self-diffusion constant and the radial distribution function of various three-point water models [23, 24]. Less is known about how well the various water models describe the liquid-vapor interface and the surface tension γ. Our own work, however, suggests that even models with very similar density and distribution profiles can have quite different predictions for surface tension.

Experimental studies demonstrate that the surface tension of water decreases with a slight quadratic dependence on temperature in the range $273\,\mathrm{K} < T < 373\,\mathrm{K}$ [25-29]. Surface tension results for higher temperatures have not been reported in the literature; we extrapolate the reported experimental data to higher temperatures. There have been a few studies for various three-point models [9, 30-36] which show that while most water models reproduce the observed

decrease in surface tension as temperature increases, they tend to underestimate γ by amounts between 25 and 50 percent. Only Alejandre *et al.*, Huang *et al.*, and Shi *et al.* [30, 36, 37] report adequate agreement with experimental data. However, as we show below, the apparent agreement of both Alejandre *et al.* [30] and Shi *et al*. [36] is the result of inadequate simulation time. Alejandre *et al.* also employ a reciprocal-space mesh that is too coarse, while Huang *et al.* report values only for the SPC and SPC/E models at $298K$ [37].

Our primary goal is to study capillary waves at the liquid-vapor interface of water, and to distinguish between various functional representations for the density profile near the interface. Additionally, we first determine the surface tension as a function of temperature for six commonly used three- and four-point models of water, in part to establish a basis for comparison with the capillary-wave simulations.

In Section II, we provide a brief overview of methods for computing the surface tension from molecular simulation data, of the various water models examined in this study, and of the simulation methods employed. Section III presents our findings on the temperature dependence of the surface tension, as well as the effects of the tail correction, interaction cutoffs, and reciprocal-space mesh refinement. We discuss the results obtained from the analysis of capillary waves at the liquid-vapor interface in Section IV before offering our conclusions in Section V.

II. Models and methodology

A. Surface tension

1. Thermodynamic method

There are two primary methods used to compute the surface tension using molecular simulations. The first approach, developed by Tolman [38] and refined by Kirkwood and Buff [39], computes the surface tension as an integral of the difference between the normal and tangential pressures $p_{\perp}(z)$ and $p_{\square}(z)$:

$$\gamma_p = \frac{1}{2}\int_{-\infty}^{\infty} \left(p_{\perp}(z) - p_{\square}(z) \right) dz, \tag{1}$$

where, in our geometry (see Figure 1),

$$p_{\perp}(z) = p_z(z),$$
$$p_{\square}(z) = \left(p_x(z) + p_y(z)\right)/2.$$

The dominant contributions to the integral in Eq. (1) occur near the interface; in the bulk away from the interface, $p_{\perp} = p_{\square}$ and the integrand vanishes. For the specific case shown in Figure 1, where the interface separates a bulk liquid from its corresponding vapor phase, the integral in Eq. (1) can be replaced with an ensemble average of the difference between the normal and tangential pressures,

$$\gamma_p = \frac{L_z}{2}\left\langle p_{\perp} - p_{\square}\right\rangle = \frac{L_z}{2}\left[\left\langle p_z\right\rangle - \frac{\left\langle p_x\right\rangle + \left\langle p_y\right\rangle}{2}\right]. \tag{2}$$

The outer factor of $1/2$ in Eq. (2) accounts for the presence of two liquid-vapor interfaces.

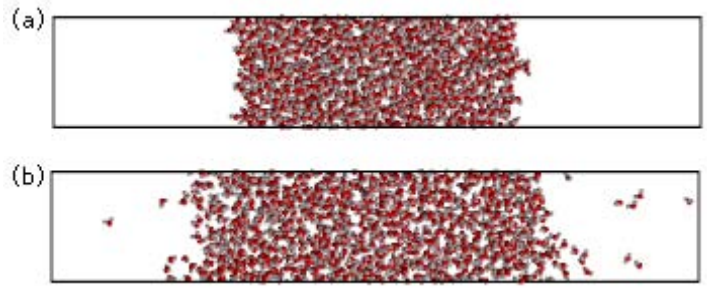

Figure 1. Sample simulation cell, showing equilibrated configurations of 1000 SPC/E water molecules at (a) 300 K and (b) 500 K. The dimensions of the cell are $L_x = L_y = L_{\square} = 2.3$ nm and $L_z = L_{\perp} = 13.5$ nm.

In most numerical simulations, the interatomic and electrostatic interactions are only applied within a cutoff range r_c. The introduction of the cutoff in the interparticle potential reduces the surface tension in much the same way that the introduction of a cutoff reduces the bulk pressure at constant density. Thus, the simulation result γ_p will underestimate the total surface tension; a better estimate of the total surface tension can be obtained from

$$\gamma = \gamma_p + \gamma_{tail}, \tag{3}$$

where γ_{tail}, the tail correction for γ_p, can be determined from [40, 41]

$$\gamma_{tail} = \frac{\pi}{2}\int_{-\infty}^{\infty}\int_{-1}^{1}\int_{r_c}^{\infty} r^3 U'(r)\, g(r)\left(1-3s^2\right)\times\left(\rho(z)\,\rho(z-sr) - \left(\rho_G(z)\right)^2\right) dr\, ds\, dz, \tag{4}$$

where $U(r)$ is the pairwise potential, $g(r)$ is the radial distribution function, $\rho(z)$ is the observed interfacial profile, and $\rho_G(z)$ is a Gibbs dividing surface:

$$\rho_G(z) = \rho_c + \frac{\Delta\rho}{2} * sgn(z). \tag{5}$$

Although the use of the tail correction in Eq. (3) improves the estimate of the surface tension, its use is restricted to systems in which the two phases contain the same components; for composite systems, such as water at the surface of a solid, only γ_p should be used.

We assume in Eqs. (4) and (5) that the liquid-vapor interface is centered at $z = 0$. In Eq. (5), $\rho_c = (\rho_l + \rho_v)/2$ is the average density of the two phases, and $\Delta\rho = \rho_l - \rho_v$ is the difference between the average densities of the two phases. Thus, $\rho_G(z) = \rho_v$ for $z < 0$ and ρ_l for $z > 0$. There are multiple possible choices for determining the observed density profile $\rho(z)$. Although it is possible to use the profile calculated from the simulation directly, both the tail-correction and capillary-wave calculations are simplified by fitting the profile to a function. In the present study, the density profile is fitted to both an error function and a hyperbolic tangent function, as discussed in the following section.

2. Capillary-wave method

The thermodynamic approach for computing surface tension assumes a sharp liquid-vapor interface when in reality it is quite rough. The roughness of the interface increases at high temperatures, as seen in Figure 1. A second method for computing the surface tension assumes that the observed magnitude of the fluctuations is derived from two sources: an intrinsic contribution plus a logarithmic term that represents broadening of the interface as a result of the capillary waves [1, 7, 42-45].

If the contributions from capillary waves can be decoupled from density fluctuations, then the surface tension can be computed by determining the interfacial profiles for a number of different system sizes. The relationship between the observed interfacial width Δ and the intrinsic interfacial width Δ_0 is given by

$$\Delta^2 = \Delta_0^2 + \frac{k_B T}{2\pi\gamma_w} \ln\left(\frac{L_\square}{B_0}\right), \tag{6}$$

where $L_\square = L_x = L_y$ is the length of the interface, and B_0 is a characteristic length scale related to the short-wavelength cutoff in the interfacial behavior. It is unnecessary to determine B_0 before computing the surface tension γ_w.

Computation of γ_w requires the scaled density profile,

$$\Psi(z) = \frac{2}{\rho_L - \rho_V}\left(\rho(z) - \frac{\rho_L + \rho_V}{2}\right), \tag{7}$$

in the z-direction. Given $\Psi(z)$, the variance in the derivative of the profile $f(z) = \Psi'(z)$ can be computed in either real or reciprocal (Fourier) space:

$$\Delta^2 = \frac{\int_{-\infty}^{\infty} z^2 f(z)\, dz}{\int_{-\infty}^{\infty} f(z)\, dz} = -\frac{1}{\tilde{f}(0)}\left[\frac{d^2\tilde{f}(q)}{dq^2}\right]_{q=0}, \tag{8}$$

where $\tilde{f}(q)$ is the Fourier transform of $f(z)$. The simple form of Eq. (8) suggests that fitting the profile $\Psi(z)$ to a functional form will be both more convenient and lead to more accurate results than using the raw profile data. Several different functional forms for $\Psi(z)$ have been proposed in the literature. Relying on mean-field arguments, most previous theoretical and computational studies of surface tension have fitted the profile to a hyperbolic tangent function [22, 30, 40-41],

$$\Psi_t(z) = \tanh\left(\frac{2z}{w_t}\right), \tag{9}$$

while Huang and Webb [42] and Beysens and Robert [43] propose the use of an error function,

$$\Psi_e(z) = erf\left(\frac{\sqrt{\pi}z}{w_e}\right). \tag{10}$$

If the density profile $\Psi(z)$ is fitted to a hyperbolic tangent function Eq. (9), then from Eq. (8) we find that [7]

$$\Delta_t^2 = \pi^2 w_t^2 / 48,$$

while for an error function Eq. (10), the interfacial width Δ_e^2 is given by

$$\Delta_e^2 = w_e^2/2\pi.$$

We will show that there is a significant discrepancy between the surface tensions obtained from the hyperbolic tangent profile, Eq. (9), and the error function profile, Eq. (10), with the error function giving results in closer agreement with Eq. (2).

B. Water models

We consider four different three-point models: the SPC/E model; the original TIP3P model; the modification of the TIP3P model [46] implemented in CHARMM (hereafter referred to as TIP3P-C); and the TIP3P-Ew model [10], a recent reparameterization incorporating the effects of Ewald summation. The parameters for the different water models are summarized in Table I.

The basic structure of the different models is similar. The common features of all models include a specified oxygen-hydrogen bond length l_{OH} and hydrogen-oxygen-hydrogen bond angle θ_{HOH}, a charge on each hydrogen atom, and a Lennard-Jones 12-6 potential describing the interaction between the oxygen atoms,

$$U_{LJ}(r_{OO}) = \begin{cases} 4\varepsilon_{OO}\left[\left(\dfrac{\sigma_{OO}}{r_{OO}}\right)^{12} - \left(\dfrac{\sigma_{OO}}{r_{OO}}\right)^{6}\right], & r_{OO} \le r_c \\ 0, & r_{OO} > r_c \end{cases}, \tag{11}$$

where ε_{OO} and σ_{OO} are the model-dependent well depth and equilibrium O-O distance, and r_{OO} is the distance between oxygen atoms. The TIP3P-C model incorporates hydrogen-hydrogen and hydrogen-oxygen Lennard-Jones interactions as well.

TABLE I: Parameters for commonly used three- and four-point models of water.

Parameter	SPC/E	TIP3P	TIP3P-C	TIP3P-Ew	TIP4P	TIP4P-Ew
q_H	0.410	0.417	0.417	0.415	0.520	0.5242
q_O	-0.820	-0.834	-0.834	-0.830		
q_M					-1.040	-1.0484
θ_{HOH} **, deg**	109.47	104.52	104.52	104.52	104.52	104.52
l_{OH} **, Å**	1.0	0.9572	0.9572	0.9572	0.9572	0.9572
l_{OM} **, Å**					0.1500	0.1250
ε_{OO} **, kcal/mol**	0.1553	0.1521	0.1521	0.102	0.1550	0.16275
σ_{OO} **, Å**	3.166	3.1506	3.1507	3.188	3.1536	3.16435
ε_{OH} **, kcal/mol**			0.0836			
σ_{OH} **, Å**			1.7753			
ε_{HH} **, kcal/mol**			0.0460			
σ_{HH} **, Å**			0.4000			

In addition to the Lennard-Jones interaction, there are electrostatic interactions between the charge sites:

$$U_{es}\left(r_{ij}\right)=\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}, \tag{12}$$

where q_α is the charge on atom α, and r_{ij} is the distance between atoms i and j in the simulation box. Previous studies have shown that significant variations in the values obtained for surface tension can occur depending upon how the sum in Eq. (12) is performed [31]. Except in Section 1.3, Ewald summations were used throughout our simulations.

We also consider a pair of four-point water models: the TIP4P model [16], and the TIP4P-Ew model [11], a recent reparameterization designed to account for the presence of long-range interactions. The four-point models introduce a bare charge at a new site, designated M, located

on the bisector of the HOH bond angle; the charge is of strength q_M. The forces acting on the massless site are distributed to the O and H atoms in the same molecule [47]:

$$\mathbf{F}_{ij,O} = (1-2a)\mathbf{F}_{ij,M},$$
$$\mathbf{F}_{ij,H} = a\mathbf{F}_{ij,M},$$

where $a = l_{OM}/\left(l_{OH}\cos(\theta_{HOH}/2)\right)$ and $\mathbf{F}_{ij,M}$ is the force acting on the massless site associated with oxygen i due to atom j. For the TIP4P model, the charge is located $l_{OM} = 0.15$ Å away from the oxygen atom. The TIP4P-Ew model changes the values of l_{OM}, the charge q_M, as well as the separation σ_{OO} and well-depth ε_{OO} of the Lennard-Jones interaction.

C. Simulation method

1. Thermodynamic method

To determine the surface tension of the various three-point water models, 1000 molecules were placed into a periodic, rectangular box of dimensions $L_x = L_y = L_\square = 2.3$ nm and $L_z = L_\perp = 13.5$ nm. The increased system size in the z-direction minimizes the interactions of water molecules in the liquid phase with their z-periodic images through the long-range Coulombic interactions in Eq. (12). Similarly, 1000 molecules of the four-point models were simulated in a box with dimensions $L_x = L_y = L_\square = 2.7$ nm and $L_z = L_\perp = 12.0$ nm, each also containing 1000 molecules. The initial configuration was constructed by placing the water molecules at the center of a simple cubic lattice with 7 molecules each in the x- and y-directions, and the z-spacing chosen to create a density of 0.98 g/cm^3 for the three-point models, and 1.00 g/cm^3 for the four-point models. The same starting configuration was used for all simulations of a given water model. At equilibrium, the thickness of the slab in the z-direction varied between approximately 5.5 nm at 300 K and 7.5 nm at 500 K.

For each of the six models examined, molecular dynamics (MD) simulations were performed in the NVT ensemble in 25-degree increments between 300 K and 500 K using the LAMMPS simulation package [48]. The cutoff for the Lennard-Jones potentials and the short-range cutoff

for the electrostatic potentials were set to 10 Å, unless otherwise specified. The bond lengths and bond angles of the various models were constrained using the SHAKE technique [49]. The equations of motion were integrated using the Verlet algorithm with velocity rescaling to control the temperature. The difference in the surface tension between simulations performed with velocity rescaling and those with a Nosé-Hoover thermostat with a damping constant of 100 ps^{-1} was significantly less than the simulation uncertainty. Each simulation was performed for a total of $2\,\mathrm{ns}$ with time step $\Delta t = 1\,\mathrm{fs}$. The system was allowed to equilibrate for $1\,\mathrm{ns}$; data from the second $1\,\mathrm{ns}$ were used to compute the surface tension.

The electrostatic interactions were calculated using the particle-particle particle-mesh (PPPM) technique of Hockney and Eastwood [50]. The mesh spacing in this work was selected to ensure that the root-mean-squared accuracy of the force calculation was within 10^{-4}; the resulting grid was of dimensions $12\times12\times48$. Most previous simulations were carried out with a maximum of $h_z^{\max} = 20$ cells in the z-direction [30, 31]. In several of those studies, simulations were carried out with $h_z^{\max} = 10$ or less, and some did not include long-range electrostatic interactions at all [32]. We consider the effects of mesh refinement on the surface tension in Section III D.

2. Capillary-wave method

Observation of capillary waves requires simulations with larger interfacial surface areas than were used in the thermodynamic method above. Consequently, we studied systems with $L_x = L_y = L_\square$ varying between $9.2\,\mathrm{nm}$ and $46.0\,\mathrm{nm}$. The resulting simulations used to compute the surface tension have surface areas between $84.6\,\mathrm{nm}^2$ and $2116.0\,\mathrm{nm}^2$, and contained between $16,000$ and $400,000$ water molecules. To construct the initial configuration, we take an equilibrated sample and replicate it multiple times in the x- and y-directions. The SPC/E model was used for this study, as it was the most computationally efficient of the models studied.

To ensure that artifacts from the replication process were eliminated, the simulation time varied between $1.0\,\mathrm{ns}$ and $6.0\,\mathrm{ns}$ as an increasing function of $L_\square$. Only the last $750\,\mathrm{ps}$ of the simulation were used for recording data; the preceding steps were used for equilibration and

discarded. We output the position of every atom after every 5000 timesteps, and then assigned each atom to one of 500 bins depending on its location in the z-direction. To ensure that the interfacial profile was not altered by drifts in the density profile, the profile was shifted so that the center of the mass was located at $z = 0$. Sample profiles for the SPC/E model of water at several temperatures are shown in Figure 2. After the average density profile $\rho(z)$ was computed, the two halves of the profile, on either side of $z = 0$, were averaged together, rescaled to values between -1 and 1 using Eq. (7), and then fitted to hyperbolic tangent and error

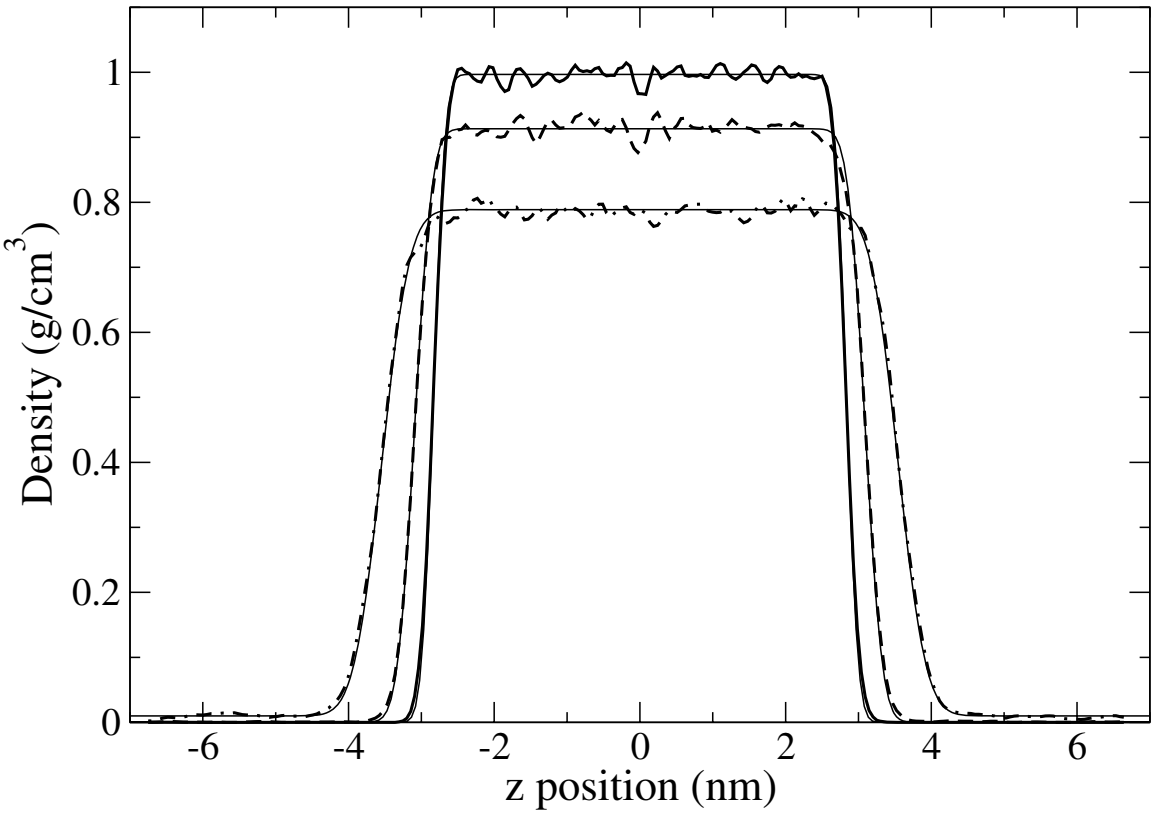

FIGURE 2. Density profiles for the SPC/E model of water at 300 K (thick solid line), 400 K (thick dashed line), and 500 K (thick dashed-dotted line). Fits to error functions are shown as thin solid lines.

III. Thermodynamic Surface Tension: Results and Discussion

A. Temperature dependence

The surface tension $\gamma = \gamma_p + \gamma_{tail}$ of the various water models, computed using Eq. (2), are collected in Table II. Results for the three-point models are shown in Figure 3(a) as a function of temperature. Using the method of Flyvbjerg and Petersen [51], the uncertainty in the results was found to be between 2.4 and 3.0 mN/m. Comparing the four three-point models, we find that SPC/E model is the closest to the experimental data, with better agreement at higher temperatures. For most temperatures, the various TIP3P models agree with the SPC/E model within the uncertainty of the simulation. Although the three-point models considered do not achieve agreement with experimental data, the overall temperature dependence of γ for the models is in good agreement with the experimental data. For the three-point models, γ is generally between 15 mN/m and 20 mN/m less than the experimental data, especially in the temperature range $300\ \text{K} < T < 425\ \text{K}$.

TABLE II: Surface tension[a] for three- and four-point water models, including tail correction.

	Surface Tension, γ (mN/m)						
T (K)	**SPC/E**	**TIP3P**	**TIP3P-Ew**	**TIP3P-C**	**TIP4P**	**TIP4P-Ew**	**Expt.**[b]
300	55.4	51.1	47.4	48.8	53.6	61.2	71.7
325	47.9	45.9	43.6	47.5	50.9	55.6	67.6
350	47.0	42.8	39.2	45.7	45.7	52.7	63.2
375	44.6	37.8	37.7	39.9	41.4	48.1	58.4
400	37.6	35.5	34.3	36.9	35.9	43.5	53.3
425	32.0	31.5	28.9	31.9	31.2	38.6	47.9
450	30.6	27.3	25.9	28.2	25.7	34.5	42.1
475	26.8	24.7	23.2	23.2	19.1	29.3	36.0
500	23.2	17.0	16.9	18.1	15.2	24.8	29.5

Notes: [a] Uncertainty for all simulation results is between 2.4 and 3.0 mN/m. [b] Experimental data taken from Refs. [25-27]; data above 400 K is extrapolated from quadratic fit provided in Ref. [26].

Alejandre *et al*. [30] and Shi *et al.* [36] reported excellent agreement with experimental results for the SPC/E model. However, our results for the SPC/E model clearly disagree with their data as well as with experimental results, although the simulations were performed under essentially identical conditions with respect to the number of molecules and the dimensions of the system, potentials employed, temperature range, and cutoffs. We study the potential causes of the disagreement in the results below.

Results for the four-point models as a function of temperature are shown in Figure 3(b). The uncertainty for the four-point models is the same as for the three-point models. Like the three-point models, the four-point models underestimate the surface tension, with the TIP4P model offering results comparable to the TIP3P-C and TIP3P-Ew models, while the performance of the TIP4P-Ew model is significantly closer to the experimental data than any of the other models examined here. Unlike the TIP3P models, at most temperatures considered here, the TIP4P and TIP4P-Ew models do not agree within simulation uncertainty.

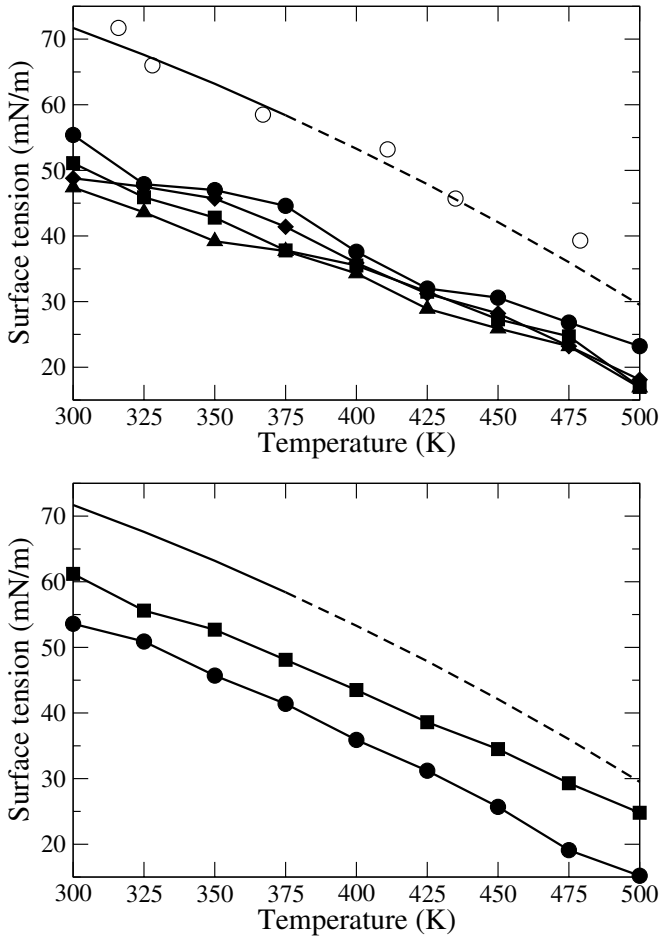

Figure 3. (Top figure) Surface tension of the three-point models of water as a function of temperature: SPC/E (filled circles), TIP3P (squares), TIP3P-C (triangles), and TIP3P-Ew (diamonds). Simulation data from Alejandre *et al.* [30] (open circles) are included for comparison. (Bottom figure) Surface tension of the four-point models of water as a function of temperature: TIP4P (circles) and TIP4P-Ew (squares). In both figures, experimental data [25-27] (solid curve) and extrapolation of quadratic fit to higher temperatures (dashed curve) are included for comparison.

A. Tail correction

While tail corrections can exist for both the Lennard-Jones and the electrostatic interactions, by using Ewald summations, we avoid the need for a Coulombic tail correction. In evaluating the tail correction, Eq. (4) for the Lennard-Jones potential in the region $r \sigma$, $dU(r)/dr \approx 24\varepsilon\sigma^6 r^{-7}$. Assuming that for $r > r_c$, the radial distribution function $g(r) \approx 1$ and that the density profile is an error function of the form Eq. (10), we can evaluate Eq. (4). After integration over s,

$$\frac{\gamma_{tail}}{12\pi\varepsilon\sigma^6} = \int_{-\infty}^{\infty}\int_{r_c}^{\infty} \frac{erf(\xi z)}{\sqrt{\pi}\xi^3 r^7} e^{-\xi^2(r+z)^2} \times$$
$$\left(\xi^2 z(r-z) + e^{4\xi^2 rz}\left(\xi^2 z(z+r)+1\right)-1\right) \qquad (13)$$
$$-\frac{z\,erf(\xi z)}{2\xi^2 r^7}\left(3-\xi^2\left(z^2-r^2\right)\right) f(\xi,r,z)\,dr\,dz,$$

where $\xi \equiv \sqrt{\pi}/w_e$ and

$$f(\xi,r,z) = erf\left(\xi(r+z)\right) + erf\left(\xi(r-z)\right).$$

Eq. (13) is then evaluated numerically for each density profile. For the SPC/E model of water, the interfacial thickness, density difference, and tail correction are shown in Table III. Because of the existence of capillary waves at the interface, as shown in Section IV, the interfacial thickness parameter w_e depends logarithmically on the length $L_{\square}$ of the interfacial surface.

The tail correction depends strongly upon the Lennard-Jones parameters, scaling as $\varepsilon\sigma^6$, and decays exponentially as a function of the chosen interaction cutoff r_c. Because ε_{OO} and σ_{OO} are approximately equal for the SPC/E and TIP3P models, the tail corrections at all temperatures are almost identical for the two models. While the TIP3P-C has additional tail corrections for the O-H and H-H interactions, their magnitudes are negligible in comparison to the correction for the O-O interaction. Only the TIP3P-Ew model, which has a significantly smaller value for the Lennard-Jones interaction strength ε_{OO}, has a noticeably different tail correction from the other three-point models. The tail corrections for the four-point models are likewise close to that of the SPC/E model, with the TIP4P model having a slightly smaller tail correction and the TIP4P-Ew model a slightly larger tail correction.

TABLE III. Interfacial properties of SPC/E water as a function of temperature, for a 1000-molecule system with r_c = 10 Å and $L_{\square}$ = 2.3 nm.

Temperature T, K	Interface thickness w_e, Å	Liquid density ρ_L, g/cm^3	Vapor density ρ_V, g/cm^3	Tail correction γ_{tail}, mN/m
300	3.12	0.990	0.0005	5.5
325	3.37	0.977	0.0008	5.2
350	3.75	0.959	0.0005	5.0
375	4.22	0.941	0.0005	4.8
400	4.63	0.913	0.0015	4.5
425	5.23	0.886	0.0023	4.1
450	5.74	0.855	0.0049	3.8
475	6.00	0.818	0.0093	3.5
500	7.54	0.779	0.0199	3.0

A. Cutoff effects

The original parameterizations for the SPC/E and TIP3P models of water employed cutoffs for both electrostatic and Lennard-Jones interactions [15, 16]. To study the effect of varying the electrostatic cutoff, we applied a short-range cutoff to both the LJ and electrostatic potentials of the SPC/E model, truncating the potentials at $r_{cut} = 10$, 12, 14, 16, 18, and $20A$. Using Eq. (2) to determine the surface tension, we found that the estimated values of the surface tension were nonsensical, ranging between -3700 and 1700 mN/m, with no value smaller in magnitude than 140 mN/m. This shows that truncated electrostatic potentials are inappropriate for use in the determination of the surface tension of water.

To determine the effect of varying only the range of the Lennard-Jones interaction on γ_p before incorporating the tail correction, we performed runs for the SPC/E model at 300 K with LJ cutoffs of 10, 12, 14, 16, 18, and 20 Å, using the PPPM Ewald technique for the electrostatic forces. The starting configuration for these runs was the final configuration from the 300 K simulation used to compute the surface tension in Section III A. The resulting values of the

surface tension are shown in Table IV. The values for the surface tension are within the simulation uncertainties, although they tend to rise with increasing r_c. This is further reflected in the density profiles, which show that the liquid-phase density ρ_L increases with r_c for values of r_c between 10 Å and 18 Å. The overall effect of the choice of r_c can be seen when comparing the surface tensions of SPC/E water in the temperature range 300 K to 500 K for LJ cutoffs of 10 Å and 16 Å. The resulting data are shown in Table V. The data demonstrate that the surface tensions for 10 Å and 16 Å cutoffs are equal, within simulation uncertainty, after the corresponding tail corrections have been applied to each set of data. Thus, the 10 Å Lennard-Jones cutoff with long-range tail corrections is sufficiently accurate for computing the surface tension.

TABLE IV: Surface tensions γ_p and γ and liquid-phase density ρ_L for the SPC/E model as a function of LJ cutoff r_c, with and without tail correction, at 300 K.

r_c **(Å)**	γ_p **(mN/m)**	γ **(mN/m)**	ρ_L **(g/cm^3)**
10.0	46.3	51.8	0.990
12.0	51.2	55.0	0.992
14.0	47.9	50.6	0.994
16.0	49.7	51.8	0.996
18.0	49.9	51.5	0.996
20.0	52.8	54.1	0.995

TABLE V: Surface tension for the SPC/E model as a function of LJ cutoff.

T **(K)**	γ_p **(mN/m)**		$\gamma = \gamma_p + \gamma_{tail}$ **(mN/m)**	
	$r_c = 10$ **Å**	$r_c = 16$ **Å**	$r_c = 10$ **Å**	$r_c = 16$ **Å**
300	49.9	50.2	55.4	52.3
325	42.8	48.7	47.8	50.7
350	42.2	47.2	47.0	49.1
375	40.0	40.3	44.6	42.1
400	33.3	37.7	37.6	39.4
425	28.1	33.5	32.0	35.0
450	27.0	29.2	30.6	30.6
475	23.6	23.6	26.8	24.9
500	20.5	22.3	23.2	23.3

A. Reciprocal space accuracy dependence

Alejandre *et al.* [30] found that the surface tension depended on the mesh refinement $\left|\mathbf{h}^{\max}\right|$ used to evaluate the long-range Coulombic interactions. To test the dependence of γ_p on $\left|\mathbf{h}^{\max}\right|$, we show results in Figure 4 for three different models as a function of $\left|\mathbf{h}^{\max}\right|$. From the figure, several trends become apparent. First, for both the TIP3P and TIP3P-Ew models, the long-time average of the surface tension depends significantly on the value $\left|\mathbf{h}^{\max}\right|$: the long-time average for a $5\times5\times20$ mesh ($\left|\mathbf{h}^{\max}\right| = 20$, rms accuracy 4.0×10^{-3}) is between 15 and 20 mN/m larger than for a $12\times12\times48$ mesh ($\left|\mathbf{h}^{\max}\right| = 50$, rms accuracy 10^{-4}). For larger values of $\left|\mathbf{h}^{\max}\right|$, there is no significant adjustment in the surface tension. Additionally, as found above, for a given value of the precision, there is little difference in the long-time average of the two models.

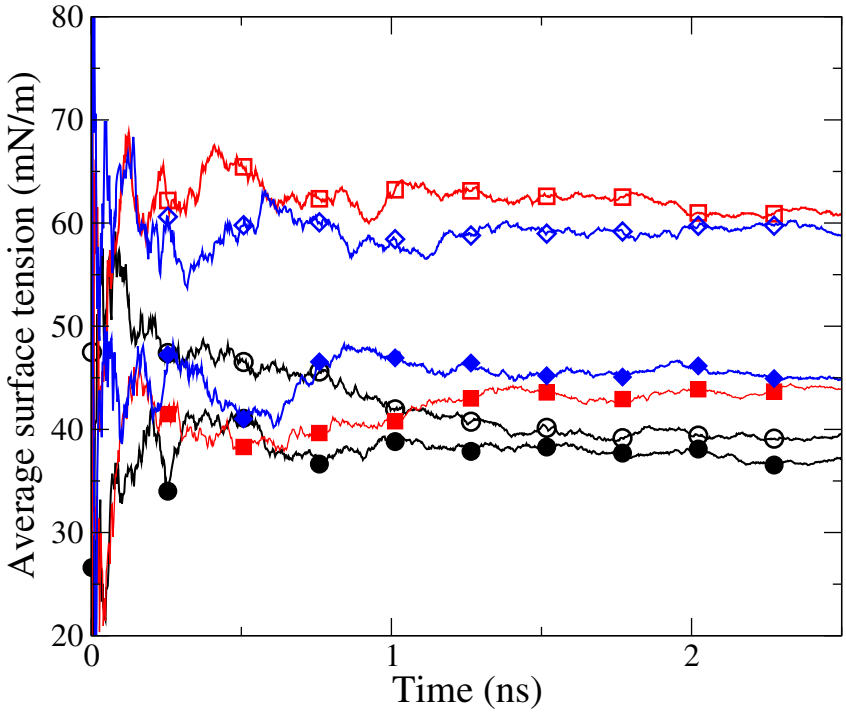

FIGURE 4. Equilibration of the surface tension of the SPC/E (black circles), TIP3P (red squares), and TIP3P-Ew (blue diamonds) models of water for rms $\mathbf{k}$ -space accuracies of 4×10^{-3} (dashed curves, open symbols) and 10^{-4} (solid curves, solid symbols).

For the SPC/E model, there is little difference between the equilibrium values for the $5\times5\times20$ mesh and the $12\times12\times48$ mesh. However, it takes approximately $1ns$ to achieve agreement between the two precision levels; before this, the less-refined mesh has a significantly greater surface tension. In the work of Alejandre *et al*., the total simulation time was only 0.375 ns. After 0.375 ns, the average surface tension from our simulation using the $5\times5\times20$ mesh was approximately $\gamma = 60$ mN/m, which closely corresponds with the result $\gamma = 60.6$ mN/m obtained in the earlier study. However, for long simulation times, the surface tension of the SPC/E model does not exhibit a strong dependence on the mesh size. Consequently, since the long-time averages are essentially equal within simulation uncertainty, we have used the finer $12\times12\times48$ mesh refinement for all of the simulations reported in this paper unless otherwise specified.

To further demonstrate that the accuracy of the reciprocal-space calculation does not affect the results for the surface tension of the SPC/E model when averaged over sufficiently long times, we performed simulations of the SPC/E model, as described above, with $\mathbf{k}$-space meshes of $5\times5\times20$, $12\times12\times48$, and $20\times20\times80$, corresponding to rms accuracies of 4×10^{-3}, 10^{-4}, and 10^{-5}, respectively. The results are shown in Figure 5 for the time ranges of 125 to 375 ps used by Alejandre *et al.*, and 1 to 2 ns used in the present work. Clearly, the results for the shorter time range (125 to 375 ps) have not converged: there is significant disagreement of as much as 20 mN/m, particularly at higher temperatures. However, for longer times, the results have converged, with the differences among the three mesh refinements being essentially within simulation uncertainty. It is interesting to note that the converged surface tensions of the different mesh refinements corresponds very closely to the profile obtained for the intermediate mesh refinement ($12\times12\times48$) after 375 ps. Our intermediate mesh refinement is comparable to

ir

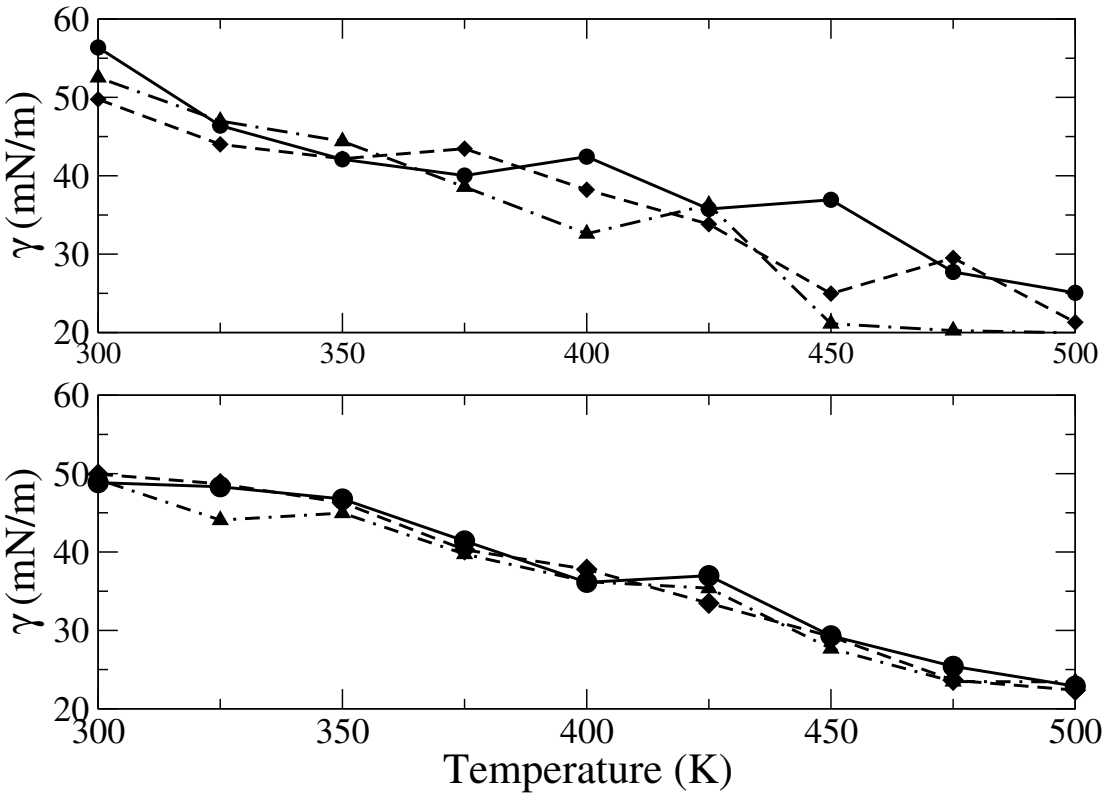

FIGURE 5. Surface tension for the SPC/E model of water as a function of mesh accuracy averaged over (top) 125-375 ps and (bottom) 1-2 ns for a $5\times5\times20$ mesh (circles), a $12\times12\times48$ mesh (diamonds), and a $20\times20\times80$ mesh (triangles).

IV. Capillary waves

The interfacial width Δ^2 was computed for various system sizes as described in Section II. The resulting plots of Δ^2 versus $\ln L$ were computed and the value of γ_w extracted using Eq. (6). A plot showing the data obtained for 300 K for the SPC/E water model is shown in Figure 6; the resulting values of γ_w for the two functional forms at 300 K, 400 K, and 500 K are shown in Table VI.

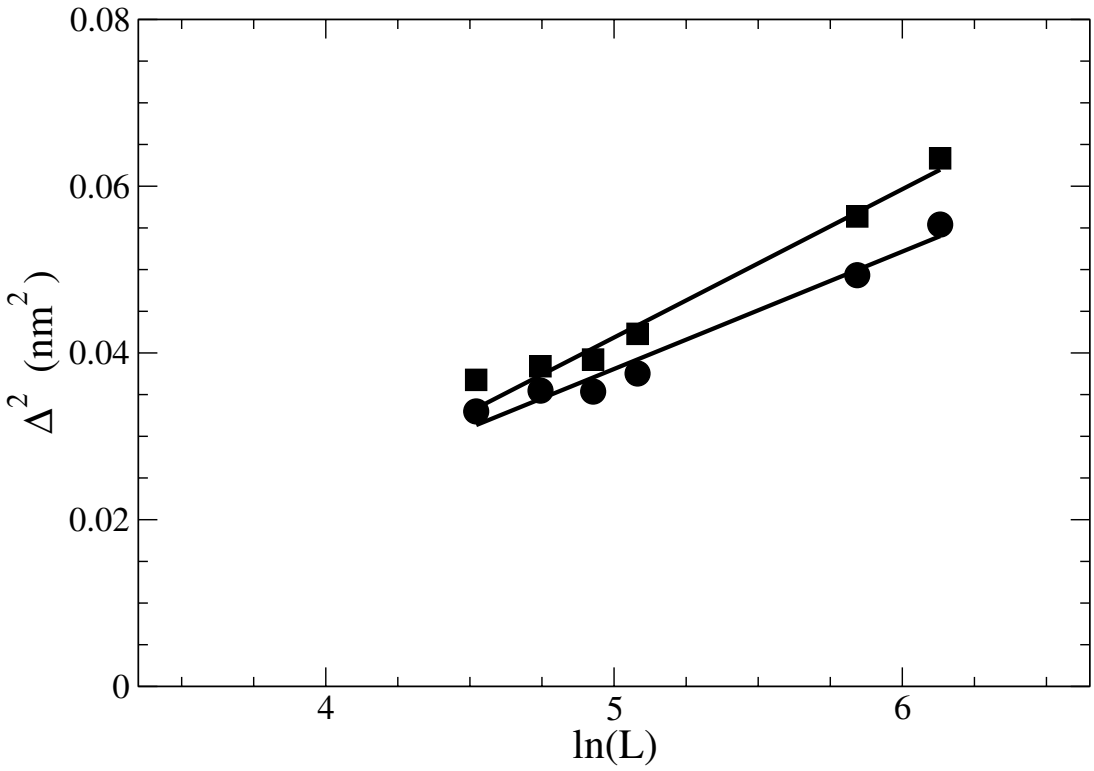

FIGURE 6. Regression fit of Eq. (6) for hyperbolic tangent (diamonds) and error function (circles) profiles at 300 K.

TABLE VI: Comparison of surface tension for the SPC/E model as a function of calculation method.

	Surface tension, mN/m		
T (K)	**Capillary-wave method**		**Pressure-integration method, γ_p**
	Erf, γ_w	**Tanh, γ_w**	
300	46.1	36.5	49.9
400	32.0	27.8	33.1
500	19.0	19.0	20.2

Our results indicate that at all three temperatures, good agreement between the pressure-integration estimate of the surface tension, γ_p, and the capillary-wave estimate of the surface tension, γ_w, is obtained only if the interfacial density profile is fit to an error function. There is substantial disagreement between the error function and hyperbolic tangent functions at lower temperatures: the hyperbolic tangent profile always yields lower estimates than the error function profile. However, the magnitude of the discrepancy between the two estimates of γ decreases the two narrows as temperature increases, and essentially vanishes at 500 K.

In addition to computing the surface tension γ_w, we can also compute an upper-bound estimate for the intrinsic interfacial width Δ_0 [7]. After obtaining the slope α_w and intercept β_w for the least-squares fit of Eq. (6), we assume that the parameter $B_0 = c\Delta_0$, where c is a constant to be specified. The intercept β_w and Δ_0 are then related by

$$\beta_w = \Delta_0^2 - \alpha_w \ln(c\Delta_0). \tag{14}$$

For values of c less than some threshold c^*, there is no real solution to (14); above the threshold, Δ_0 quickly decays as c increases. Thus, Δ_0 has a maximum at the threshold value $c = c^*$ where the imaginary part of the solution vanishes. For the system sizes under consideration, we find that the maximum intrinsic width $\Delta_0 \approx 0.8$ Å, 1.0 Å, and 1.5 Å at T = 300 K, 400 K, and 500 K.

Further evidence that fitting the density profile to an error function yields more accurate results than fitting to a hyperbolic tangent function can be seen by comparing the density fits themselves. As shown in Figure 7, although the two profiles are similar, the error function fit more closely adheres to simulation results than the hyperbolic tangent profile. Although the χ^2 parameter for both functional fits was relatively small, the coefficient for the hyperbolic tangent χ^2 parameter ($\sim 10^{-3}$) was approximately two orders of magnitude larger than the corresponding error function χ^2 parameter ($\sim 10^{-5}$). The greater accuracy of the error function is further seen by comparing the magnitude of the differences between the simulation results and the fitted functional profiles, as shown in Figure 8.

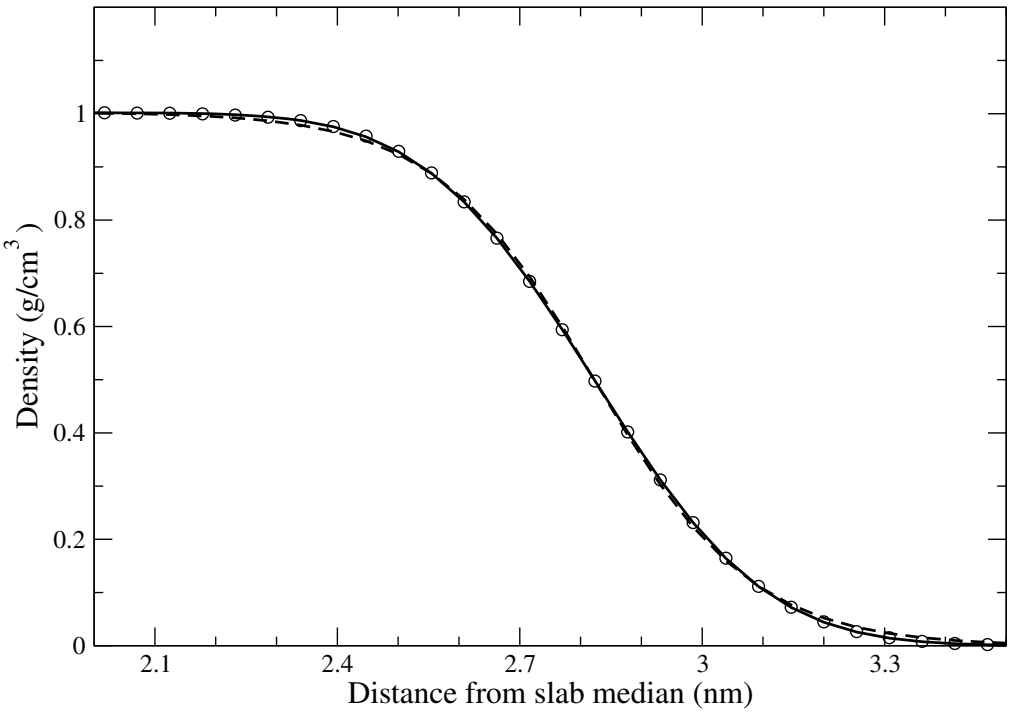

FIGURE 7. Density profile in the interfacial region for a slab of 400,000 SPC/E water molecules at $300K$ ($L_z = 46.0$ nm). Simulation results are shown as circles; fits to error function (solid curve) and hyperbolic tangent function (dashed curve) are also included.

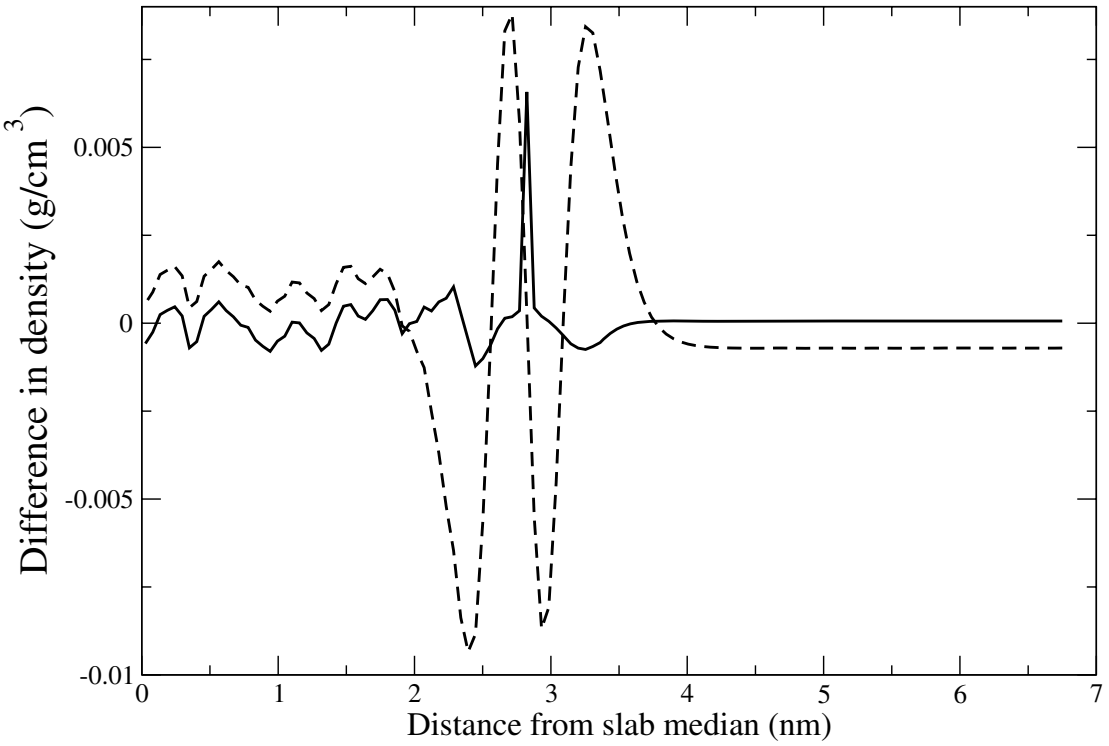

FIGURE 8. Difference between simulation results and the error function (solid curve) and the hyperbolic tangent (dashed curve) fits for the same system shown in Figure 2.

In addition to observing capillary wave behavior, the larger simulations can be used to study the effect of the interfacial area $L_\square^2$ on the surface tension computed using Eqs. (2) and (13). Since surface tension is an intensive property, it should be independent of $L_\square^2$. Although Eq. (13) is a slowly decreasing function of the parameter w_e, over the range of values for w_e and Δ^2 considered here, the tail contribution γ_{tail} varies by less than 2%. Thus, we expect the simulation values for the surface tension to remain constant, independent of $L_\square$. Examining the results for the surface tension γ_p (without tail correction) versus the system size N, as shown in Table VII, we find a slight increase in γ as $L_\square$ increases, although the results remain within simulation uncertainty, even for large values of $L_\square$.

TABLE VII: Surface tension for the SPC/E model as a function of system size at 300 K.

Interfacial length $L_{\square}$, Å	γ_p **(mN/m)**	γ **(mN/m)**
2.3	49.9	55.4
9.2	52.0	57.5
11.5	51.5	57.0
13.8	51.8	57.3
16.1	51.9	57.4
23.0	51.6	57.1
34.5	52.0	57.5
46.0	51.8	57.3

V. Conclusions

An investigation of surface tension as a function of temperature for a number of popular three- and four-point water models shows systematic disagreement between experimental and simulation results for all six models considered. All six models considered consistently underestimate the surface tension relative to the experimental data. The TIP4P-Ew model is closest to the experimental data, although not in quantitative agreement.

Examining the SPC/E model in greater detail, we have illustrated the importance of having a sufficiently long simulation time and sufficiently fine $\mathbf{k}$-space mesh: significant variations in the surface tension can result if the equilibration period is too short. In the SPC/E, TIP3P, and TIP3P-Ew models, significant variations were also observed if a small number of $\mathbf{k}$-space vectors are used.

The present study also includes the first in-depth study of the effect of large system sizes on interfacial properties, studying capillary waves for systems of up to 4×10^5 molecules. We have demonstrated that examining capillary waves at the liquid-vapor interface can be used to determine the surface tension of real fluids, and that the use of an error function profile offers

better results in comparison to experimental data than the use of a hyperbolic tangent profile. Finally, we note that the interfacial width Δ depends on the interfacial length $L_{\square}$, and that attempting to extract an interfacial width without taking into account the effect of capillary waves [22] is incorrect.

Acknowledgments

The authors would like to thank Steve Plimpton, Paul Crozier, and Amalie Frischknecht for helpful discussions in implementing the TIP4P model.

References

[1] F. P. Buff, R. A. Lovett, and F. H. Stillinger, Phys. Rev. Lett., 15, 621-623 (1965).

[2] J. Rowlinson and B. Widom, Molecular Theory of Capillarity, Clarendon Press, Oxford, 1982.

[3] S. Senapati and M. L. Berkowitz, Phys. Rev. Lett., 87, 176101 (2001).

[4] J. L. Rivera, C. McCabe, and P. T. Cummings, Phys. Rev. E, 67, 011603 (2003).

[5] M. Nijmeijer, A. Bakker, C. Bruin, and J. Sikkenk, J, Chem. Phys., 89, 3789 (1988).

[6] P. Adams and J. Henderson, Mol. Phys., 73, 1383 (1991).

[7] M.-D. Lacasse, G. S. Grest, and A. J. Levine, Phys. Rev. Lett., 80, 309-312 (1998).

[8] B. Guillot, J. Molec. Liq., 101, 219-260 (2002).

[9] J. L. Rivera, M. Predota, A. A. Chialvo, and P. T. Cummings, Chem. Phys. Lett., 357, 189-194 (2002).

[10] D. J. Price and C. L. Brooks III, J. Chem. Phys., 121, 10096-11003 (2004).

[11] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madera, T. J. Dick, G. L. Hura, and T. Head-Gordon, J. Chem. Phys., 120, 9665-9678 (2004).

[12] W. L. Jorgensen and J. Tirado-Rives, Proc. Natl. Acad. Sci., 102, 6665-6670 (2005).

[13] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, and J. Hermans In Intermolecular Forces, B. Pullman, Ed.; Reidel, Dordrecht, 1981; page 331.

[14] H. A. Stern, F. Rittner, B. J. Berne, and R. A. Friesner, J. Chem. Phys., 115, 2237-2251 (2001).

[15] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem., 91, 6269-6271 (1987).

[16] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys., 79, 926-935 (1983).

[17] K. Watanabe and M. L. Klein, Chem. Phys., 131, 157-167 (1989).

[18] M. W. Mahoney and W. L. Jorgensen, J. Chem. Phys., 112, 8910-8922 (2000).

[19] J. D. Bernal and R. H. Fowler, J. Chem. Phys., 1, 515 (1933).

[20] F. H. Stillinger and A. Rahman, J. Chem. Phys., 60, 1545-1557 (1974).

[21] S. W. Rick, S. J. Stuart, and B. J. Berne, J. Chem. Phys., 101, 6141-6156 (1994).

[22] I.-F. W. Kuo, C. J. Mundy, B. L. Eggimann, M. J. McGrath, J. I. Siepmann, B. Chen, J. Vieceli, and D. J. Tobias, J. Phys. Chem. B, 110, 3738-3746 (2006).
[23] P. Mark and L. Nilsson, J. Phys. Chem. A, 105, 9954-9960 (2001).
[24] P. Mark and L. Nilsson, J. Comp. Chem., 23, 1211-1219 (2002).
[25] G. J. Gittens, J. Coll. Interf. Sci., 30, 406-412 (1969).
[26] R. Cini, G. Loglio, and A. Ficaldi, J. Coll. Interf. Sci., 41, 287-298 (1972).
[27] J. J. Jasper, J. Phys. Chem. Ref. Data, 1, 841 (1972).
[28] K. Johansson and J. C. Eriksson, J. Coll. Interf. Sci., 40, 398-405 (1972).
[29] D. K. Thakur and K. Hickman, J. Coll. Interf. Sci., 50, 525-531 (1975).
[30] J. Alejandre, D. J. Tildesley, and G. A. Chapela, J. Chem. Phys., 102, 4574-4583 (1995).
[31] S. E. Feller, R. W. Pastor, A. Rojnuckarin, S. Bogusz, and B. R. Brooks, J. Chem. Phys., 100, 17011-17020 (1996).
[32] R. S. Taylor, L. X. Dang, and B. C. Garrett, J. Phys. Chem., 100, 11720-11725 (1996).
[33] L. X. Dang and T.-M. Chang, J. Chem. Phys., 106, 8149-8159 (1997).
[34] V. V. Zakharov, E. N. Brodskaya, and A. Laaksonen, J. Chem. Phys., 107, 10675-10683 (1997).
[35] V. V. Zakharov, E. N. Brodskaya, and A. Laaksonen, Mol. Phys., 95, 203-209 (1998).
[36] B. Shi, S. Sinha, and V. K. Dhir, J. Chem. Phys., page In press.
[37] D. M. Huang, P. L. Geissler, and D. Chandler, J. Phys. Chem. B, 105, 6704-6709 (2001).
[38] R. C. Tolman, J. Chem. Phys., 16, 758-774 (1948).
[39] J. G. Kirkwood and F. P. Bu¤, J. Chem. Phys., 17, 338-343 (1949).
[40] G. A. Chapela, G. Saville, S. M. Thompson, and J. S. Rowlinson, J. Chem. Soc. Farad. Trans., 73, 1133-1144 (1977).
[41] E. M. Blokhuis, D. Bedeaux, C. D. Holcomb, and J. A. Zollweg, Mol. Phys., 85, 665-669 (1995).
[42] J. S. Huang and W. W. Webb, J. Chem. Phys., 50, 3677-3693 (1969).
[43] D. Beysens and M. Robert, J. Chem. Phys., 87, 3056-3061 (1987).
[44] A. N. Semenov, Macromolecules, 27, 2732-2735 (1994).
[45] S. W. Sides, G. S. Grest, and M.-D. Lacasse, Phys. Rev. E, 60, 6708-6713 (1999).
[46] S. R. Durell, B. R. Brooks, and A. Ben-Naim, J. Phys. Chem., 98, 2198-2202 (1994).
[47] K. A. Feenstra, B. Hess, and H. J. C. Berendsen, J. Comp. Chem., 20, 786-798 (1999).
[48] S. J. Plimpton, J. Comp. Phys., 117, 1-9 (1995).

[49] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comp. Phys., 23, 327 (1977).

[50] R. W. Hockney and J. W. Eastwood, Computer Simulation Using Particles, Adam Hilger-IOP Publishing, Bristol, 1988.

[51] H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.*, 91, 461-466 (1989).

[52] In the case of four-point water models, while the contribution to the virial from the Lennard-Jones interactions can be computed using the more computationally efficient form $\sum_i \mathrm{r}_i \cdot \mathrm{f}_i$, the contribution to the virial from short-range electrostatic forces must be computed using the pairwise form $\sum_{i>j} \mathrm{r}_{ij} \cdot \mathrm{f}_{ij}$.

Chapter 4: Hamiltonian molecular dynamics for computational mechanicians and numerical analysts

Andrei Drăgănescu and Rich Lehoucq
Sandia National Laboratories, Albuquerque, New Mexico 87185

Abstract

Molecular dynamics (MD) is a computer simulation technique where the time evolution of a set of interacting atoms is followed by integrating their equations of motion. Our goal is to motivate the recipes of MD for practitioners and researchers in numerical analysis and computational mechanics. The vast majority of these practitioners and researchers work with continuum mechanics. In contrast, an atomistic method such as MD is both culturally and intellectually distinct. The recent interest in multiscale analysis, in particular, Atomistic-to-Continuum coupling necessitates a sophisticated understanding of MD, its goals, limitations and computation.

1. Introduction

Molecular dynamics (MD) is a computer simulation technique where the time evolution of a set of interacting atoms is followed by integrating their equations of motion. Our goal is to motivate the recipes of MD for practitioners and researchers in numerical analysis and computational mechanics. The vast majority of these practitioners and researchers work with continuum mechanics. In contrast, an atomistic method such as MD is both culturally and intellectually distinct. The recent interest in multiscale analysis, in particular, Atomistic-to-Continuum coupling necessitates a sophisticated understanding of MD, its goals, limitations and computation.

In contrast to continuum methods where accurate trajectories are of interest, MD is rarely concerned with accurate trajectories. In point of fact, accurate trajectories are not, in general, possible. Instead, quantities of interest are statistical averages computed during the sampling of

phase space. Primarily, our manuscript indicates why MD works from a mathematical perspective. We point out that our manuscript does not focus on geometric time integration issues, rather what are the relevant features of a time integrator that results in an acceptable sampling of phase space. Much of our discussion revolves around explaining in a mathematical fashion, both formally and informally, what the MD community of users and researchers has learned via dint of hard work and careful physical reasoning.

Key concepts. We identified in the literature a number of concepts that play key roles in assessing the numerical integration of a Hamiltonian system in order to ascertain whether phase space is sampled accordingly.

1. **Measure preservation:** This concept relates a map (transformation) $T : \mathrm{X} \to \mathrm{X}$ to a probability-measure μ on the space X : T is said to preserve the measure μ if for all sets $A \subset \mathrm{X}$, the preimage[1] of A has the same measure as A, that is: $\mu(T^{-1}(A)) = \mu(A)$. If T is differentiable, and μ is available in analytic form, then it is analytically verifiable whether T preserves μ or not (see Section 3).
2. **Ergodicity:** A measure-preserving map $T : \mathrm{X} \to \mathrm{X}$ is said to be ergodic if the only T-invariant[2] sets are trivial, in the sense that their measure is either zero or one. Alternatively, X cannot be split into disjoint sets of non-zero measure, each being invariant under T. In the statistical physics literature ergodicity is typically defined to be the property that for any statistical quantity Ψ, if we view the map T as a transition from one step to the next, the "time"-averages of T converge to the average of T over the space X, that is:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Psi(T^i x) = \int_{\mathrm{X}} \Psi(y) d\mu(y) \tag{1.1}$$

for almost all $x \in \mathrm{X}$. As usual, $T^i(x)$ means $\overbrace{T(T(...(T}^{i\,times}(x)...)$, with $T^0(x) = x$. Indeed, the property expressed in (1.1) is the desired one because it expresses the potential for computing the average on the right-hand side of (1.1) by means of the available time-averages on the left-hand side. However, verifying (1.1) may be extremely difficult,

[1]The preimage of A (through T) is the collection of points $x \in \mathrm{X}$ such that $T(x) \in A$.

[2]The set A is T-invariant if $T(A) \subset A$.

mathematicians regard (1.1) rather as a consequence of ergodicity, and not as ergodicity itself. We emphasize that verifying whether a map is ergodic (or not) may be nontrivial.

3. **Sensitive dependence on initial conditions (SDIC):** A map $T : \mathrm{X} \to \mathrm{X}$ is said to have SDIC if for every starting point $x \in \mathrm{X}$ we can find arbitrarily close points x', such that the positive orbit[3] starting at x' eventually diverges from the one starting at x. As a consequence, trajectories for a map having SDIC cannot be computed numerically for long periods of time in a classical sense (that is, point-to-point matching between exact value and computed value), because even the unavoidable round-off error will naturally be magnified, thus potentially rendering irrelevant computed values.
4. **Symplecticness:** A differential map $T : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ is symplectic if at each point $x \in \mathbb{R}^{2d}$ its Jacobian $T'(x)$ satisfies the algebraic equation (2.11). If an analytic expression is given for T, then the symplecticness of T is normally analytically verifiable.

There are various links between these concepts in the context of Hamiltonian systems. Firstly, the flow map of a Hamiltonian system is symplectic. A symplectic map from $\mathbb{R}^{2d}$ to itself preserves the Lebesgue measure. However, symplecticness is a more involved kind of measure preservation than just global measure preservation, as discussed in Section 2.3. In particular, the flow map preserves the projection of the Lebesgue measure on the constant energy-surface. Generically[4] measure-preserving maps generically have SDIC and are ergodic. The former property implies that we cannot expect to compute accurate trajectories of the Hamiltonian flow map. However, ergodicity allows us the extract statistical information from the same computations. The current state of affairs is that, if a discretization of the flow map $T \approx \Phi_{\Delta t}$ is symplectic, then time-averages of a statistic ϕ computed with T will converge to the desired spatial average as the time step $\Delta t \to 0$. Our thesis statement is that if T approximately conserves energy and preserves the projection onto S of the Lebesgue measure, then accurate statistical information is extracted from T.

The paper is organized as follows: In Section 2 we review the basic properties of Hamiltonian systems. We provide more mathematical background and language for the discussion above in

[3]The positive orbit of T starting at x is the set $\{T^n(x) : n \geq 0\}$.

[4]*Generically* is meant in a precise mathematical sense (see Section 3 and Appendix B).

Section 3. The last section is devoted to the question of what are sufficient conditions for numerical integrators to give correct answers, in light of the discussion from Section 3.

2. Background Information

This section quickly provides some background information on Hamiltonian dynamics. The reader is referred to [5] and the sources cited for an excellent introduction to Hamiltonian dynamics.

2.1. Mathematical Model

Our object of this discussion is the n-body problem for the regime of MD-simulations. As usual, if we denote the positions of n particles at time t particles by $\mathbf{q}(t) \in \mathbb{R}^{3n}$, then Newton's second law is

$$\mathbf{M}\ddot{\mathbf{q}} = \mathbf{F}(\mathbf{q}), \quad \mathbf{q}(0) = \mathbf{q}_0, \ \mathbf{p}(0) = \mathbf{p}_0 \tag{2.1}$$

where $\mathbf{M}$ is a diagonal matrix consisting of the particle masses. We assume that the force field is conservative, that is, there exists a *potential* $V : \mathbb{R}^{3n} \to \mathbb{R}$, such that

$$\mathbf{F}(\mathbf{q}) = -\nabla_{\mathbf{q}} V(\mathbf{q}).$$

An additional level of abstraction is achieved by introducing the *Hamiltonian* or *energy* functional

$$H(\mathbf{q}, \mathbf{p}) = \frac{1}{2}\mathbf{p}^T \mathbf{M}^{-1}\mathbf{p} + V(\mathbf{q}), \tag{2.2}$$

where the momenta of the particles is denoted by $\mathbf{p} = \mathbf{M}\dot{\mathbf{q}}$. This notation implies that the system (2.1) is equivalent to

$$\begin{cases} \dot{\mathbf{q}} &= \nabla_{\mathbf{p}} H(\mathbf{q}, \mathbf{p}) \\ \dot{\mathbf{p}} &= -\nabla_{\mathbf{q}} H(\mathbf{q}, \mathbf{p}) \end{cases}. \tag{2.3}$$

We define the matrix $\mathbf{J} = \begin{bmatrix} 0 & \mathbf{I}_d \\ -\mathbf{I}_d & 0 \end{bmatrix} \in \mathbb{R}^{2d \times 2d}$, and vector $\mathbf{z} = \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} \in \mathbb{R}^{2d}$, respectively, where $d = 3n$. Therefore, (4) can be rewritten as

$$\dot{\mathbf{z}} = \mathbf{J}\nabla_{\mathbf{z}} H(\mathbf{z}). \tag{2.4}$$

Under suitable assumptions[5] on (2.4), there is a unique solution for every initial condition, and we define the *flow map*

$$(\mathbf{z}_0, t) \mapsto \mathbf{z}(t) \overset{def}{=} \Phi_t(\mathbf{z}_0).$$

The flow map $t \mapsto \Phi_t(\mathbf{z})$ satisfies (2.3) with initial condition $\Phi_0(\mathbf{z}) = \mathbf{z}$, i.e.,

$$\frac{d}{dt}\Phi_t(\mathbf{z}) = \mathbf{J}\nabla_{\mathbf{z}} H(\Phi_t(\mathbf{z})). \tag{2.5}$$

A basic property satisfied by a solution of (2.4) is energy conservation. A simple calculation yields

$$\frac{d}{dt}H(\mathbf{z}) = (\nabla_{\mathbf{q}} H)^T \dot{\mathbf{q}} + (\nabla_{\mathbf{p}} H)^T \dot{\mathbf{p}} = (\nabla_{\mathbf{q}} H)^T \nabla_{\mathbf{p}} H + (\nabla_{\mathbf{p}} H)^T (-\nabla_{\mathbf{q}} H) = 0,$$

therefore $H(\mathbf{z})$ is constant of motion. A functional that is constant on solutions of (2.4) is called a *first integral*, therefore the Hamiltonian is a first integral. However, energy conservation is only necessary for Hamiltonian dynamics. The necessary and sufficient condition of Hamiltonian dynamics is that the flow (or trajectory) is *symplectic*, a concept discussed in Section 2.3. In other words, Hamiltonian dynamics occurs if and only if the flow map (2.5) is symplectic[6].

2.2. Phase space conservation

The Hamiltonian set of equations (2.3) gives rise to an important conservation law that has an analogy in continuum mechanics.

We first introduce some terminology. The space given by the vectors $\mathbf{z}$ defined by (2.4) is called phase space. Phase points are parameterized by time t and so the flow (2.5) lies in phase space. Note that each phase point defines a unique trajectory because the associated $\mathbf{z}$ represent a solution of Hamilton's equations and are uniquely determined. Therefore each phase point defines a state of a mechanical system. We can consider a continuum of points in phase space occupying a volume V in phase space. This notion of phase space volume is well considered because in reality, we cannot identify the precise positions and momentums of the Hamiltonian

[5] for example, the Lipschitz property

[6] The precise statement discusses locally Hamiltonian flow. See Theorem 2.6 [4, pp. 173-174].

system. Instead, each of the points in V represents a possible state of the system. The collection of states is called an *ensemble* and represents a useful idealization. The following theorem proves central to our discussion, and provides an extremely useful analogy with incompressible fluid flow. The theorem describes the motion of an ensemble in phase space.

Theorem 2.1 (Liouville) *Let the Hamiltonian set of equations (4) be given. If* Ω *denotes a region, or volume, of phase space, denote* $\Omega_t = \Phi_t(\Omega)$, *and by* $V(t) = vol(\Omega_t)$. *Then*

$$\frac{dV(t)}{dt} = 0. \tag{2.6}$$

Proof. We have

$$\frac{dV}{dt} = \frac{d}{dt}\int_{\Omega_t} d\mathbf{z} = -\int_{\partial\Omega_t} \dot{\mathbf{z}} \cdot \mathbf{n} dS = -\int_{\Omega_t} \nabla_{\mathbf{z}} \cdot \dot{\mathbf{z}}\, d\mathbf{z}.$$

Liouville's theorem is established because

$$\int_{\Omega_t} \nabla_{\mathbf{z}} \cdot \dot{\mathbf{z}}\, dz = \int_{\Omega_t} \nabla_{\mathbf{z}} \cdot \mathbf{J}\nabla_{\mathbf{z}} H(\mathbf{z})\, dz = \int_{\Omega_t} (\nabla_{\mathbf{q}} \cdot \nabla_{\mathbf{p}} - \nabla_{\mathbf{p}} \cdot \nabla_{\mathbf{q}}) H(\mathbf{q},\mathbf{p})\, d\mathbf{q}d\mathbf{p} = 0, \tag{2.7}$$

where we used Hamiltonian set of equations (2.4). □

Because the volume V is arbitrary, phase space flow is incompressible, i.e.

$$\nabla_{\mathbf{z}} \cdot \dot{\mathbf{z}} = 0.$$

If $\rho \equiv \rho(\mathbf{z})$ is the density of system points in phase space per unit volume, then $\rho\Delta V$ is the amount of system points in an infinitesimal volume ΔV. Because the amount of system points remains constant, we have

$$0 = \frac{d}{dt}\rho\Delta V = \Delta V \frac{d}{dt}\rho + \rho \frac{d}{dt}\Delta V = \Delta V \frac{d}{dt}\rho = \Delta V(\frac{\partial}{\partial t}\rho + \dot{\mathbf{z}} \cdot \nabla_{\mathbf{z}}\rho),$$

where we used Liouville's Theorem in going from the second to third equality. If we divide through by ΔV, and use the incompressibility of phase space flow, then

$$\frac{\partial}{\partial t}\rho + \dot{\mathbf{z}} \cdot \nabla_{\mathbf{z}}\rho + \rho\nabla_{\mathbf{z}} \cdot \dot{\mathbf{z}} = \frac{\partial \rho}{\partial t} + \nabla_{\mathbf{z}} \cdot (\rho\dot{\mathbf{z}}).$$

Therefore we have the Lagrangian and Eulerian version of the continuity equation

$$\frac{d}{dt}\rho = \frac{\partial \rho}{\partial t} + \nabla_{\mathbf{z}} \cdot (\rho\dot{\mathbf{z}}) = 0, \tag{2.8}$$

for phase space density. In words, the amount of system points in any volume ΔV moving with the fluid remains constant.

In reality, as explained before Theorem 2.1, precise positions and momentums of the Hamiltonian system are not available. And so $\mathbf{z}$ can be identified with a random variable with distribution ρ that evolves in time via (2.8). If we define statistical *equilibrium* as the distribution, or measure, where ρ is constant at every point in phase space, then (2.8) gives that incompressible phase space flow is a necessary and sufficient condition for statistical equilibrium.

Hamiltonian set of equations (2.4) imply that we can rewrite (2.8) as

$$\frac{\partial \rho}{\partial t} = -\nabla_{\mathbf{z}} \cdot (\rho \dot{\mathbf{z}}) = -\dot{\mathbf{z}} \cdot \nabla_{\mathbf{z}} \rho = \mathsf{L}\rho(\mathbf{z})$$

where $\mathsf{L} \stackrel{def}{=} \nabla_{\mathbf{p}} H \cdot \nabla_{\mathbf{q}} - \nabla_{\mathbf{q}} H \cdot \nabla_{\mathbf{p}}$ is called the Liouville operator. Because the operator is linear, we have the formal solution

$$\rho(\mathbf{z}, t) = e^{-\mathsf{L}t} \rho(\mathbf{z}, 0).$$

The Liouville operator is skew-symmetric, and so the eigenvalues all lie on the imaginary axis. Hence the phase space density $\rho(\mathbf{z}, t)$ oscillates and does not have a limit as t increases. The formal solution also allows us to specify an important property that any density ρ must satisfy. Using (2.8) gives that

$$0 = \frac{d\rho}{dt} = -\mathsf{L}\rho = \nabla_{\mathbf{p}} H \cdot \nabla_{\mathbf{q}} \rho - \nabla_{\mathbf{q}} H \cdot \nabla_{\mathbf{p}} \rho$$

so that ρ must commute with the Hamiltonian. If ρ is selected equal to a function of H, then this function of ρ commutes with H.

2.3. Symplectic Flow

By differentiating (2.5) with respect to $\mathbf{z}$ we conclude that the derivative with respect to the initial condition, or Jacobian

$$\mathbf{F}_t(\mathbf{z}) = \frac{\partial}{\partial \mathbf{z}} \Phi_t(\mathbf{z}) \tag{2.9}$$

satisfies the linear differential equation

$$\frac{d}{dt} \mathbf{F}_t(\mathbf{z}) = \mathbf{J} \partial^2_{\mathbf{zz}} H(\Phi_t(\mathbf{z})) \mathbf{F}_t(\mathbf{z}), \tag{2.10}$$

with $\mathbf{F}_0(\mathbf{z}) = \mathbf{I}_{2d}$. A differentiable map $\mathbf{T}: \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ is called symplectic if the Jacobian satisfies the matrix relation

$$\mathbf{A}^T \mathbf{J} \mathbf{A} = \mathbf{J}, \quad \mathbf{A} \equiv \frac{\partial}{\partial \mathbf{z}} \mathbf{T} \tag{2.11}$$

Theorem 2.2. *The Jacobian of the flow map (given by (1)) is a symplectic map, i.e.*

$$\left(\mathbf{F}_t(\mathbf{z})\right)^T \mathbf{J} \mathbf{F}_t(\mathbf{z}) = \mathbf{J} \quad \forall t \in , \forall \mathbf{z} \in \mathbb{R}^{2d}, \tag{2.12}$$

and so the flow map Φ_t is symplectic.

Proof. For example, see [5, p. 54]. □

It follows that a symplectic map preserves the Lebesgue measure. However, symplecticness is much more than measure-preservation. In order to describe symplecticness in geometrical terms we extract the following discussion from [5, pp 56-61]. Consider the bilinear form

$$\Omega(\mathbf{y}, \mathbf{x}) \stackrel{def}{=\!=} -\mathbf{y}^T \mathbf{J} \mathbf{x}, \text{ for } \mathbf{y}, \mathbf{x} \in \mathbb{R}^{2d}. \tag{2.13}$$

If $d = 1$, then

$$\Omega(\mathbf{y}, \mathbf{x}) = \mathbf{y}_1^T \mathbf{x}_2 - \mathbf{y}_2^T \mathbf{x}_1, \quad \mathbf{x} \equiv \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}, \quad \mathbf{y} \equiv \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \quad \mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^3. \tag{2.14}$$

The right-hand side expression from (2.14) represents the signed area of the parallelogram spanned by the vectors $\mathbf{x}$ and $\mathbf{y}$. In general, we have that

$$\Omega(\mathbf{y}, \mathbf{x}) = \sum_{i=1}^{d} \mathbf{y}_i^T \mathbf{x}_{d+i} - \mathbf{y}_{d+i}^T \mathbf{x}_i, \quad \mathbf{y}, \mathbf{x} \in \mathbb{R}^{2d}. \tag{2.15}$$

In words, $\Omega(\mathbf{y}, \mathbf{x})$ is obtained by summing up the signed areas of the projections on the $(\mathbf{q}_i, \mathbf{p}_i)$-planes of the parallelogram spanned by $\mathbf{y}$ and $\mathbf{x}$. It is the preservation of **this** quantity that characterizes symplectic maps. Preservation of the Ω-form is more restrictive than measure-preservation in full-space, e.g. that given by (2.8). In other words, the incompressibility of phase space flow is a weaker condition than that of symplectic flow.

3. Computing Statistics, Ergodicity And Chaotic Behavior

The flow map associated with a Hamiltonian is a family of measure-preserving transformations in phase-space. Generically, measure-preserving maps are both chaotic and ergodic. While the former property inhibits the possibility of following trajectories for a long time, the latter allows for retrieving global statistical information by following almost all trajectories for infinite time. Symplecticness is intimately related both to measure-preservation and to the property of a system to be Hamiltonian, therefore it plays a key role in our discussion of numerical integration for Hamiltonian ordinary differential equations. In this section we attempt to introduce the mathematical framework in which the ideas above reside.

A critical component in our exposition is the idea that the preserved measure of interest is a measure on the constant-energy surface, with the density given by (3.5). A statement common to the statistical mechanics and MD literature is that *all states on the energy surface are equally probable*. If the surface under discussion were, say, the unit cube in $\mathbb{R}^{2d}$, then a well-intended mathematician would read in this statement that the underlying probability is the Lebesgue measure[7]. That is, the probability density with respect to the natural measure on the set of interest (here, the cube) has a constant value of 1. Since every subsurface of $\mathbb{R}^{2d}$ is naturally equipped with a measure inherited from the Lebesgue measure of $\mathbb{R}^{2d}$, namely the one given by the *volume-form*[8], one may be tempted to read in the aforementioned statement that the probability density with respect to this inherited measure on the surface has a constant value of 1. As we will see below, this is not the case here. We refer the unfamiliar reader to [8] for a definition of the volume-form.

As mentioned in Section 1, given a probability space $(\mathsf{X}, \Sigma, \mu)$[9], a map $T : \mathrm{X} \to \mathrm{X}$ is said to preserve the measure μ if

$$\mu(T^{-1}(A)) = \mu(A), \forall A \subset \mathsf{X}, A \text{ measurable}, \tag{3.1}$$

[7] An ill-intended mathematician might say: 'Of course they are equally probable, each individual state has probability zero'.

[8] For example, the inherited measure of the sphere of radius r regarded as a subsurface of $\mathbb{R}^3$ is $4\pi r^2$, that is, the known surface-area of the sphere.

[9] Σ is a σ-algebra of subsets of X, and $\mu : \Sigma \to [0, \infty)$ is a probability measure.

where $T^{-1}(A)$ is the preimage of A, that is, $T^{-1}(A) = \{x \in \mathsf{X} : T(x) \in A\}$. If T is one-to-one, onto, and T^{-1} is measurable, then (3.1) is equivalent to

$$\mu(T(A)) = \mu(A), \forall A \subset \mathsf{X}, A \text{ measurable}, \tag{3.2}$$

where $T(A)$ is the image of A. If X is equipped with a differentiable structure as well, and if the probability measure is given by a density ρ with respect to a volume-form $d\sigma$, i.e. $\mu = \rho d\sigma$, then the change-of-variable formula (see [8]) implies that

$$\int_A \rho(x)\, d\sigma(x) = \mu(A) = \mu(T(A)) = \int_{T(A)} \rho(y)\, d\sigma(y) \overset{y=T(x)}{=} \int_A \left|\det T'(x)\right| \rho(T(x))\, d\sigma(x),$$

where $T'(x)$ is the Jacobian matrix of T at x. Since the above holds for every measurable set A, we conclude that T preserves the measure μ if and only if the integrants above are equal a.e., that is,

$$\det T'(x) = \frac{\rho(x)}{\rho(T(x))}, \text{ a.e. } [d\sigma]. \tag{3.3}$$

For example, if $\mathsf{X} = \mathbb{R}^m$ and $\rho \equiv 1$, then μ is the Lebesgue measure, and T preserves μ if and only if $|\det T'(x)| = 1$ a.e. We should point out that the definition of measure-preserving maps does not require them to be one-to-one, however, since the flow-map of a Hamiltonian is one-to-one and onto, we restrict the discussion to the latter case.

Let $\mathsf{S}_E = \{z \in \mathbb{R}^{2d} : H(z) = E\}$ be the constant-energy surface at level energy level E. The fact that the Hamiltonian flow Φ_t preserves energy may be restated as

$$\Phi_t(\mathsf{S}_E) \subseteq \mathsf{S}_E.$$

Since the dynamics is restricted to the surface S_E, a crucial question is to identify a nontrivial measure **on** S_E that is preserved by Φ_t. This measure is provided both in the statistical mechanics and MD literatures (e.g., [7, 9]) and the dynamical systems literature [3], but the notation differs between the two. The statistical mechanics and MD communities describe the invariant measure by

$$\mu(A) = \int_A \delta(H(X) - E) dX, \text{ for } A \subseteq \mathsf{S}_E, \tag{3.4}$$

with δ being the Dirac delta-distribution. Mathematicians might prefer the description of μ as provided in [3] because a measure is defined explicitly on the surface S_E. If we denote by $d\sigma_{2d-1}$ the $(2d-1)$-dimensional volume-element of S_E, then (3.4) can be written as

$$\mu(A) = \int_A \frac{d\sigma_{2d-1}}{\| \nabla_z H \|}, \tag{3.5}$$

where $\| \nabla_z H \|$ is the Euclidian norm of the gradient of the Hamiltonian. An excellent intuitive explanation of why μ is preserved by Φ_t can be found in [7], Chapter 8. For the sake of completeness we provide a formal mathematical argument in Appendix A. We should stress that the formal proof in Appendix A relies on verifying an equation similar to (3.3), however, on the left-hand side we have the Jacobian of the restriction of Φ_t to the hyperplane tangent to S_E. Assuming that $\mu(\mathsf{S}_E) < \infty$ we rescale μ in order to produce an invariant probability measure

$$\mu(A) \overset{redefine}{=} \mu(A)/\mu(\mathsf{S}_E). \tag{3.6}$$

We will return to this probability shortly.

We used the term *generically* in the introductory paragraph of this section in a **topological sense**. Technically, a property is called *generic* if the set of all elements satisfying that property is "rich" in the sense that it contains a countable intersection of dense open sets. To the reader who is less familiar with the subject this may seem an odd way of a regarding a set as rich. We restrict our comments on this issue to two items. Firstly, if a property is generic then it is satisfied by a dense set of elements[10]. Secondly, if properties (P_1) and (P_2) are generic then so is the property $(P_1 \wedge P_2)$. For example, the property of a number in $[0,1]$ to be rational is not generic, but the property of being irrational is. This is an example where topological genericity coincides with the measure-theoretical genericity, since almost all numbers in $[0,1]$ are irrational. However, there are examples of generic sets that have zero Lebesgue measure (see Example B.1). We provide exact definitions and basic results that justify the use of topological genericity in Appendix B.

The topological context of interest for the present discussion is the metric space $M(\mathsf{S}_E, \mu)$ of measure-preserving maps from (S_E, μ) onto itself. We define the *uniform metric* on $M(\mathsf{S}_E, \mu)$:

$$\delta(T_1, T_2) \overset{def}{=} \operatorname{ess\,sup}_{x \in \mathsf{S}_E} d(T_1(x), T_2(x)). \tag{3.7}$$

[10]A set A is called dense in a metric space T if for every element in T we can find arbitrarily close elements that belong to A.

Recall that (3.7) implies that

$$d(T_1(x), T_2(x)) \le \delta(T_1, T_2) \text{ a.e. } [\mu].$$

Two properties of interest to us are generic in this metric space: *ergodicity*, and the property of being *chaotic* (see [1] and Theorem B.1). There are several definitions of what it means for a map to be chaotic, nevertheless, they all have in common *sensitive dependence on initial conditions* (SDIC). A map T has SDIC if small perturbations in the initial state are magnified by propagation through T. A more precise statement is that for each initial condition, there exist arbitrary close initial conditions whose orbits eventually diverge from that of x_0. A linear map T has SDIC if it has an expanding direction, i.e., an eigenvector y associated with an eigenvalue λ with $\lambda > 1$[11]. For a nonlinear map, a strong indication of SDIC is when it's linearization $T'(x)$ has an expanding direction at each point x. If T is measure-preserving in $\mathbb{R}^{2d}$, then $\det(T'(x)) = 1$, therefore

$$\prod_{\lambda \in \sigma(T'(x))} |\lambda| = 1.$$

Hence, unless all eigenvalues in $\sigma(T'(x))$ have absolute value 1, we will find at least one $\lambda \in \sigma(T'(x))$ with $\lambda > 1$. Loosely speaking, the linearization of a measure-preserving map has an expanding direction around each point. We remark that this heuristic argument is not meant to replace the rigorous result expressed in Theorem B.1(ii), however, it should suffice to convince the reader that numerically following individual orbits of a measure-preserving map such as Φ_t for a long time period is a futile task.

As previously mentioned, the value of long-time numerical solutions for measure-preserving maps resides in retrieving statistical information. So what does it mean to retrieve statistics? The basic assumption is that the flow Φ_t correctly samples the constant-energy surface S_E. Thus, it is natural to define empirically a probability on the measurable subsets of S_E; namely the probability of a set A is the fraction of time that an orbit spends in A:

$$\nu(A) = \lim_{t \to \infty} \frac{\lambda(\{s \in [0,t] : \Phi_s(z_0) \in A\})}{t},$$

[11] For any x_0 and ε we have $T^n(x_0) - T^n(x_0 + \varepsilon y) = \varepsilon T^n(y) = \varepsilon \lambda^n y \to_{n \to \infty} \infty$.

where λ is the Lebesgue measure on the real line. Given a statistic ψ, i.e., an integrable function on S_E, we expect the long-term averaged value of ψ on the orbit to be equal to the spatial average value of ψ with respect to ν:

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \psi(\Phi_s(z_0))ds = \int_{\mathsf{S}_E}\psi(z)d\nu(z). \tag{3.8}$$

The question remains of whether such a probability measure ν can be defined independently of the starting point of the orbit. In particular, the average on the right-hand side of (3.8) should be the same, independent on the moment in time when we start averaging on the left-hand side of (3.8). This shows the necessity for the measure ν to be preserved by the flow map Φ_τ. However, we have already encountered such a (nontrivial) measure, namely μ defined in (3.6). Retrieving statistical information in MD is based on the fact that the *ergodic hypothesis* holds, that is, the equality (3.8) is valid with μ in place of ν, for almost all points $z_0 \in \mathsf{S}_E$:

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \psi(\Phi_s(z_0))ds = \int_{\mathsf{S}_E}\psi(z)\mu(z). \tag{3.9}$$

The equality (3.9) is typically regarded in the MD literature as the **definition** of ergodicity for the flow map. In general, (3.9) is not directly verifiable. Finding hypotheses under which (3.9) holds forms the object of ergodic theory. According to ergodic theory, given a probability space (X,Σ,μ), a map $T:\mathrm{X}\to\mathrm{X}$ is defined to be ergodic (with respect to the probability measure μ) if it is measure-preserving, and if $T^{-1}(A)\subseteq A$ implies $\mu(A)=0$ or $\mu(A)=1$. Perhaps the most celebrated result in ergodic theory is Birkhoff's strong ergodic theorem, which states that the averages of a measurable function ψ on sets of consequent iterates of an ergodic map converge to the spatial average of ψ for almost every orbit:

Theorem 3.1 (Birkhoff's strong ergodic theorem) *Let T be an ergodic transformation of the probability space (X,Σ,μ), and let ψ be integrable. Then for almost every $x\in\mathsf{X}$*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\psi\circ T^j(x) = \int_{\mathsf{X}} Td\mu. \tag{3.10}$$

For $T=\Phi_{\Delta t}$ ($\Delta t>0$) the statement (3.10) reads

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\psi(\Phi_{j\Delta t}(x)) = \int_{\mathsf{S}_E} Td\mu, \tag{3.11}$$

provided $\Phi_{\Delta t}$ is ergodic. Note that the expression on the left-hand side of (3.11) is an approximation of the integral on the left-hand side of (3.9), with exact integration being replaced by the trapezoidal rule. The question of why $\Phi_{\Delta t}$ is ergodic cannot be answered in general. With the exception of simple examples, (see Examples B.2 and B.3) we have to rely on the fact that generic measure-preserving maps are ergodic (see Theorem B.1(i)). However, there is one instance when we know a priori that $\Phi_{\Delta t}$ is not ergodic on S_E, namely when there is an additional first integral G, independent of H, i.e., $\nabla_z H(z)$ and $\nabla_z G(z)$ are linearly independent for all $z \in \mathsf{S}_E$. In this case sets of the type $\{z \in \mathsf{S}_E : l < G(z) < r\}$ may have nontrivial measure, and are invariant under the flow. It is easy to see that in fact the flow is restricted to a subsurface of S_E of the form $\{z \in \mathsf{S}_E : G(z) \equiv C\}$, therefore one orbit cannot sample the whole of S_E. However, in this case, an invariant measure that plays the same role as μ can be explicitly defined on the $(2d-2)$-dimensional surface

$$\mathsf{S}_{E,C} = \{z \in \mathbb{R}^{2d} : H(z) = E \text{ and } G(z) = C\}.$$

In conclusion, the main reason for which the continuous flow Φ_t is expected (but not guaranteed) to sample correctly the constant-energy surface S_E lies in the fact that is preserves the measure μ.

Acknowledgments

We thank Bob Skeel (Purdue University), Don Estep (CSU), Paul Crozier, Mike Parks, Steve Plimpton, Mark Sears, Mark Stevens, Aidan Thompson for many helpful discussions.

References

[1] Steve Alpern and V. S. Prasad. Typical dynamics of volume preserving homeomorphisms, volume 139 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 2000.

[2] V. I. Arnold. Mathematical Methods of Classical mechanics. Number 60 in Springer Graduate Texts. Springer Verlag, second edition, 1989.

[3] V. I. Arnold, editor. Dynamical systems. III, volume 3 of Encyclopaedia of Mathematical Sciences. Springer-Verlag, Berlin, second edition, 1993. Mathematical aspects of classical and celestial mechanics, A translation of Current problems in mathematics. Fundamental directions, Vol. 3 (Russian), Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1985 [MR0833508 (87i:58151)], Translation by A. Iacob, Translation edited by V. I. Arnold.

[4] Ernst Hairer, Christian Lubich, and Gerhard Wanner. Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations. Springer-Verlag, Berlin, 2002.

[5] Benedict Leimkuhler and Sebastian Reich. Simulating Hamiltonian dynamics, volume 14 of Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 2004.

[6] J. C. Oxtoby and S. M. Ulam. Measure-preserving homeomorphisms and metrical transitivity. Ann. of Math. (2), 42:874-920, 1941.

[7] L. E. Reichl. A Modern Course in Statistical Physics. University of Texas Press. W. A. Benjamin, Inc., New York-Amsterdam, Austin, TX, 1988.

[8] Michael Spivak. Calculus on manifolds. A modern approach to classical theorems of advanced calculus. W. A. Benjamin, Inc., New York-Amsterdam, 1965.

[9] Mark E. Tuckerman and Glenn. J. Martyna. Understanding molecular dynamics: techniques and applications. J. Phys. Chem. B, 104:159-178, 2000.

Appendix A: An Invariant Measure On The Constant-Energy Surface

Lemma A.1 (invariant measure on the surface). *The measure* μ *defined in (3.5) is preserved by the flow restricted to* S_E.

Proof. Let $z_0 \in \mathsf{S}_E$. Since S_E is flow-invariant, the Jacobian $\partial_z \Phi_t(z_0)$ maps the tangent plane $\mathsf{T}_{z_0}\mathsf{S}_E$ at z_0 to the surface S_E onto the tangent plane $\mathsf{T}_{\Phi_t(z_0)}\mathsf{S}_E$ at $\Phi_t(z_0)$. If we choose an orthonormal basis in each of the two planes, the relevant quantity for volume preservation on the surface S_E is

$$\det\left(\partial_z \Phi_t(z_0)|_{\mathsf{T}_{z_0}\mathsf{S}_E}\right).$$

Let $N(t) = \nabla_z H(\Phi_t(z_0))$, be the unnormalized vector orthogonal to S_E at the point $\Phi_t(z_0)$, and $n(t) = \nabla_z H(\Phi_t(z_0))/\|\nabla_z H(\Phi_t(z_0))\|$ be the unit normal vector. We have

$$\begin{aligned}
&\partial_t\left(\partial_z \Phi_t(z_0)\cdot n(0), N(t)\right)\\
&= \left(\partial_t(\partial_z \Phi_t(z_0))\cdot n(0), N(t)\right) + \left(\partial_z \Phi_t(z_0)\cdot n(0), \partial_t N(t)\right)\\
&\overset{(2.5),(2.10)}{=} \left(\mathbf{J}\cdot\partial^2_{zz} H(\Phi_t(z_0))\cdot\partial_z \Phi_t(z_0)\cdot n(0), \nabla_z H(\Phi_t(z_0))\right) +\\
&\qquad \left(\partial_z \Phi_t(z_0)\cdot n(0), \partial^2_{zz} H(\Phi_t(z_0)\cdot\mathbf{J}\cdot\nabla_z H(\Phi_t(z_0))\right)\\
&\overset{\mathbf{J}^T=-\mathbf{J}}{=} 0.
\end{aligned}$$

Since $N(t) = \|N(t)\|\cdot n(t)$ and $\partial_z \Phi_t(z_0)|_{t=0} = I$ we obtain

$$\|N(0)\| = \left(\partial_z \Phi_t(z_0)\cdot n(0), n(t)\right)\cdot\|N(t)\|. \tag{A.1}$$

Therefore

$$\det(\partial_z \Phi_t(z_0)) = \det\left(\partial_z \Phi_t(z_0)|_{\mathsf{T}_{z_0}\mathsf{S}_E}\right)\cdot\left(\partial_z \Phi_t(z_0)\cdot n(0), n(t)\right) = 1. \tag{A.2}$$

This together with (A.1) implies

$$\frac{\det\left(\partial_z \Phi_t(z_0)|_{\mathsf{T}_{z_0}\mathsf{S}_E}\right)}{\|\nabla_z H(\Phi_t(z_0))\|} = \frac{1}{\|\nabla_z H(z_0)\|}, \tag{A.3}$$

which, by (3.3), completes the proof. □

It should also be mentioned that an invariant measure similar to μ can be given explicitly when more than one first integral is known.

Lemma A.2. *Assume* $H_1 = H, H_2, ..., H_k$ *are constants of the motion that are linearly independent, and denoted by* $\mathsf{S}_{E_1,...,E_k} = \cap_{i=1}^{k}\{z \in \mathbb{R}^{2d} : H_i(z) = E_i\}$ *the submanifold on which the flow is restricted. We define the measure*

$$\nu(A) = \int_A \frac{d\sigma_{2d-k}}{vol_k(\nabla_z H_1, ..., \nabla_z H_k)}, \text{ for } A \subset \mathsf{S}_{E_1,...,E_k}. \tag{A.4}$$

where $vol_k(\nabla_z H_1,...,\nabla_z H_k)$ *is the* k*-dimensional volume of the parallelepiped formed by the vectors* $\nabla_z H_1,...,\nabla_z H_k$. *Then* Φ_t *preserves* ν.

Appendix B: Topological considerations

Definition B.1 A map $\phi: X \to X$ is said to have *sensitive dependence on initial conditions* (SDIC) if there exists $D > 0$, such that $\forall x \in \mathsf{X}$ and $\varepsilon > 0$ there exists $x' \in X$ with $d(x, x') \le \varepsilon$ and $n > 0$ such that $d(\phi^n x, \phi^n x') > D$.

In words, given any orbit O, there are orbits that start arbitrarily close but eventually diverge away from O.

Theorem B.1 (genericity) *Let $X = [0,1]^N$ be the unit cube equipped with the Lebesgue measure and the usual Euclidian metric. Then*

i. the set of ergodic maps is generic in $\mathsf{M}(X, \mu)$;
ii. the set of chaotic maps is generic in $\mathsf{M}(X, \mu)$.

Part (i) of Theorem B.1 is due to Oxtoby and Ulam [6] and is the main result of their theory. A proof of the statements in Theorem 2, nearby results, and further discussions of the subject can be found in [1]. We should point out that from the fact that generically measure-preserving maps are chaotic and ergodic, we cannot automatically come to the same conclusion if additional information is known about the map. In particular no conclusion can be drawn about an individual map other than that there exists a nearby map that is both ergodic and chaotic. An extreme example is the identity map: it preserves any measure, but it is neither chaotic nor ergodic. (extension to compact manifolds)

Example B.1 (A generic set of zero measure). Arrange the rationals in [0,1] in a sequence $r_1, r_2, \ldots$, and define the sets

$$Q_\varepsilon = \left(\cup_{k=1}^{\infty} (r_k - \frac{\varepsilon}{2^{k+1}}, r_k + \frac{\varepsilon}{2^{k+1}}) \right) \cap [0,1].$$

Then Q_ε is a dense open set in $[0,1]$, with Lebesgue measure

$$\lambda(Q_\varepsilon) \le \varepsilon \sum_{k=1}^{\infty} \frac{1}{2^k} = \varepsilon.$$

The set

$$Q_0 = \cap_{n=1}^{\infty} Q_{\frac{1}{n}}$$

is generic (a dense set of type G_δ). Since $Q_{\frac{1}{n+1}} \subseteq Q_{\frac{1}{n}}$, $\lambda(Q_0) = \lim_{n \to \infty} \lambda(Q_{\frac{1}{n}}) = 0$.

The dyadic permutations provide a simple example of an ergodic map:

Example B.2 (dyadic permutations of unit interval) We divide the unit interval in N equal intervals $I = [0,1] = \cup_{i=1}^{N} I_k$, with $I_k = [(k-1)/N, k/N]$. For a permutation $\sigma \in \mathsf{S}_N$ we define T_σ to map I_k onto $I_{\sigma(k)}$ by translation. Then T_σ is measure-preserving. If σ is cyclic, then T_σ is also ergodic.

Theorem B.2 (Poincaré's recursion theorem). *Let T be an ergodic transformation of the probability space X. Then for almost every point $x \in \mathsf{X}$, for every neighborhood V of x, the orbit of x crosses V infinitely many times.*

Distribution

1	MS 1322	John Aidun, 1435
1	MS 1316	William Brown, 1412
5	MS 1322	Paul Crozier, 1435
1	MS 1303	Gary Grest, 1114
1	MS 1395	Ahmed Ismail, 6711
1	MS 1320	Rich Lehoucq, 1414
1	MS 0895	Marcus Martin, 8333
1	MS 0321	Jennifer Nelson, 1430
1	MS 1316	Steve Plimpton, 1412
1	MS 1316	Mark Rintoul, 1412
1	MS 1411	Mark Stevens, 8332
1	MS 1322	Aidan Thompson, 1435
2	MS 9018	Central Technical Files, 8944
2	MS 0899	Technical Library, 4536