UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Molecular dynamics simulations on parallel computers

Permalink
https://escholarship.org/uc/item/6hw083v3

Author
DeBolt, Stephen E.

Publication Date
1993

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Molecular Dynamics Simulations on Parallel Computers:

A Study of Polar versus Nonpolar Media Effects in Small Molecule Solvation

by

Stephen E. DeBolt

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Chemistry

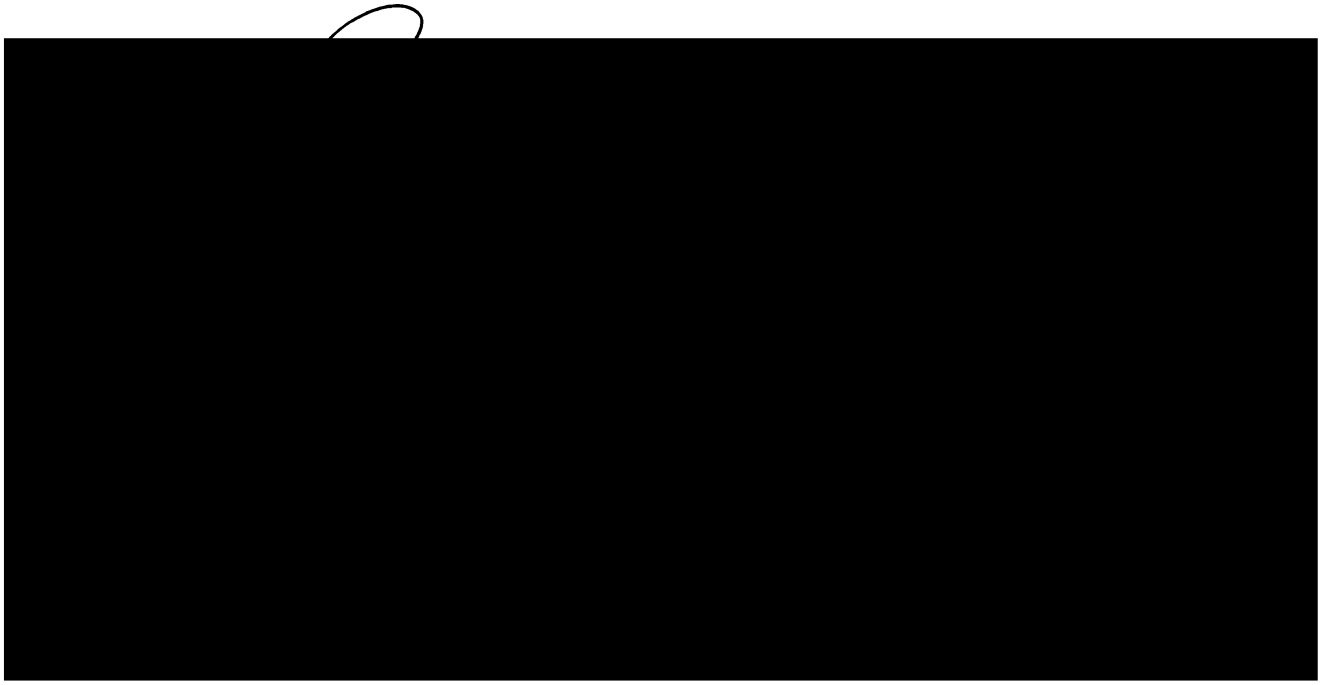
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



Copyright 1993
by
Stephen E. DeBolt

Dedicated

to

Marty Stouffer

World Renowned Naturalist
And Nature Photographer

Yes, that's right. I've never met him. He doesn't know me. He is the fellow who does the "Marty Stouffer's Wild America" programs frequently broadcast on public television. He films creatures in their natural habitats, going about their natural behaviors. I really *like* that show. It is not so much in regarding him as a hero, as it is in respecting him for what he does and how he goes about it. Some happy times were spent when I was a boy, tramping freely about the fields and creeks with my dog, through the woods to secret ponds and places where I watched animals, explored, collected things, and generally indulged in Nature. I was constantly wide-eyed with awe in confronting some new intricacy or beauty of Nature, like suddenly coming across a giant irridescent dragonfly daubbing in eggs at the edge of a Spring brook on a sunny day. That frame of mind served as the doorway through which I came into Science.

As host of his Nature program, Mr. Stouffer presents and inculcates this same kind of perspective. Everyone can get in touch with the Nature Spirits by spending some time with the facinating wilder side of the world, even vicariously, through the TV. Now, rather than a boy blasing trails through the brush and stalking painted turtles at the edge of some backwater, I find that I am a scientist, deciphering the microscopic details of molecular solvation by observing Nature at the level of molecular interactions. Instead of going frog huntin', I go "investigating", and continue to enjoy, as I always have, the rewards of doing *serious research.*

ACKNOWLEDGEMENTS

I want first to thank my primary research advisor, Professor Peter Kollman, without whose input and steadfast support none of this research could have been done. In all of our discussions and interactions I have learned so many things, and not a few of them were outside the field of Science. It's unlikely that I could have found another research guide who was as suited to my independent temperament.

I am also grateful to Professor I. D. (Tack) Kuntz, who has also provided research counsel. On those occasions when we discussed my research, he had the most remarkable ability to immediately focus on and point out the crucial issues. In no other of my encounters with scientists has the quality of a quick bites-worth of information proven so useful.

I must point out that it was a series of stimulating discussions with Professor Frank Szoka that were instrumental in leading to the investigation of the structure and dynamics of 1-octanol. That also helped lead to the recognition of the potential usefulness of parallel computers in molecular simulations.

The people of Professor Robert Langridge's Computer Graphics Lab (Tom Ferrin, Conrad Huang, Eric Petterson, Al Conde, Greg Couch, Leslie Taylor, and Graham Redgrave) at UCSF have always been helpful. The color reproductions in Chapter II of this dissertation attest to the quality of their lab.

I gratefully acknowledge and appreciate the assistance provided by administrative secretaries Willa Crowell and Norma Belfer over these several years while I was doing graduate research.

If I did not thank Sherry Baker for standing with me as my companion, friend, and helper during my graduate years, I would be remiss. During this endeavor she has "been there" for me so many times; nothing and no one could possibly substitute. And I should credit Mom and Dad DeBolt for all of their encouragement throughout my life. It's probably no small part that upbringing plays. They have made it easier by having instilled in me an inner strength and determination that undoubtly made the completion of this PhD project less of an ordeal than it might otherwise have been.

DISSERTATION ABSTRACT

Molecular Dynamics Simulations on Parallel Computers: A Study of Polar versus Nonpolar Media Effects in Small Molecule Solvation

Stephen E. DeBolt

Solvent effects were studied and decribed via molecular dynamics (MD) and free energy perturbation (FEP) simulations using the molecular mechanics program AMBER. The following specific topics were explored:

Polar solvents cause a blue shift of the $n \rightarrow \pi^*$ transition band of simple alkyl carbonyl compounds. The ground- versus excited-state solvation effects responsible for the observed solvatochromism are described in terms of the molecular level details of solute-solvent interactions in several modeled solvents spanning the range from polar to nonpolar, including water, methanol, and carbon tetrachloride.

The structure and dynamics of octanol media were studied to explore the question: "why is octanol/water media such a good biophase analog?". The formation of linear and cyclic polymers of hydrogen-bonded solvent molecules, micelle-like clusters, and the effects of saturating waters are described.

Two small drug-sized molecules, benzene and phenol, were solvated in water-saturated octanol. The solute-solvent structure and dynamics were analysed. The difference in their partitioning free energies was calculated.

MD and FEP calculations were adapted for parallel computation, increasing their "speed" or the time span accessible by a simulation. The non-cyclic polyether ionophore salinomycin was studied in methanol solvent via parallel FEP. The path of binding and release for a potassium ion was investigated by calculating the potential of mean force along the "exit vector".

TABLE OF CONTENTS

Molecular Dynamics Simulations on Parallel Computers:
A Study of Polar versus Nonpolar Media Effects
in Small Molecule Solvation.

CHAPTER I

A Theoretical Examination of Solvatochromism and Solute-Solvent Structuring in Simple Alkyl Carbonyl Compounds. Simulations Using Statistical Mechanical Free Energy Perturbation Methods.

CHAPTER II

A Comprehensive Investigation of Structure and Dynamics in 1-Octanol and Water-Saturated 1-Octanol, Molecular Dynamics and Free-Energy Perturbation Studies.

CHAPTER II (CONTINUED)

CHAPTER III

AMBERCUBE MD,
Parallelization of AMBER's Molecular Dynamics Module
for Distributed-Memory Hypercube Computers

CHAPTER IV

Free Energy Perturbation Calculations
On Parallel Computers:
Demonstrations Of Scalable Linear Speedup

LIST OF TABLES

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

LIST OF FIGURES

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

DISSERTATION INTRODUCTION

Molecular Dynamics Simulations on Parallel Computers:
A Study of Polar versus Nonpolar Media Effects
in Small Molecule Solvation

DISSERTATION INTRODUCTION

This research project has a bipartite theme.

(1) The first objective was to use the techniques of computational chemistry to explore the thermodynamics of solvation and the structural features of solute-solvent interactions in aqueous versus non-aqueous medias. "Solvent effects" in membrane-mimicking medias were of particular interest. Molecular dynamics (MD) simulations and statistical free energy perturbation (FEP) calculations were carried out in explicitly modeled liquid water, methanol, carbon-tetrachloride, hexane, octane, and pure or water-saturated 1-octanol.

(2) The second objective was to facilitate these computationally intensive simulations by adapting the MD and FEP source modules of the AMBER program to parallel computers. This entailed the development, optimization, testing, and benchmarking of a parallel version of the molecular dynamics algorithm and auxiliary support routines.

Molecular Mechanical Modeling

Let us begin by reviewing the fundamentals of the modeling methodologies utilized in this project. A classical *molecular mechanical force field* function can be used to describe a system of interacting particles. In this project, a molecular system's

potential energy is divided into intramolecular and pairwise intermolecular terms, each depending on the coordinates of the individual particles.

$$\upsilon_{Total} = \sum_{bonds} K_b (b - b_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left\{ \varepsilon_{ij} \left[\left[\frac{r^*_{ij}}{r_{ij}} \right]^{12} - 2 \left[\frac{r^*_{ij}}{r_{ij}} \right]^{6} \right] + \frac{q_i q_j}{\varepsilon r_{ij}} \right\} . \quad (1)$$

In the AMBER[1] molecular mechanical (MM) potential function, υ_{Total} is the potential energy of the system; K_b and b_{eq} are the bond stretching constant and equilibrium bond distance; K_Ω and Ω_{eq} are the bond angle stretching constant and equilibrium bond angle; V_n, n, and γ are the torsional force constant, periodicity of the torsional term, and the phase angle; r_{ij}^* and ε_{ij} are the (Lennard-Jones) nonbonded interaction parameters for particles i and j; r_{ij} is the interatomic distance between centers i and j; q_i and q_j are the atomic partial charges on i and j; ε is the effective dielectric constant. The component of potential energy due to the presence of a given particle is taken as the simple additive sum of all terms involving the particle. Taking the derivitive of this function with respect to coordinates provides the force on each particle by virtue of its position within the "field" of surrounding particles.

Once the potential energy and forces are known for a system of interacting particles, the time evolution of a system can be projected. A sufficiently long trajectory through phase space will give a representative sample, a *time-average ensemble*, from which thermodynamic, mechanical, and structural properties can be computed.

Molecular Dynamics

Molecular dynamics algorithms permit the modeling of particle interaction dynamics by integrating Newton's equations of physical motion. A system of particles (say atoms) interact by virtue of their potential and kinetic energies. Such a system can be characterized at a given instant in time by the position $\mathbf{r}_i$ and momentum $\mathbf{p}_i$ variables associated with each particle *i*. MD algorithms use the equation

$$\mathbf{F}_i = m_i\mathbf{a}_i = \frac{\partial \mathbf{p}_i}{\partial t} = m_i\frac{\partial \mathbf{v}_i}{\partial t} = m_i\frac{\partial^2 \mathbf{r}_i}{\partial t^2} = -\nabla \mathcal{U}(\mathbf{r}) \qquad (2)$$

where $\mathbf{F}_i$ is the force on particle *i*, m_i is the mass of *i*, $\mathbf{v}_i$ is the velocity, **r** and **p** are as above, and $\nabla\mathcal{U}(\mathbf{r})$ represents the spacial derivitive of the potential energy with respect to Cartesian coordinates.

A finite time-step method solution to equation 2 is used to smoothly vary the positions of the particles with time. Given the atomic forces, positions, and velocities at time *t*, the positions and velocities at a later time $t+\delta t$ can be predicted. After adjustment of the particle positions, the new forces at time $t+\delta t$ can be calculated via the spatial derivitive of the classical molecular mechanical force field equation. Reiterating this process on a step-by-step basis produces a particle trajectory.

The most direct method of integrating the equations of motion is via the Verlet algorithm[2,3]. The method uses the positions at **r**(t), accelerations **a**(t), and the positions **r**(t-δt) from the previous step. Recall the equation of motion for predicting particle positions is just a Taylor expansion about **r**(t):

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 + \cdots \quad (3)$$

and analogously for predicting prior positions:

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 - \cdots \quad (4)$$

Subtracting equation 4 from equation 3 removes the term involving velocity, giving

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \mathbf{a}(t)\delta t^2 \quad (5)$$

from which new positions $\mathbf{r}(t+\delta t)$ are available when the current positions, the immediately prior positions, and the current accelerations (obtained from the forces) are known. Using this formula, velocities are not needed to produce a trajectory, but for computing the kinetic energy $1/2m\mathbf{v}^2$ (and thus the total system energy when added to the potential energy), they are available as

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)}{2\delta t} \quad . \quad (6)$$

The MD module of AMBER uses a modification to the basic Verlet algorithm, the so-called half-step "leap-frog" algorithm, which provides more accurate velocities (on the fly), and generally better precision[4,5]. There are many finite-difference algorithms (of varying accuracy and programming complexity) for predicting particle motions[6,7].

Free Energy Perturbation

The free energy perturbation methodology (FEP), a method for calculating free energy *differences* is discussed in detail in chapters I, II, and IV. In chapter I, FEP is applied to compute differences in solvation free energies. In chapter II, FEP is used to compute relative octanol/water partition coefficients for two small solutes. The particulars of FEP as it relates to parallel processing are discussed in chapter IV, where the method is used to demonstrate the calculation of a potential of mean force along a geometric constraint trajectory for the decomplexation of potassium from a carboxylic ionophore, the free energy of binding for a ligand-enzyme interaction, and differences in small molecule octanol/water partition coefficients computed in parallel.

Generally, application studies employing FEP forgo a discussion of the statistical mechanical formalism underlying the method; more commonly, it is merely referenced. Rather than assuming the reader's understanding of the theory, an appropriately brief derivation will be given here as preparation (or a refresher) for chapters I, II, and IV. The FEP formalism used in the AMBER program can be derived from the statistical mechanical partition function for an ensemble of particles. The free energy F of a system is related to the statistical mechanical partition function Q,

$$F = -k_B T \ln Q \tag{7}$$

where k_B is the Boltzman constant (abbreviated as k hereafter), T represents temperature in Kelvin, and the partition function Q is

$$Q = \sum_{i=1}^{states} e^{-\left[\frac{H(p,r)}{kT}\right]} \tag{8}$$

where H(p,r) is the system's Hamiltonian, the sum of kinetic and potential energies with respect to momenta and position

$$H(p,r) = \sum_{i=1}^{states} \frac{\mathbf{p}^2}{2m} + \upsilon(r) \quad . \tag{9}$$

The free energy can be expressed as

$$F = -kT\ln\int e^{-\left[\frac{H(p,r)}{kT}\right]}\partial p\partial r = -kT\ln\int e^{-\left[\Sigma\frac{p^2}{2m}+\upsilon(r)\right]\cdot\frac{1}{kT}}\partial p\partial r = \tag{10}$$

$$= -kT\ln\left[\int e^{-\left[\Sigma\frac{p^2}{2mkT}\right]}\partial p \int e^{-\left[\frac{\upsilon(r)}{kT}\right]}\partial r\right] \quad .$$

and the difference in free energies between systems A and B is

$$\Delta F_{AB} = -kT\ln\left[\frac{Q_A}{Q_B}\right] \quad , \tag{11}$$

which can be expressed as

$$\Delta F_{AB} = -kT\ln\left[\frac{\int e^{-\left[\Sigma\frac{p_A{}^2}{2mkT}\right]}\partial p \int e^{-\left[\frac{\upsilon_A(r)}{kT}\right]}\partial r}{\int e^{-\left[\Sigma\frac{p_B{}^2}{2mkT}\right]}\partial p \int e^{-\left[\frac{\upsilon_B(r)}{kT}\right]}\partial r}\right] \quad . \tag{12}$$

Now the integral expressions containing the momentum components in equation 12

cancel each other since a "heat bath" is implemented to hold T constant during the MD simulation. Therefore, the kinetic energy expressions for systems A and B (assuming A and B possess the same number of degrees of freedom) are equivalent because

$$\frac{\mathbf{p}_A^2}{2m} = \frac{1}{2}m_A\mathbf{v}_A^2 = \frac{3}{2}Nk_BT = \frac{1}{2}m_B\mathbf{v}_B^2 = \frac{\mathbf{p}_B^2}{2m} \quad . \tag{13}$$

This brings us to the resulting statistical mechanical FEP equation

$$\Delta F_{AB} = -kT \ln < e^{-\left[\frac{\upsilon_A - \upsilon_B}{kT}\right]} >_A = \tag{14}$$

$$= -kT \ln \left[\frac{1}{N} \sum_i^N e^{-\left[\frac{\Delta\upsilon_{AB}}{kT}\right]} \right]_A \quad , \tag{15}$$

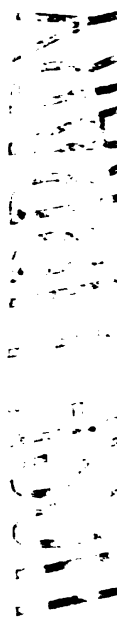
where $< >_A$ in equation 14 indicates the *A* state(s) are used to generate the time-average "reference" ensemble. The *B* state is superimposed on the *A*-ensemble to obtain the exponentiated difference in *A* and *B* state potential energies. This is the "perturbation" in the procedure. Note that state *B* has no effect on the time evolution of the system. In practice, the reference *A*-ensemble is generated by collecting data over a series of finite steps "through" time via the molecular dynamics methodology, then averaging as in equation 15 to get the expectation value. Commonly, the entire calculation will be redone using the *B* states as the reference ensemble to determine whether any hysteresis results from this approximation. The results of the *A*-reference and *B*-reference calculations are then averaged to "dilute" (hopefully) any hysteresis in the computed free energy values.

So long as the two states *A* and *B* are not two dissimilar, i.e. their potential energy surfaces nearly overlap, and their net free energy difference ΔG_{AB} is less than about k_BT, and sampling is sufficient, the method is rigourous. This assumption of a *small* perturbation must be met. When the energy difference between states *A* and *B* is great enough to contradict this assumption, intermediate states can often be interposed, e.g. *A*, $I(AB)_1$, $I(AB)_2$, ..., $I(AB)_N$, *B*. A *small* perturbation can now take place between each of the intermediate states (including the end-states). The accumulation of the free energy differences from all the intermediate states $\Sigma_j \Delta G(I_{AB})_j$ is equivalent to ΔG_{AB}. This gives the difference free energies of two generic systems *A* and *B*, e.g. two similar but slightly differing forms of a molecule, bringing us to the FEP equation as it is applied in chapters I, II, and IV.

Aqueous Versus Non-Aqueous Solvation

To facilitate molecular dynamics calculations exploring non-aqueous solvation phenomena, the EDIT source module of the program AMBER (where the solvent topology is built) was first modified to handle arbitrary solvents as opposed to water only. Previously, AMBER was "hard-coded" to accept only a 3-center solvent, namely 3-point water. The FORTRAN source code for GSEDIT (General Solvents EDIT) is included here in Appendix I. With this extension of the molecular mechanics/dynamics program AMBER, computations involving solubilization in arbitrary solvents are more conveniently undertaken.

Prior to engaging in such studies, an appropriate set of classical force-field parameters must be developed for each solvent, and refined by verifying their ability to reproduce experimentally observable physical quantities. Several model organic solvents were parameterized for use with AMBER's united-atom and all-atom molecular mechanical potential energy functions. Solvent parameters for water, methanol, 1-octanol, carbon-tetrachloride, hexane, and octane are reported in the chapters discussing their utilization. Also, although the study is not reported as part of this dissertation, AMBER/MD parameters for fluorinated alkanes are contained in "Derivation of Fluorine and Hydrogen Atom Parameters Using Liquid Simulations" by C. A. Gough, S. E. DeBolt, and P. A. Kollman[8]. In each case, it was found that many experimentally observable physical properties of pure liquids and solvent-bathed solutes are reproducible by computer modeling at this level of resolution. When molecular level structural, energetic, and/or dynamic characteristics of solvated systems are experimentally inaccessible, qualitative predictions made on the basis of well parameterized MD model systems are often a credible alternative.



Solvatochromism From MD/FEP

The first study of this project is an application of computational methodologies to investigate solvatochromism due to differential solvation energies of the ground versus excited states of solutes. The $n \rightarrow \pi^*$ electronic transition of simple alkyl ketones was examined in several solvents spanning the range from polar to nonpolar, including water, methanol, and carbon-tetrachloride. Which solute-solvent association

phenomena cause the observed spectroscopic blue shift for the n→π* excitation (relative to the gas phase transition energy) in more polar hydrophilic solvents? In polar media, solute-solvent hydrogen bonding is stronger in the ground-state where the molecular dipole along the carbonyl axis is larger than in the excited singlet or triplet state solutes. But upon excitation, the system favors a more desolvated π* state because of competition between solute-solvent and solvent-solvent hydrogen bonding in the surrounding bulk water. In a nonpolar non-hydrogen bonding solvent, the difference between the solvation energies of ground and excited states is of less consequence. In fact, overall solvation free energy may be lower when the excited state is present; in carbon-tetrachloride the transition is red shifted relative to the gas phase transition. This study provided a molecular level test system for investigating the role of hydrogen bonding in the hydrophobic effect. It was found that after solvent relaxation, an enthalpicly stronger clathrate structure, built of a water-water hydrogen bond network, is formed around the excited state. The more hydrophobic excited state is essentially "walled-out" of solution.

This solvatochromism study highlighted the necessity of running the simulations involving methanol solvent media over significantly longer time spans relative to water or carbon-tetrachloride. Often, non-aqueous solvents reorient more slowly than water, particularly when the individual solvent molecules possess both polar and a nonpolar groups, such as the alcohols. Methanol simulations typically require in the neighborhood of 200 picoseconds to reach convergence; most water studies approach convergence in about 20-30 psec. Satisfactory parameters for model non-aqueous solvents are determined by following a series of many simulations until the best least-squares fit to a

set of calibrating physical properties is reached. The simulation time-length required for convergence of calibrating properties will vary with the particular properties chosen and the intrinsic relaxation time for the solvent being parameterized. Longer sampling times are also needed when an MD simulation is used to generate an ensemble of states from which an average property will be predicted (such as the differential solvation energies of solute electronic states).

MD Over Long Time Scales; Membrane Analog Systems

There are several issues of concern in dealing with MD simulations involving systems of large size, complexity, and inherently slow relaxation times. These distill down to the following problems:

(1) Arriving at an accurate set of interaction potentials.
(2) Ensuring the attainment of equilibrium in complex modeled systems.
(3) Adequately sampling the distribution of states over "phase space" to provide sufficient statistics.

Originally, at the inception of this project, the idea of modeling the interactions of phospholipid biomembranes (a "patch" of bilayer) with drug-sized molecules was considered. In order to elucidate the telling molecular-level details regarding the interactions of small molecules with membranes, especially those details which are unavailable to experimentalists, it would be useful to simulate these phenomena via high resolution computer modeling. However, the physical reorientation times of membranes are so slow relative to the time scale accessible to MD simulations that this idea was ruled

out as unfeasible at the time. Liquid water's rotational relaxation time is about 2.5 picoseconds, and its dielectric relaxation time of about 8-10 picoseconds (25°C), easily spanned by an MD simulation. The solvation of a small solute in a membrane bilayer requires on the order of microseconds -way out of reach for all-atom MD, where a reasonable maximum time span for simulating a system of 2000-5000 atoms is only a nanosecond or two.

The task in modeling such interactions involves overcoming what might be called *the computer modeler's dilemma.* When considering molecular-level interaction details, a model accounting for nearly all of the atom-wise interactions is required. Also, a thorough statistical sampling of a modeled system's distribution among it's thermodynamically accessible states must be obtained. A high-resolution model of a large molecular mechanical system, with its many degrees of freedom, would require an exorbitant amount of computer-time to achieve completeness in sampling. The reality of finite computer processing speed imposes a limitation on the degree of resolution that can be used in a molecular model. If a high-resolution all-atom model is used, all of state-space can not be spanned during a simulation using the current generation of digital computers. Alternatively, when a very low-resolution model containing fewer degrees of freedom is used, state-space may be well sampled. But, a low-resolution model cannot explore the relevant questions regarding the specific molecular-level details of interest.

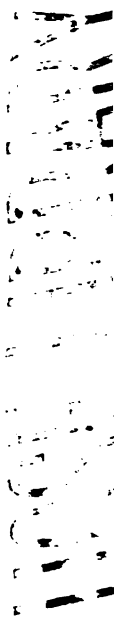
Two avenues of pursuit were followed to help bring long time-scale MD simulations (such as the explicit non-aqueous solvation of small to medium sized molecules

mimicking solvation in membranes) closer within the realm of computational feasibility.

(1) Lowering the number of degrees of freedom in the system, that is, to model at a level of *moderately less detailed representation* (united-atom model), using simpler analog molecules to play the part of the amphiphilic molecules found in biomembranes.

(2) Physically speed up the calculation by taking advantage of the new and developing *parallel supercomputer technology*.

Considerable success was acheived in these dual efforts to increase the sampling range of MD simulations. Even so, these studies highlight the clear need to develop alternative modeling methods to study molecular phenomena which occur over a time-span considerably longer than that currently accessible via MD simulations, e.g. the partitioning of solutes in biomembranes, the conformational behavior of peptides, the folding of proteins, and intermolecular complexation of macro-molecules.



Octanol Media As Membrane Mimetic

The octanol/water partitioning system has become the standard reference for investigating the distribution behaviors of pharmaceutical agents in the body[9] and of organic contaminants in the environment[10]. Octanol/water media also occasionally serves in the lab as a lipid-like substitute for predicting the pharmacokinetic charac-

teristics of drug compounds in body tissues. Octanol media is probably the closest analog to a biological membrane that can be studied over a long time span via high-resolution MD simulations. *And* much experimental data is available regarding the solvation and partitioning of small molecules in octanol.

Interaction potentials for 1-octanol were based on those of the separately parameterized methanol, hexane, and octane. These solvents contain the polar "head" and long chain hydrocarbon "tail" which are present in octanol. This work has resulted in a reliable set of intra- and interatomic united-atom-model MD interaction potentials for amphiphilic long-chain alcohols. Observables used for calibration of the force field included the experimentally determined pure-liquid density, heat of vaporization, and the distribution of chain conformations in alkane chains. The models reproduced the experimental values of these quantities computed as an average over a molecular dynamics time ensemble.

Chapter II is a comprehensive study of the internal structure of pure 1-octanol and water-saturated octanol media via radial distribution functions, and detailed analyses of hydrogen bonding, cluster formation, and regions of preferential hydrophilic versus hydrophobic solvation. The specific mechanisms of dielectric response phenomena in long-chain amphiphiles (1-octanol) are investigated. Analysis of the trajectories generated during these studies have enabled further characterization of the model and comparisons with relevant experimental physical data such as self-diffusion coefficients. The results support (or negate) the various explanations offered by experimentalists to account for the multiple frequency dielectric responses measured in long-chain

alcohols, phenomena which have not been entirely understood to this date. These studies are at the cutting edge of modern liquid simulations.

The effects of saturating waters on the structure and dynamics of 1-octanol was investigated. Calculations of the octanol/water transfer free energies for small solutes serve as an indication of the degree to which these new interaction potentials facilitate the simulation of hydrophobic versus hydrophilic solvation phenomena. Another important result of these simulations is the detailed information regarding specific atomic level interactions which are responsible for the differences in the partition coefficients of drug molecules.

Parallel Processing Strategy

Adaptation of MD/FEP algorithms for the current generation of high performance parallel supercomputers has become extremely relevant as theoretical chemists continue to treat larger and more complex molecular systems. There is an ever increasing need for more rapid processing of the numerical computations in particle interaction simulations. As the speed of computing increases with the onset of concurrent supercomputing, computational chemists can begin to investigate processes for which the inherent relaxation times are comparatively "slow". Examples of such systems include biomembrane component interactions, membrane-to-solvent-surface interface interactions, and the interactions of membranes with solutes, peptides, drugs, ionophores, etc, as well as the long-time effects in the interactions of proteins, dna, and other polymers.

Chapter III discusses in technical detail the algorithmic strategy used in parallelizing MD for distributed memory computers. For a scientific programmer interested in parallelizing a program that simulates a time-series of dependent events, particularly particle simulations, chapter III could provide invaluable information and strategies. If the reader is looking for an overview of the motivation, general approach, and results from the MD parallelization effort, the Introduction, Parallel Performance, and Summary sections should serve nicely. An overall rethinking of the algorithms, data-structures, and subroutine interplay, was necessary to effectively parallelize the MD module of AMBER. Parallel adaptation has proven to require a significant amount of recoding and writing of auxiliary support procedures. The approaches to particle simulation which were outlined in the parallel programming literature were not well suited for the adaptation of molecular dynamics as implemented in AMBER. In order to adapt AMBER MD to distributed memory machines, and preserve the program's broad generality of function, it was necessary to develop custom subroutines. Generality of function in the parallel version is demonstrated by running a set of applications of graduated complexity.

In adapting AMBER MD/FEP for parallelism, I have explored both fine-grain and coarse-grain "work distribution" approaches. A fine-grain approach is required to enable continuous long-time-scale simulations of serially dependent molecular events. This facilitates more thorough and representative sampling of the molecular systems of interest. This approach is especially important when the objective is to reproduce or predict time-dependent properties where the time-scale of the events mediating such properties are near the MD accessibility limit for sequential computer technology. The

distribution of such a calculation over many processors requres a fine-grain re-tooling of the traditional MD algorithms in order to obtain the best possible scale-up with added processors.

It was possible to get very good speedup of the most computationally intensive MD subroutines when using a fine-grain parallelism approach. The speedup was not purely linear when many (over 16) processors were used, but further speedup was obtained even up to 64 processors on the nCUBE hypercube machine. For simple molecular systems such as pure liquid water, the most intensive part of the simulation (the evaluation of pairwise nonbonded interactions over a system of atoms) can give approximately 80% to nearly 100% of linear speedup depending on the number of processors used. This was achieved by implementing a simple *domain decomposition/recomposition* scheme for distributing the computational work over multiple processors. But to obtain this kind of performance in more complex systems, such as a solvated protein, coding of a pairlist load-balancer was required. I have also subjected the nonbonded interactions pairlist generator to parallelization, developed a "smart" automatic decomposer of system topologies so that the SHAKE routine (which maintains rigid bonds between atoms) can be parallelized, and so that the dihedral-angle, 1-4 nonbonded interaction terms, angles, and bond energies/forces are executed in parallel.

On the other hand, in molecular systems where extended sampling is required, but the evolution of a continuous trajectory is *not* critical, or in any case where the simulation can be broken down into discrete independent pieces, a coarse-grain approach to

parallelization can be taken. Inherent in this approach is a very efficient scaling-up of the speed, closely approaching linear proportionality with the number of processors used. When a free energy perturbation between states *A* and *B* is treated as a sum of N independent intermediate states, the methodology lends itself perfectly to coarse-grain parallelism. This has been demonstrated with a variety of molecular systems in chapter IV which discusses the parallelization of FEP.

A particularly interesting parallel MD/FEP calculation involves the carboxylic ionophore salinomycin. Salinomycin$^{-}K^{+}$ binding is studied in methanol, a solvent which mimics the aqueous/lipid interface boundary. A potential of mean force (PMF) calculation using the FEP methodology is carried out in coarse-grain parallel by driving the ion from the ionophore's center of mass, into the solvent media. The ion-ionophore mechanism/path of binding and release is observed as a series of associated conformational changes and progressive decomplexations of liganding oxygen groups.

We will eventually need on the order of 100 to 1000 times greater processing speed to simulate interesting drug/membrane and protein/membrane systems at the all-atom level of resolution. The use of high performance parallel computers will extend the time-span accessible to MD and FEP simulations. This will undoubtedly lead to valuable molecular level information concerning interaction specificities, information which should be pertinent in the design of new pharmaceutical agents.

REFERENCES: DISSERTATION INTRODUCTION

(1) D.A Pearlman, D.A. Case, J.A. Caldwell, G.L. Seibel, U.C. Singh, P. Weiner, P.A. Kollman, *AMBER 4.0*, University of California, San Francisco (1991). U.C. Singh, P.K. Weiner, J.W. Caldwell, P.A. Kollman, **AMBER (UCSF)**, version 3.0, Department of Pharmaceutical Chemistry, University of California, San Francisco, (1986). S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A Case, *J. Comp. Chem.*, **7**, 230 (1986).

(2) L. Verlet, "Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules", *Phys. Revs.*, **159**, 98-103 (1967).

(3) C. W. Gear, "Numerical Initial Value Problems in Ordinary Differential Equations of Various Orders", Prentice-Hall, Englewood Cliffs, New Jersey (1971).

(4) R. W. Hockney, *Methods Comput. Phys.*, **9**, 136-211 (1970).

(5) D. Potter, "Computational Physics", Wiley, New York (1972).

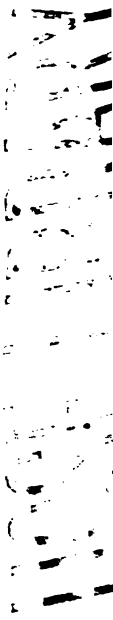
(6) M. P. Allen, D. J. Tildesley, "Computer Simulation of Liquids", Chapt. 3, Clarendon Press, Oxford, (1987).

(7) W. F. van Gunsteren, H. J. C. Berendsen, "Algorithms for Macromolecular Dynamics and Constraint Dynamics", *Mol. Phys.*, **34**, 1311-1327 (1977).

(8) C. A. Gough, S. E. DeBolt, P. A. Kollman, *J. Comp. Chem.*, **13**, 963-970 (1992).

(9) C. Hansch, "Quantitative approach to biochemical structure-activity relations", *Accounts.Chem. Revs.*, **2**, 232-239 (1969).

(10) S. W. Karichkhoff, D. S. Brown, T. A. Scott, "Sorption of hydrophobic pollutants on natural sediments", *Water Res.*, **13**, 241-248 (1979).



CHAPTER I

A Theoretical Examination of Solvatochromism and Solute-Solvent Structuring in Simple Alkyl Carbonyl Compounds.

Simulations Using Statistical Mechanical Free Energy Perturbation Methods.

The text of this dissertation chapter I is a reprint of a very similar article that appeared in The Journal of the American Chemical Society in 1990. The coauthor listed in the JACS article, Peter A. Kollman, directed and supervised the research which forms the basis of this chapter.

Permission has been granted from the ACS provided the following line appears on the first page of this article (chapter):

Reprinted with permission from: S. E. DeBolt, P. A. Kollman, *Journal of the American Chemical Society*, **112**, 7515-7524 (1990). Copyright 1990 American Chemical Society.

A Theoretical Examination of Solvatochromism and Solute-Solvent Structuring in Simple Alkyl Carbonyl Compounds. Simulations Using Statistical Mechanical Free Energy Perturbation Methods.

ABSTRACT

The difference in Gibbs free energies and potential energies of solvation ($\Delta\Delta G_s$ and $\Delta\Delta V_s$) for the ground state versus the $n\rightarrow\pi^*$ excited state were calculated for the carbonyl containing chromophores formaldehyde and acetone. $\Delta\Delta V_s$ corresponds to the difference in (ground state to Franck-Condon state transition) excitation energy for a chromophore in the vapor phase versus when solvated, i.e. the solvent shift energy. $\Delta\Delta G_s$ corresponds to the solvated versus vapor phase energy difference between the ground and solvent-equilibrated excited state, i.e. the difference in free energy between the two adiabatic states. By the term "solvent equilibrated excited state" we refer to the thermally relaxed, solvent equilibrated system containing the electronically excited solute molecule, where Franck-Condon strain has been alleviated. Results were obtained for the carbonyl solutes in TIP3P water, modified-OPLS methanol, and five-center carbon tetrachloride model solvents. Although experimentally measured U.V. solvent shifts were not quantitatively reproduced in all cases, the calculated difference-energy values exhibited the proper qualitative trend in magnitudes with respect to this series of solvents. The calculated difference-energy values were found to compare particularly well with experimentally observed U.V. solvent shifts for the $n\rightarrow\pi^*$ electronic

transitions in water and methanol.

Characteristics of solute-solvent orientational structuring were examined for both ground and excited state solutes, especially in their relationship to solute-solvent interaction energies. Solute-solvent radial distribution functions provide interesting insight into the different characteristics of the average solvent structures around the ground and excited state solutes. Intermolecular energy pair distribution functions reported for the ground, the Franck-Condon, and the equilibrated-excited states illustrate a progressive loss of solute-solvent hydrogen bonding and indicate further desolvation of the solute after excitation in polar solvents. The solvent's response upon $n \rightarrow \pi^*$ solute excitation is consistent with current ideas regarding the solvation of hydrophobic moieties. Solvent-solvent interactions in the solute's near-shell and in the bulk solvent were examined. The change in the relative strengths of solvent-solvent versus solute-solvent interactions drives the restructuring of the near-shell solvent cage, thus contributing to the differential solvation of ground- and excited-state solutes.

BACKGROUND AND INTRODUCTION, THE N→Π* TRANSITION

Solvatochromism refers to the phenomena wherein solvents may affect the position, intensity and shape of U.V. absorption (or emission) spectra relative to the vapor phase. Qualitatively, the n→π* transition involves the promotion of an n-electron from the carbonyl oxygen's lone-pair orbital to the empty antibonding π* orbital which is delocalized over the carbonyl group. This results in an overall reduction of the magnitude of the molecular dipole moment in the excited state.[1,2]

The solvent dependent shift of the n→π* electronic transition band of carbonyl compounds toward shorter wavelengths in polar solvents is largely due to the differential solvation of the ground versus the excited state of the solute. This has been attributed to factors such as specific interactions, e.g. hydrogen bonding, in hydroxylic solvents, and more generally to solvent dielectric effects such as dipole-dipole and dipole-induced dipole interactions as well as packing and orientation strain in the Franck-Condon excited state. For all of the n→π* transitions considered in this study, the ground state solute has a larger permanent dipole moment than in the excited state, and is thus expected to exhibit a more well defined solute-solvent orientation in polar solvents due to stronger interactions. Such solutes are more favorably solvated in their ground state than in their excited state, thereby increasing the energy difference between states. This corresponds to a blue shift of transition frequency in polar solvents.[3-9]

A shift of the carbonyl n→π* band toward longer wavelengths is seen in some nonpolar nonorienting solvents. Bayliss and McRae have attributed this "polarization

red shift" to the electronically polarizable solvent's ability to facilitate the formation of the transition moment which develops in the process of absorption or emission.[10a,b,c] This effect is present in all solvents, but it figures most prominently in the red shift of the carbonyl n→π* bands in such solvents as carbon tetrachloride, benzene, and alkanes[10c] where orientational polarization is minimal. An alternative contribution to this effect is the decreased stabilization of the ground state relative to the excited state due to more favorable solute-solvent dispersion interactions involving the electronically diffuse excited state.

Ab initio CI calculations of optimally oriented, hydrogen-bonded, formaldehyde-water and acetone-water dimers have given dimerization energies which are nearly equal to the energy of the blue shift of these solutes in bulk water.[11a] Conclusions were drawn from these calculations as to the number of hydrogen bonds that form between these solutes and water. Also, the blue shift was attributed to the energy required to (largely) break the hydrogen bond between a solute's carbonyl oxygen lone-pair orbital and the hydroxyl group of water, a phenomenon that occurs on n→π* electronic transition.

After experimentally determining enthalpies of transfer of ketones and other organic solutes, and coupling these results with spectroscopic information, Haberfield concluded that the above explanation of solvent-induced frequency shifts was overly simplistic.[12] It was found that the n→π* blue shift of carbonyl compounds on going to more polar solvents was due not only to superior solvation of the ground state, but also in significant part to the desolvation of the nonpolar excited state in the polar solvent.

The experimental absorption data for the $n \rightarrow \pi^*$ transition in formaldehyde and acetone is most representative of the $S_0 \rightarrow S_1$ excitation. λ_{max} is in the vicinity of 3400 (formaldehyde) or 2800 (acetone) angstroms, depending on the solvent and temperature.[13-17] For acetone in several solvents, Borkman and Kearns's experiments have shown that, after excitation to the singlet, rapid intersystem crossing (ISC) to the triplet ($S_1 \rightarrow T_1$) occurs, at a rate which is independent of temperature between 77K and 298K, and is approximately 100 times the rate of $S_1 \rightarrow S_0$ radiation. ($\Phi_{ISC} = 1.0 \pm 0.1$; $k_{ISC} = 4 \times 10^7$ sec^{-1}) This has been shown to account for acetone's measurably low fluorescence quantum yield ($\Phi_F = 0.01 \pm 0.003$) at 25 °C.[17] The lifetime of the lowest $n \rightarrow \pi^*$ singlet S_1 was measured as $\tau_{S1} = 2.5 \times 10^{-8}$ sec^{-1}. This allows plenty of time for the solvent to reorient (which occurs within picoseconds) after excitation, prior to emission.

The intent of this study has been two-fold: first, to determine how well a molecular dynamics (MD) implementation of the statistical mechanical perturbation methodology, using a two-body interaction potential and a simple electrostatic model for the ground and excited states, can reproduce the solvent dependent frequency shift energies for the well characterized $n \rightarrow \pi^*$ electronic transitions of simple alkyl carbonyls and second, to examine the intermolecular orientational structuring which occurs at room temperature in fluid media between the ground and excited states of these carbonyl solutes and a series of several different solvents spanning the range from polar to nonpolar.

METHODS AND MODELS

All MD calculations were carried out using contant temperature (25 °C for water and methanol, 20 °C for carbon tetrachloride), constant pressure (1 atm), and composition (single solute) ensembles by coupling to a heat and pressure bath.[18] The original thicknesses of the solvent shells (radius out from the solute) before the application of cubic periodic boundary conditions were (in Å) : 10.0 for all simulations in water, using a nonbonded cutoff of 8.0, 14.0 for all simulations in methanol, using a 9.5 cutoff, and 15.0 for all simulations in carbon tetrachloride, with a 10.0 cutoff. For simulations run in TIP3P water and carbon tetrachloride solvents, a molecular dynamics timestep of 2 fsec was used (3 fsec for formaldehyde in water), and for the methanol simulations a timestep of 1 fsec was used. Physically realistic atomic masses were used, e.g. 12.01 for carbon, on all centers. Van der Waals parameters for the solutes were taken directly from the Wiener et al. all-atom force field.[19]

Molecular geometries and molecular orbital (MO) charge distributions used for the ground- and excited-state solutes formaldehyde and acetone were generated from *ab initio* calculations using the Gaussian-82 program[20a], except for the acetone $n\rightarrow\pi^*$ S_1 excited state, for which the optimized geometry and MO electronic distribution were obtained using the GAMESS program[20b]. This corresponds to isolated monomers at 0.0 K at their zeroth vibrational levels. Geometry optimizations of the monomers were at the 6-31G** basis set level for formaldehyde, and 6-31G* for acetone. R groups attached to the carbonyl carbon bent out of plane about 40 degrees for both formaldehyde and acetone during excited-state optimization, attaining the well-documented

pyramidal structure. Atom-centered point charges were obtained by fitting to the electrostatic potentials generated with the above basis sets using the program G80UCSF[21a], via the fitting procedure described previously.[21b]

For the molecular dynamics, no attempt was made to explicitly model electronic distributions in terms of directed spatial spin orbitals nor to include differences in ground-state versus excited-state solute van der Waals interaction parameters. Notice (Table I) that while the magnitude of the formaldehyde dipole moment resulting from the *ab initio* triplet calculation (1.55 Debye) is closer to that experimentally observed for S_1 (1.56 D), the difference in dipole moment magnitudes for ground state versus triplet is well preserved (2.33 - 1.29 = 1.04 D, experiment ; 2.66 - 1.55 = 1.11 D, calculated). We will occasionally refer to the π^* states modeled in this study as simply "the excited state".

Bond and angle force constants for the molecular mechanical formaldehyde were iteratively least squares fit to reasonably reproduce the known ground-state vibrational frequencies[22] in normal mode calculations. Excited-state force constants were extrapolated from these, e.g. the carbonyl force constant was slightly reduced corresponding to the longer, weaker C-O bond. An "improper" 180 degree dihedral (e.g. H1-H2-C-O) force constant was used to constrain the formaldehyde and acetone ground states more firmly to the planar configuration. Force constants used for acetone were similar to those in formaldehyde (see Table II).

Three different solvent models were used: (1) TIP3P waters[23], (2) modified-OPLS methanol[24] composed of 3 centers: H, O, and CH_3, where the methyl group is a united

atom center, and (3) five-center carbon tetrachloride. All solvents were equilibrated as pure solvents prior to solute insertion and are known (TIP3P water) or were required (carbon tetrachloride and modified-OPLS methanol) to reproduce experimental heats of vaporization and density to within a few percent. For simulations run in TIP3P water and carbon tetrachloride solvents, all solute and solvent bonds were constrained to remain rigid by using SHAKE, an algorithm that permits retention of fixed internal geometries during MD.[25] Simulations run in modified-OPLS methanol did not use SHAKE; rather, a loop-closing "bond" across the methanol hydroxyl hydrogen and united methyl group was used to maintain a rigid CH_3-O-H angle. The bond-stretching force constant used on the loop-closing bond was 900 kcal/mol, 320 for the Me-O bond, and 553 for the O-H bond. All other geometric and interaction parameters for methanol were identical to those used by Jorgensen[24], except in the later simulations involving the acetone $S_0 \rightarrow S_1$ transition in methanol, where the van der Waals parameters used on the united methyl group ($R^* = 4.00$ Å , $\varepsilon = 0.190$ kcal/mol) were slightly different than those of Jorgensen. We found that (after extensive MD equilibration of the pure methanol) the use of this modification gave better accord with the experimental volume while retaining the proper internal energy. This makes physical sense in that since this methyl group carries a partially positive charge (+0.265), due to its being immediately adjacent to an electronegative oxygen, one would expect its VDW radii to be somewhat smaller than that of a strictly alkyl methyl group, such as found in ethane. For the pure modified-OPLS methanol, heat of vaporization, and density were correct to within a few percent of experiment, and radial distribution functions (not shown) were essen-

tially identical to those reported by Jorgensen. For carbon tetrachloride the experimentally determined C-Cl and loop-closing Cl-Cl bond lengths (Table II) were used.[26]

No attempt was made to prefit the interaction parameters so as to specifically reproduce experimental solvent shifts. A simple molecular mechanical potential model was parameterized in a straightforward "assembly line" fashion, using *ab initio* calculations and some values already in the literature. This turned out to be adequate to qualitatively reproduce experimentally observed solvent shifts for electronic transitions in several solvents.

STATISTICAL MECHANICAL FREE ENERGY PERTURBATION

The molecular dynamics implementation of the FEP method in the program AMBER[27a,b] was used for the calculation of the difference in Gibbs free energy of solvation for a ground-state solute versus an n→π* excited state (the recent AMBER 3.0 Revision A was used only in simulations involving S_1 acetone). The perturbation is facilitated by changing or "mutating" the values of the parameters of a classical molecular mechanical potential energy function which mediates intermolecular interactions:[28]

$$V_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right]$$

The form and parameterization of this potential function has been discussed in detail.[19]

For a list of additional specific parameters used in this study see Tables I and II.

Figure 1 depicts the various paths of the thermodynamic cycle used to calculate the difference in energy for two electronic states of a solute *A* in a solvent versus gas phase. The differential solvation energies of ground versus solvent equilibrated excited states, $\Delta\Delta G_s$, path 4 minus path 3, could be calculated with FEP theory. But FEP theory requires a "small" perturbation in order to ensure adequate ensemble sampling and to minimize computational artifacts that could be generated by physically unrealistic systems. A solute can be "vanished from" or "grown into" solution in order calculate absolute solvation energies. However, this approach would require long sampling times and could lead to artifacts. Alternatively, paths 2 and 1 can be directly calculated using the perturbation methodology, so that a "small" perturbation is maintained. The (generic) transition energy ΔE_{TR}(gas) represented by path 1 in the above scheme (Figure 1) is simply equal to the isolated solute excitation energy, and that of path 2 ΔE_{TR}(solv) is the solute excitation energy in solution. The difference between path 2 and path 1, $\Delta E_2 - \Delta E_1$, gives the difference in excitation energy due to the influence of the solvent, which we refer to as either $\Delta\Delta V_s$ or $\Delta\Delta G_s$. When A(solv)* is considered as the Franck-Condon excited state, path 2 minus path 1 is given by $\Delta\Delta V_s$ and directly corresponds to the UV solvent frequency shift. When A(solv)* is considered as the solvent equilibrated excited state, path 2 minus path 1 is given by the difference in free energy $\Delta\Delta G_s$ between the ground and solvent equilibrated excited state.

There could be two contributions to $\Delta\Delta E_{TR}$, one coming from a difference in intrasolute energies, and another from solute-solvent interactions. We have assumed

that to a good approximation the intrasolute energies (mainly vibrational energies) cancel in the subtraction ΔE_2 - ΔE_1, and thus a simulation of path 2 gives the change in transition energy due to differential solvation of the solute *A* its ground versus excited state.

Theoretical aspects of free energy perturbation methods have been discussed in detail elsewhere[29], and only the fundamental equations will be recalled here as comments regarding our simulations will relate to these equations. Free energy was calculated via the continuous integration or "slow growth" method as

$$\Delta G = \int_0^1 \left\langle \frac{\delta V(\lambda)}{\delta \lambda} \right\rangle_\lambda d\lambda$$

where δV is the difference in potential energies between the reference and the perturbed systems when both are evaluated at the coordinates of the reference system according to the set of states generated by the coupling parameter λ. λ. varies along the integral [0,1] as the trajectory progresses. The free energy via single step (one-window) was also calculated in the usual way

$$\Delta G = -RT \ln \langle \exp(- (V_{pert} - V_{ref}) / RT) \rangle_{ref}$$

where the brackets $\langle\rangle_{ref}$ again denote a time average, *R* is the gas constant, and *T* is temperature. These two methods should give identical free energy differences in the limit of infinite sampling.

The point charges in the coulombic part of the nonbonded parameters and the parameters for bond lengths and angles were mutated as a function of trajectory time

(slow growth) or as a discrete single-interval "flash" perturbation (one-window), using a mutational coupling function of the form

$$P(\lambda) = (\lambda)\, P_{ref} + (1-\lambda)\, P_{pert}$$

where P_{ref} is a parameter assigned to the reference state solute and P_{pert} is the corresponding parameter which is assigned to the perturbed state solute.

When calculating the $\Delta\Delta G_s$ using the "slow growth" method, the FEP theory's assumption of $\Delta G < RT$ for each (micro) window is rigourously met. Using this method, and assuming representative sampling, one may expect to obtain a reliably quantitative $\Delta\Delta G_s$ value, but where the effects of Franck-Condon solute-solvent "strain" (nonoptimal packing and orientation at the instance of transition) are largely lost. These simulations do not correspond directly to an electronic transition, i.e. the transition is artificially "slowed" down, so that solute-solvent equilibration occurs throughout the MD perturbation trajectory, during the transition. These simulations should provide reliable values for the difference in free energy of solvation of the solvent-equilibrated ground state versus solvent-equilibrated excited state.

When sampling over a single perturbation window (one-window method) where the ΔG involved is greater than several RT, FEP theory begins to break down as the fundamental assumption of a "small" perturbation is no longer rigorously met. Some incompleteness of ensemble sampling is probably inevitable in such a situation. In light of this, the calculated ΔG is usually expected to be less quantitative for one-window perturbations where λ jumps directly from $1\rightarrow0$ or $0\rightarrow1$, e.g. ground state $\rightarrow$ excited, or excited state $\rightarrow$ ground state all in one step. However, these one-window

calculations do facilitate the direct simulation of an electronic transition event in a solvent environment, as $\Delta\Delta V_s$ is also calculated for the ensemble over the MD trajectory. During these simulations the solvent equilibrates about and orients with respect to the reference solute, e.g. the ground state for an absorption. In this example the excited state "flashes" into the solvent cage structure which has been oriented about the ground-state solute, directly modeling what is customarily referred to as the Franck-Condon excited state. Recall that the time required for an electronic transition is of the order of 10^{-15} seconds, while intramolecular relaxation occurs more on the order of 10^{-14} to 10^{-13} seconds, and 10^{-13} to 10^{-12} for intermolecular reorientations.[30] Consequentially, internuclear distances do not have time to change much during an electronic transition. Only electronic polarization of the solvent responds immediately on excitation. The transition can be considered (prior to nuclear relaxation) to be the superpositioning of the excited-state electronic distribution onto the solute geometry and potential energy surface which was prepared as a result of ground-state solute-solvent interactions. A single window perturbation treatment closely corresponds to this physical situation, although we have omitted solvent electronic polarization. The thermally equilibrated excited state is subsequently reached through relaxation processes, including reorientation of the solvent. A Franck-Condon nonequilibrum state also appears (in the reverse sense) during an emission event. The ensemble configurations generated during the single step, one-window perturbation calculations will be recalled later on as their analysis provides valuable qualitative insight into the influence of solute-solvent and solvent-solvent structuring on the energetics involved in the $n\rightarrow\pi^*$ transition process.

For the one-window runs the $\Delta\Delta V_s$ was calculated (in both forward and reverse directions) at every MD step and averaged over the entire trajectory, facilitating the calculation of the difference in solvation free energies of the ground- and excited-state solutes. Several individual calculations were performed for each perturbation, each from different pre-equilibrated starting configurations, and the values of each of the separate calculations were averaged (time-weighted) to give the overall $\Delta\Delta V_s$ and $\Delta\Delta G_s$.

The usual procedure when using windows is to equilibrate about and perturb with respect to one reference state and then to do the same for the opposite state, taking the average of these two calculations as the ΔG value for that window. One expects this to give a more quantitative value for one-window calculations. In particular, such double-wide sampling ordinarily leads to a more thorough representation of the surface and hopefully at least a partial cancelling of hysteresis effects. It is the averaged $\Delta\Delta G_s$ of the forward and reverse single-step simulations that should be considered as the value corresponding to the difference in free energy between the solvent-equilibrated ground state and the solvent-equilibrated excited state. In short, for our two roughly similar systems, i.e. a single ground- or excited- state solute in a periodic box of solvent, the generation of samples using either electronic state as the reference monomer essentially provides a complementary ensemble which enhances information about the difference in energy between different areas of the potential surfaces.[31] One may also note (Table III) the general agreement between the one-window (forward and reverse averaged) calculations with the more rigourous slow growth method.

FREE ENERGY AND POTENTIAL ENERGY RESULTS

The calculated average $\Delta\Delta V_s$ and $\Delta\Delta G_s$ values are compared to spectroscopically measured solvent dependent frequency shift values which were available (see Table III). The calculated values exhibit the proper trend, i.e. the $n\rightarrow\pi^*$ transitions shift further toward the blue as successively more polar solvents are considered. The experimental shift values in terms of kcal/mol for acetone solute in water, in methanol, and in carbon tetrachloride are listed in table III (column 8) as 4.46, 2.40, and -1.43 respectively, along with an alternative set of values (5.78, 4.18, 0.00) from a second source (which may be slightly less rigourous) listed in parentheses. The forward (e.g. A $\rightarrow$ A*) one-window $\Delta\Delta V_{1W}$ values are comparably close to those of experiment, and quantitative within a single standard deviation (1 SD) for water and methanol solvents when the $S_0\rightarrow S_1$ transition is considered. The average $\Delta\Delta V_{1W}$'s calculated (column 7) are 4.75 ± 1.38 and 3.96 ± 1.89 for acetone's $S_0\rightarrow S_1$ transition in water or in methanol, respectively. For carbon tetrachloride, $\Delta\Delta V_{1W}$ is 0.25 ± 0.28 using the 6-31G* charge model. Regardless of the carbon tetrachloride charge model used, $\Delta\Delta V_{1W}$ values were always slightly positive, with oscillations occasionally carrying over into negative shift energies for acetone's $S_0\rightarrow T_1$ transition.

Using data tabulated in table III, one can graphically illustrate (Figure 2) the effects of solvent on the shifts of absorption or emission frequency. The arrow-headed "transition" lines are not drawn to full scale since the absolute magnitude of the $n\rightarrow\pi^*$ transition is near 102.1 kcal/mol (2800 cm^{-1}), and the solvent effects are on the order of kilocalories, thus the broken transition lines. However, the calculated state-to-state

endpoint levels (represented by the horizontal lines), the calculated (1 SD) standard deviations (represented by rotated parentheses), and the experimental data points (represented by asterisks), are all correctly positioned relative to one another on this scale. All of the calculated $\Delta\Delta V_{1W}$ results are in reasonable agreement with experiment. The calculated values are within the range of the dual sets of experimental values for water and methanol, but they blue shift slightly too far in the carbon tetrachloride case. Note that the calculations never give a shift that is too far toward the red when compared to their corresponding experimental values. Also note that the experimental data points are within the reported calculated standard deviations for water and methanol.

One can see at a glance (as per Haberfield's earlier experimental findings) that the $n\rightarrow\pi^*$ blue shift that occurs on transferring acetone from carbon tetrachloride to a dipolar, hydrogen-bonding solvent is due not only to superior solvation of the ground state, but also to the inferior solvation of the excited state in these dipolar hydrogen-bond-donor solvents (Figure 2). This is especially true in the case of methanol (keep in mind that the true shift in carbon tetrachloride is even further to the red than our calculated value).

The calculation of $\Delta\Delta G_s$, represented by the dashed arrow-headed lines, has facilitated the determination of the solvent-equilibrated excited-state solvation energy and its positioning within this scheme. Knowing this also permits the determination of the Franck-Condon radiative emission state solvation energy. The interesting structural changes that occur on solvent equilibration of the Franck-Condon states will be examined in detail later in this paper.

The corresponding (forward/reverse averaged) $\Delta\Delta G_{1W}$ values (Table III, column 6) are 4.03 ± 1.27, 2.70 ± 1.32, and 0.22 ± 0.23 in water, methanol, and in carbon tetrachloride respectively, where again the (1 SD) standard deviations indicate the partial range of free energy values encountered over the ensemble. The $\Delta\Delta G_s$ between the ground *equilibrated* and excited *equilibrated* states (the value obtained from FEP calculations) should always be less blue shifted compared to the $\Delta\Delta V_s$ between the ground *equilibrated* state and *Franck-Condon* excited state (the value analogous to spectroscopic data). This must be the case since on equilibration from the Franck-Condon excited (solvent-strained) state, the overall system free energy optimizes, lowering the excited-state solvation free energy. This behavior is reproduced in all of our calculations, i.e. $\Delta\Delta G_s < \Delta\Delta V_s$.

There are several possible reasons why the FEP calculations (particularly in the methanol and carbon tetrachloride cases) give an average $\Delta\Delta G_s$ which corresponds to a frequency shift slightly further into the blue than the experimental shift. One reason is the neglect of explicit electronic polarizability in the potential function and the related fact that solvent models were designed with enhanced permanent polarities. Molecules are assigned enhanced dipole moments, causing their effective two-body pair potentials to account (in an average way) for the neglect of explicit electronic polarizability. Also, differences in the dispersion interactions between solute and solvent for ground versus excited states were not explicitly modeled. The unavoidable incompleteness in sampling is always an issue, and our comparison of $n\rightarrow\pi^*$ (0-0) calculated difference-energy values with spectroscopic blue shift values reported for $n\rightarrow\pi^*$ (λ_{max}) may have

some bearing on the agreement between theory and experiment.

There is a physically relevant reason involving the relative differences between the ground- and excited-state solute dipole moments when considering the S_1 versus T_1 excited states. For formaldehyde, $\Delta\mu_{(GS\text{-}S1)expt}$ = 0.77 Debyes $\Delta\mu_{(GS\text{-}T1)expt}$ = 1.04 D, and $\Delta\mu_{(GS\text{-}T1)calc}$ = 1.11 D as per experimental measurements or *ab initio* wave functions used here for formaldehyde (see Table I and Methods section). Although no experimental measurement of the excited state dipole moment was found for acetone, we would also expect that $\Delta\mu_{GS\text{-}T1}$(acetone) > $\Delta\mu_{GS\text{-}S1}$(acetone). The *ab initio* calculations do give this trend for the acetone states: $\Delta\mu_{(GS\text{-}S1)calc}$ = 1.26 D, and $\Delta\mu_{(GS\text{-}T1)calc}$ = 1.36 D. On the basis of the above listed $\Delta\mu$'s, an $S_0 \rightarrow T_1$ absorption should then be expected to produce an $n \rightarrow \pi^*$ band which shifts slightly further toward the blue when compared with absorption to an S_1 state. Our calculations reflect this difference of T_1 versus S_1 dipole moment magnitudes in the larger difference energies calculated for the $S_0 \rightarrow T_1$ transition which comes about as a result of the less favorable solvation of the less (di-)polar T_1 state. Although solvent shift data for carbonyl $n \rightarrow \pi^*$ transitions directly to the triplet are not available, the ordinarily spin-"forbidden" $S_0 \rightarrow T_1$ absorption does occur as a weak magnetic dipole transition (via spin-orbit coupling) for formaldehyde, and carbonyl compounds in general[13a,b]. Recall that the spectroscopically measured values (solvent shifts) which we use for comparisons correspond to the $S_0 \rightarrow S_1$ transition. The $\Delta\Delta G_s$ (forward/reverse, averaged) calculated for the $S_0 \rightarrow T_1$ in methanol (2.43 kcal/mol) is apparently slightly out of order with this trend, a result

which may be due to insufficient sampling, and/or use of the different united-methyl VDW parameters as described in the Methods section.

No reliable spectroscopic shift value was found for formaldehyde which is 99.95% diol in aqueous media. Acetone remains 99.8% ketone, i.e. less than 0.2% diol and less that 0.0001% enol in aqueous media (25 °C).[32] On the basis of the absolute magnitudes of their experimentally measured dipole moments (see Table I), and the $\Delta\mu_{GS\text{-}XS}$ given by the calculated dipole moments, one naturally expects a smaller effect in the differential solvation of ground versus excited states for formaldehyde as compared to acetone. It follows that $\Delta\Delta E_{tr}$ should be of smaller absolute value for formaldehyde than for acetone, and this is predicted by the calculations, i.e. for the $S_0 \rightarrow T_1$ transition in water (Table III): $\Delta\Delta V_s$(formaldhyde) $\approx$ 4.4 kcal/mol, $\Delta\Delta V_s$(acetone) $\approx$ 6.2 kcal/mol, and $\Delta\Delta G_s$(formaldhyde) $\approx$ 3.0 kcal/mol, while $\Delta\Delta G_s$(acetone) $\approx$ 4.6 kcal/mol (free energies as per the "slow growth" calculations). Blair et al. argue on the basis of related available data, that the true solvent shift exhibited by monomeric formaldehyde in water is likely to be in the 600-1900 cm^{-1} range. Our calculated value of $\Delta\Delta V_s$, 4.4 kcal/mol (1530 cm^{-1}) is consistent with their estimate.[34a]

The absence of polarizability terms in the molecular mechanical potential, along with the omission of explicit modeling of the differences in the solute-solvent van der Waals interactions for ground versus excited states, accounts for the inability of the simulation to reproduce the red shift of $n \rightarrow \pi^*$ carbonyl transitions in carbon tetrachloride. As mentioned in the introduction, *all* solvents electronically polarize (to varying degrees), facilitating the formation of an electronic transition moment in a solute

and thus contribute toward an n→π* red shift relative to the vapor-phase transition. Also, since the electronically more diffuse excited-state solutes are "larger" and more polarizable, attractive solute-solvent dispersion interactions may be stronger even in the non-equilibrium Franck-Condon excited state, again contributing to a red shift relative to the vapor phase. Because carbon tetrachloride is a nonorienting, nonpolar solvent, these contributions to a red-shift due to the interaction of the solute with the large polarizable chlorine atoms are the major solvent influences on carbonyl n→π* transitions in carbon tetrachloride.

Several carbon tetrachloride models were examined in this study (see Table I). The STO-3G basis set gives a carbon tetrachloride with slightly negative chlorines and a positive carbon, whereas the 6-31G* basis set gives charges of the opposite sign and slightly greater magnitude. The van der Waals parameters used were those of McDonald et al.[33] with slight modifications to the chlorine parameters to ensure proper fit to the heat of vaporization and density for pure carbon tetrachloride (we adopted the formula Jorgensen uses[24] for determination of ΔH_{vap}). These simulations resulted in a energy values which oscillated near 0.0 kcal/mol (Table III). A third carbon tetrachloride model used the permanent point charges (+0.30 Cl, -1.20 C) that McDonald et al. have suggested should reproduce a reasonable octapole moment, estimated as Ω = 15DÅ^2, for carbon tetrachloride.[33] The strong permanent (non-induced) field created by these charges caused the FEP calculations to reproduce even less adequately the red-shift character of the carbonyl n→π* transition in carbon tetrachloride (Table III). A simulation using these same charges but with reversed signs gave approximately the

same results. Thus, using a simple (no explicit polarization) two-body potential we found that neither charge-charge interactions could not account for the experimentally observed red shift. The red shift of the carbonyl n→π* band in carbon tetrachloride appears to be an induction and/or dispersion effect. These results indirectly indicate theoretical support for Bayliss's elucidation[10a] of a plausible mechanism for the polarization red shift, although we feel that the effect of differential dispersion interactions between the ground- and excited-state solutes and the solvent (which Bayliss ignores) may also figure prominently in the production of the n→π* red shift in nonpolar solvents.

Note that induced-field polarization effects are also neglected in the simulations involving the polar solvents. In reality, these effects are operating but are far overshadowed by the presence of stronger dipole-dipole interactions. Also, since water and methanol are much less polarizable that carbon tetrachloride, this red-shift contribution is expected to be smaller in these solvents. If the shift toward the red contributed by a (Bayliss type) polarization response were "added in" as a correction to the results from the calculations presented here, all of the values would shift in the correct direction, toward closer agreement with the (presumedly) more rigourous spectroscopic data. The molecular electronic polarizabilities of the three solvents used in this study go as carbon tetrachloride > methanol > water. In this regard, it is interesting to note that our $\Delta\Delta V_s$ values (taking the acetone $S_0 \rightarrow S_1$ values for water and methanol and the either the 6-31G* or octapole reproducing carbon tetrachloride model) are furthest from the experimental values for the solvent ordering carbon tetrachloride > methanol > water. Notice

(in Figure 2) that across the progression from water to carbon tetrachloride the calulated shift values proceed toward the "blue" end of the range of experimental values, and that the experimental values "drift" out of the range of the standard deviations in this same order. Thus it appears that where our calculations are not in strict agreement with experiment, they are "off" in the sense that they *should* be, in light of our having neglected explicit electronic polarizability and quantum effects.

DESOLVATION OF SOLUTE ON SOLVENT REORIENTATION

During solvent rearrangement about the excited state prior to emission, further alteration of the solute's energy occurs. On release of Franck-Condon orientation strain, the overall system solvation free energy decreases. Commonly, it is assumed that when solvent relaxation occurs around an excited molecule, this brings about more favorable solvation of that particular molecule. However, for systems in which the solute-solvent interaction is mediated primarily by dipole-dipole (orientational) electrostatics, and where the solute is more strongly polar in the ground state than in the excited state, consider the following data:

A diagram of the solute's relative energy levels can be constructed from the absolute potential energies of solvation for the ground, Franck-Condon, and excited-state solutes (<monomer-bulk energy> presented in Table IV). The diagram presented for acetone in water (Figure 3) illustrates quantitatively that, at least in enthalpic terms, desolvation of the Franck-Condon excited-state solute occurs during thermal equilibration after absorbance. Also, at the first instance upon emission, a Frank-Condon ground

state occurs in which the monomer solvation energy is higher (less favorable) relative to the solvent rearranged state, which occurs within picoseconds later. The energy levels in this diagram represent the solute-solvent interaction potential energies during the various stages of an absorption/emission event and not the levels of overall system energy.

For acetone (and formaldehyde which is not shown, but see ref 34a) the radial distribution functions g(O-H) for the carbonyl oxygen and water hydrogen (Figure 4) illustrate that on average the hydrogen of water, which is in position to hydrogen bond with the carbonyl oxygen, rotates back toward the bulk solvent during thermal equilibration of the excited state. Note (Figure 4) the well-defined hydrogen bond which forms between the ground-state solute and water. Acetone's carbonyl oxygen to first shell water oxygen distance is about 3.0 Å on average (origin to highest first shell oxygen peak, longest dashed line) and about 3.1 Å for formaldehyde (not shown), typical for hydrogen bonds. Note that since the first shell water hydrogens are located at about 2.1 Å (solid line), and since the equilibrium O-H bond length for TIP3P waters is 0.96 Å, the O... H-O angle must be near 180 degrees on average, marking the presence of linearly oriented hydrogen bonds between the solute and solvent. This ground state average solute-solvent orientation, optimally oriented for hydrogen bonding and dipolar interactions, applies equivalently to the Franck-Condon excited state. But after solvent reorientation about the excited state, the near-shell hydrogen peak is substantially reduced (Figure 4, short dashes); only a hydrogen bond remnant remains. The number of hydrogen bonds (defined as less than or equal to -2.5 kcal/mol dimer pair potential energy) for each solute state are quantitated in Table IV. In our simulations, acetone

formed 1.9 hydrogen bonds to water on average, and these are (mostly) lost on excitation, leaving 0.50 of a hydrogen bond in the Franck-Condon state, and only 0.18 of a hydrogen bond remaining after thermal equilibration of the excited state. Even though the near-shell solvents reorient to take best advantage of interactions with other (bulk) solvents, a near-shell water's "second" hydrogen still faces toward the excited state carbonyl oxygen a certain fraction of the time.

The strong ground-state dipole (as opposed to the weaker excited-state dipole) confers upon the ground-state solute a greater ability to command the near-shell solvent configuration on the basis of strong solute-solvent interactions. On solvation of the ground state, the solvent sacrifices some attractive solvent-solvent interactions in order to accommodate the dipolar, hydrogen bonding solute. The solute-solvent interaction energy lowers overall system energy, while the average solvent-solvent interaction energy becomes somewhat higher. The solvent becomes orientationally saturated for solute-solvent dipolar and hydrogen bonding interactions during equilibration of the ground state.

On absorption to the Franck-Condon state, the excited solute thus enjoys a solvent configuration which is optimally oriented so that it can take full advantage of any weak hydrogen-bonding or dipolar interactions involving its reduced dipole. Then during equilibration in the presence of the now weakly dipolar, relatively non-hydrogen bonding excited state solute, the system free energy minimizes when the solvent reorients to regain dipolar and hydrogen bonding interactions with other solvent molecules, those interactions which were traded away when the ground-state solute was originally sol-

vated. The solvent relaxation proceeds in order to minimimize the overall system free energy, not to accommodate the excited-state solute. If the acetone excited state were to then undergo a radiative transition back to the ground state, approximately 1.35 (table IV) hydrogen bonds could form in the Frank-Condon emission state which occurs prior to equilibration of the renewed ground state.

The monomer-to-bulk energy distribution function for water (Figure 5) graphically displays this loss of solute monomer to bulk solvent interaction energy on excited-state equilibration. A plot for water-water monomer to bulk energy distribution is also shown in the inset (light dotted line) for comparative purposes. This curve was generated with one solute present but should essentially reflect the pure water energy distribution function. These curves are interpreted as the number of solute-solvent dimers (y-axis) with a particular interaction energy (x-axis) that the solute of interest forms on average. These are solvent-solvent dimers in the case of the pure solvent. The large peaks centered around zero energy are due to the many long-range interactions. The area under the peak in the negative energy region (solid line) reflects hydrogen bonding interactions between the ground-state solute and equilibrated waters. Note that due to optimal solvent orientation in the Franck-Condon excited state (dots), electrostatic interactions are enhanced in the area of -2 to -3 kcal/mol. The thermally equilibrated π^* excited-state solute (longest dashes) curve indicates the loss of strong hydrogen bonds. Only minimal hydrogen bonding or dipolar interactions remain in the area of -2 to -3 kcal (see also Table IV for quantitation.) Finally, in the Franck-Condon emission state (short dashes) some hydrogen bonding is immediately recovered, even in a solvent that has equilibrated about the excited-state solute. Illustrated in these plots is a progressive

loss of hydrogen bonds and dipolar interactions during the transition process from ground state to Franck-Condon state and to equilibrated excited state, and the progressive recovery of these interactions upon emission and reequilibration about the ground state.

In methanol, as in water, the excited-state solute molecule is desolvated relative to the ground-state solute (Figure 6). Again, there is strongly oriented hydrogen bonding between ground-state acetone and the near-shell methanol (see figure 7). The acetone-methanol oxygen-to-oxygen distance is 2.7 Å, with the methanol hydrogen at 1.75 Å, giving a closer hydrogen bond than in TIP3P water. Because the equilibrium O-H bond length for modified-OPLS methanol is 0.95 Å, the O... H-O angle is again linear on average. Solute-solvent hydrogen bonds, numbering 1.13 in the ground state, are diminished to 0.53 in the Franck-Condon excited state, and totally lost on equilibration, as indicated in the Table IV.

The radial distribution functions for acetone in methanol g(O-H), g(O-O), g(O-Me) for ground (Figure 7) versus excited state (Figure 8) illustrate that the hydroxyl group of methanol has made an about-face away from solute-solvent hydrogen bonding configuration during excited-state equilibration. Although coordination number plots have been omitted for clarity, their calculation confirmed that (as indicated in the rdfs) methanol's first shell atoms line up nearest the carbonyl oxygen in the order H, O, Me for the ground state and as Me, O, H in the excited state. This precludes the possibility of the excited state forming even an opportunistic hydrogen bond remnant as occurs in water.

The monomer-to-bulk energy distribution function for acetone in methanol (Figure 9) displays the pair energies for ground-state acetone-methanol (solid), Franck-Condon excited acetone-methanol (dots), equilibrated excited-state acetone-methanol (longest dashes), and Franck-Condon emission state (short dashes), along with the pure methanol-methanol (dotted line in inset) distribution function for comparison. Note that there is formation of some strong hydrogen bonds between ground state acetone and the methanol solvent (peak at about -7 kcal/mol, solid line), as well as some general electrostatic interactions (beginning at about -3 kcal/mol) that are more favorable than those for methanol-methanol (light dotted line in inset). There is a progressive diminishing of these general electrostatic interactions as the monomer dipoles decrease in the series ground-state acetone, methanol, excited-state acetone. The Franck-Condon excited state acetone (dotted in main plot) benefits energetically from the optimally oriented methanol solvent shell. Some weak hydrogen bonding (about -3 kcal/mol) is in evidence (also see Table IV), as well as general solute-solvent electrostatic interactions that are more favorable than those after excited-state equilibration. As would be expected on consideration of the near-shell solvent configuration about the equilibrated excited-state acetone, very little recovery of electrostatic interactions occurs at the Franck-Condon emission state in methanol (short dashes); only 0.18 worth of a hydrogen bond is indicated (Table IV).

From the viewpoint of a polar solvent, acetone appears hydrophobic in its π^* excited state relative to the ground state. We have demonstrated how the solute-solvent potential energy of interaction becomes progressively less favorable on excitation and equilibration of the Frank-Condon excited state. We have also reasoned that we expect

near-shell solvent-solvent potential energy of interaction to grow more favorable on equilibration of the Franck-Condon excited state. A plot of solvent-solvent energy as a function of radial distance out from the center of mass of the solute to the center of mass of the methanol solvent molecules (Figure 10) shows that near-shell methanol solvent-solvent potential energy of interaction is indeed stronger when the excited state (solid line) is being solvated. Note that the average potential energy of solvent-solvent interaction is attained (where V/<V> approaches 1.0) at a radial location nearer to the excited acetone than for the ground-state acetone, illustrating (as has already been discussed) the change in the nature of near-shell solvent-solvent interactions. This also indicates a that a more rigid solvent "cage" structure forms around the more hydrophobic form (excited state) of the solute acetone in methanol. A similar though less pronounced result (plot not shown) was calculated for acetone in water solvent.

SUMMARY AND CONCLUSION

Calculations using a molecular dynamics implementation of the statistical mechanical perturbation method have satisfactorily reproduced the spectroscopically observed trend for the shifting of an alkyl carbonyl compound's $n \rightarrow \pi^*$ transition frequency further toward the blue in progressively more polar solvents. We found that simple point charge modeling of the electronic distributions in the ground and excited states of the solutes acetone and formaldehyde and the use of a classical two-body potential energy function enabled the nearly quantitative determination of shift energies when the predominant solute-solvent interactions were due to non-induced electrostatics (as when these solute excitations were considered in water). This further validates our approach to the development of charge models for molecular mechanical/dynamical simulations of molecules.[21b]

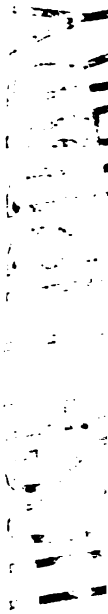
Where previous researchers have studied hydrogen bonding orientation and energetics for these solute-solvent interactions using *ab initio* methods which considered isolated vapor-phase dimers or clusters[11a,b], and determined that this approach is inadequate for studying solvent spectral shifts[34a], we have investigated these issues within the context of "infinite" solvation (using a variety of solvents), where many explicit solvent molecules interact simultaneously with the (ground or excited state) solute of interest. This led to reasonable quantitations of the numbers of hydrogen bonds formed, their strengths, and orientational behaviors where the ground-state, Franck-Condon states, and thermally equilibrated excited-state solutes were examined. We were able to illuminate the relation of solvent restructuring around the excited-state solute to the dif-

ferential solvation of ground and excited states and demonstrated that less-polar Franck-Condon excited-state solutes are actually further desolvated during thermal equilibration in polar solvents. This result concurs with and extends Haberfield's[12] experiments which demonstrated that when a carbonyl solute is transferred from a non-polar to a polar hydrogen-bonding solvent, the blue shift of the $n\rightarrow\pi^*$ band is due not only to superior solvation of the ground state but also in substantial part to the desolvation of the excited state in the polar solvent. A key realization here is that simple ketone groups' excited-state charge distributions render them hydrophobic compared to the ground-state solutes and that their mainly nonpolar character encourages the optimization of solvent-solvent interactions in polar solvents such as water and methanol.

Using this approach, we were unable to reproduce the red shift of acetone's $n\rightarrow\pi^*$ absorption band in the nonpolar solvent carbon tetrachloride. The rationale for this involves the neglect of explicit polarizability in the solvent and solute models. It warrants comment that although the effects of differential dispersion interactions and polarizability-mediated electro-inductive energies are expected to be relatively small in most perturbation calculations, they may contribute significantly in specific situations; the case of $n\rightarrow\pi^*$ shifts in nonpolar solvents is a case in point. Inclusion of these terms in an accurate way is likely to require either a quantum dynamical treatment, or quantum mechanical/molecular mechanical coupling,[34] where the diffuse nature of the excited-state solute can be accounted for, and solvent polarizability is explicitly considered. When differential dispersion interactions between ground and excited states with the solvent contribute heavily to a red shift, the effect could also probably be

reproduced empirically, at least in part (albeit somewhat crudely), by using a simple point charge model along with specifically fit van der Waals potential parameters R^* (or σ) and ε, which are nonidentical for the ground and excited states of the solute. The difference in transition-dipole formation energy might also be estimated from Bayliss's formula[10a] and added as a correction.

Further research, such as presented here for small solutes, which gives a qualitative, or even semiquantitative, description of solvation energetics and structure can naturally be extended to interesting and useful analyses of environmental effects on the electronic excitations of biologically important chromophores.[35]



REFERENCES FOR CHAPTER I

(1) Freeman, D. E.; Klemperer, W. *J. Chem. Phys.* 1966, 45, 52-57

(2) Streitwieser, A.; Kohler, B. *J. Am. Chem. Soc.* 1988, 110, 3769-3772

(3) Scheibe, G.; Felger, E.; Rossler, *G. Ber.* 1927, 60, 1406

(4) Barltrop, J. A.; Coyle, J. D. *Excited States in Organic Chemistry*, pg. 60, 1975, John Wiley & Sons, Ltd.

(5) Halverson, F.; Hirt, R. C. *J. Chem. Phys.* 1951, 19:6, 711-718

(6) McConnell, H. *J. Chem. Phys.* 1952, 20:4, 700-704

(7) Brealey, G. J.; Kasha, M. *J. Am. Chem. Soc.* 1955, 77, 4462-4468

(8) Pimentel, G. C. *J. Am. Chem. Soc.* 1957, 79, 3323-3326

(9) Krishna, V. G.; Goodman, L. *J. Chem. Phys.* 1960, 33:2, 381-386

(10) (a) Bayliss, N. S. *J. Chem. Phys.,* 1950, 18:3, 292-296 (b) Bayliss, N. S.; McRae, E. G. *J. Phys. Chem.* 1954, 58,1002-1006 (c) Bayliss, N. S.; McRae, E. G. *J. Phys. Chem.* 1954, 58,1006-1011

(11) (a)Del Bene, J. E. *J. Chem. Phys. Lett.* 1973, 23:2, 287-291. Del Bene, J. E. *J. Am. Chem. Soc.* 1973, 95:20, 6517-6522. Del Bene, J. E. *J. Am. Chem. Soc.* 1974, 96:17, 5643-5644. Del Bene, J. E. *J. Chem. Phys.* 1975, 62:2, 666-669. Del Bene, J. E. *J. Chem. Phys.* 1975, 63:11, 4666-4671. (b)Blair, J. T.; Westbrook, J. D.; Levy, R. M. Krogh-Jespersen, K. *Chem. Phys. Lett.*, 1989, 154:531-535.

(12) Haberfield, P. *J. Am. Chem. Soc.* 1974, 96:20, 6526-6527 Haberfield, P.; Lux, M. S.; Rosen, D. *J. Am. Chem. Soc.* 1977, 99:21, 6828-6831 Haberfield, P. J.; Rosen, D.; Jasser, I. *J. Am. Chem. Soc.* 1979, 101:12 3196-3199

(13) (a)Clouthier, D. J.; Ramsay, D.A. *Ann. Rev. Phys. Chem.* 1983, 34, 31-58 (b)Turro, N. J. *Modern Molecular Photochemistry*, chapt. 5, 1978, The Benjamin/Cummings Publishing Co., Inc.

(14) Cohen, A. D.; Reid, C. *J. Chem. Phys.* 1956, 24:1, 85-88

(15) (a)Lombardi, J. R.; Freeman, D. E.; Klemperer, W. J. *J. Chem. Phys.* 1967, 46, 2746 (b)Freeman, D. E.; Klemperer, W. J. *J. Chem. Phys.* 1964, 40, 604 (c)Freeman, D. E.; Klemperer, W. J. *J. Chem. Phys.* 1966, 45:1, 52-57

(16) Buckingam, A. D.; Ramsay, D. A.; Tyrrell, J. *Canadian J. of Physics* 1969, 48,1242-1253

(17) Borkman, R. F.; Kearns, D. R. *J. Chem. Phys.* 1966, 44, 945-949

(18) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* 1984, 81, 3684 (Molecular Dynamics with coupling to an external T & P bath.)

(19) Weiner, S. J.; Kollman, P. A.; and Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* 1986, 7:2, 230-252

(20) (a)Binkley, J. S.; Whiteside, R. A.; Raghavachari, K.; Seeger, R.; DeFrees, D. J.; Schlegel, H. B.; Frisch, M. J.; Pople, J. A.; Kahn, L. R. **Gaussian-82**, Carnegie-Mellon University, 1982 (b)Dupuis, M.; Spangler, D.; Wendoloski J. J. **GAMESS**, National Resource for Computations in Chemistry Software Catalog, University of California: Berkeley, CA 1980, Program QG01. Also see: Schmidt, M. W.; Boatz, J. A.; Baldridge, K. K.; Koseki, S.; Gordon, M. S.;

Elbert, S. T.; Lam, B. *QCPE Bulletin*, 1987, 7:115

(21) (a)Singh, U. C.; Kollman, P. A. **G80-UCSF**, QCPE Program #446, 1982. (b)Singh, U. C.; Kollman, P. A. *J. Comput. Chem.*, 1984, 5, 129-145

(22) Pulay, P.; Fogarasi, G.; Pongor, G.; Boggs, J.E.; Vargha, A. *J. Am. Chem. Soc.* 1983, 105:24, 7037-7047 Duncan, J. L.; Mallinson, P. D. *Chem. Phys. Lett.* 1973, 23:4, 597-599

(23) Jorgensen W. L. *J. Phys. Chem.* 1986, 90, 1276-1284

(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* 1983, 79, 926

(25) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* 1977, 34, 1311-1327

(26) Narten, A. H.; Danford, M. D.; Levy, H. A. *J. Chem. Phys.* 1967, 46:12, 4875-4880

(27) (a)Singh, U. C.; Weiner, P. K.; Caldwell, J. W; Kollman, P. A. **AMBER (UCSF)**, version 3.0, 1986, Department of Pharmaceutical Chemistry, University of California, San Francisco. See also: Weiner, P. K.; Kollman P. A. *J. Comp. Chem.* 1981, 2, 287. (b)Seibel, G.; **AMBER** 3.0 Revision A, 1990, Department of Pharmaceutical Chemistry, University of California, San Francisco.

(28) In the AMBER potential function V_{totl} is the potential energy of the system; K_r and r_{eq} are the bond stretching constant and equilibrium bond distance; K_Ω and Ω_{eq} are the bond angle stretching constant and equilibrium bond angle; V_n, n, and γ are the torsional force constant, periodicity of the torsional term, and the phase angle; A_{ij} and B_{ij} are the (Lennard-Jones) non-bond interaction parame-

ters for particles i and j; R_{ij} is the interatomic distance between i and j; q_i and q_j are the atomic partial charges on i and j; ε is the effective dielectric constant.

(29) (a)van Gunsteren, W. F.; *Methods for Calculation of Free Energies and Binding Constants: Successes and Problems*, Proceedings of the Free Energy Perturbation Colloquium, Princeton, Dec. 1987; Weiner, P., ed. (b)Pearlman, D. A.; Kollman, P. A. In: *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, 1989, pp. 101-119; van Gunsteren, W. F. and Weiner, P. K., eds., Escom Science Publishers, Netherlands. (c)Singh, U. C.; Brown, F. K.; Bash, P. A.; Kollman P. A. *J. Amer. Chem. Soc.*, 1987, 109:6, 1607-1614

(30) (a)Ohmine, I.; *J. Chem Phys.*, 1986, 85:3342 (b)Levy, R. M.; Kitchen, D. B.; Blair, J. T.; Krogh-Jespersen, K. *J. Phys. Chem.* (submitted 1989). In particular, Levy, et al. have found that in the solvent (water) equilibration of S_1 excited state formaldhyde, the major part of the relaxation requires only about 200 femtoseconds.

(31) Due to the incompleteness of sampling over the reference and perturbed potential surfaces (because $\Delta\Delta G$ is large for one window), and the biasing of free energy differences due to the one-window FEP equation's treatment of the Franck-Condon surfaces as if they were populated in a boltzman distribution, the one-window free energy values are not reliable when considered in only one perturbation direction, i.e. they must be averaged. The one-direction values should not therefore be used as evidence for any solvent mediated shifting of the $n\rightarrow\pi^*$ 0-0 band in an emission versus an absorbance, nor are they interpretable as the difference in free energy of solvation between a thermally equilibrated state and a Franck-Condon state.

(32) (a)Bell, R. P. *Adv. Phys. Org. Chem.*, 1966, 4, 1 (diol percentages for F and A) (b)Bell, R. P.; Smith, P. W. *J. Chem. Soc.*, 1966, B, 241

(33) McDonald, I. R.; Bounds, D. G.; Klein, M. L. *Mol. Phys.* 1982, 45:3, 521-542

(34) (a)Blair, J. T.; Krogh-Jespersen, K.; Levy, R. M. *J. Am. Chem. Soc.*, 1989, 111:6948-6956. (b)Herman, M.; Berne, B. *J. Chem. Phys.*, 1983, 78:4103. (c)Warshel, A.; Levitt, M.; *J. Mol. Biol.*, 1976, 103:227. (d)Warshel, A.; Russell, S. T. *Q. Rev. Biophys.*, 1984, 17:283. (e)Cieplak, P.; Bash, P.; Singh, U. C.; Kollman, P. A. *J. Am. Chem. Soc.*, 1987, 109:6283. (f)Madura, J. D.; Jorgensen, W. L. *J. Am. Chem. Soc.*, 1986, 108:2517. (g)Bash, P. A.; Field, M. J.; Karplus, M. *J. Am. Chem. Soc.*, 1987, 109:8092.

(35) (a)Warshel, A.; *J. Phys. Chem.*, 1979, 83:1640. (b)Warshel, A.; Weiss, R. M. *J. American Chem. Soc.*, 1980, 102:6218, and 1981, 103:446.

Table I. Interaction Potential Parameters

atom	$Q(S_0)^b$	$Q(T_1)$	$Q(S_1)$	$R^*/2^c$	ϵ
		Acetone			
$\mu_{GS} = 3.13_{calc}, 2.88_{expt}$ D;[a] $\mu_{XS} = 1.77_{calc,T1}, 1.87_{calc,S1}$ D					
O	−0.5812	−0.1944	−0.2114	1.60	0.20
C	0.8177	0.1482	0.1652	1.85	0.12
methyl C	−0.5252	−0.3003	−0.3003	1.80	0.06
H	0.1356	0.1078	0.1078	1.54	0.01
		Formaldehyde			
$\mu_{GS} = 2.66_{calc}, 2.33_{expt}$ D;[d] $\mu_{XS} = 1.55_{calc,T1}, 1.29_{expt,T1}, 1.56_{expt,S1}$ D[e]					
O	−0.4604	−0.1029		1.60	0.20
C	0.4421	−0.2674		1.85	0.12
H	0.0091	0.1851		1.54	0.01
		Carbon Tetrachloride			
atom	Q(STO-3G)	Q(6-31G*)	$Q(\Omega=15)^f$	$R^*/2$	ϵ
C	+0.3560	−0.5872	−1.20	2.5817	0.1017
Cl	−0.0890	+0.1468	+0.30	1.9900	0.2270

[a] The experimental ground-state dipole moment value μ_{GS} (D) is from: Weast, R. C., Ed. *Handbook of Chemistry and Physics*, 55th ed.; CRC Press: Cleveland, 1973. [b] $Q(S_0)$, $Q(T_1)$, and $Q(S_1)$, are atom-centered charges used for the ground state and the excited states, respectively. [c] R* is the van der Waals radius (Å) on the atom, and ϵ is the van der Waals equilibrium well depth energy (kcal/mol) used in pairwise atom–atom interactions. [d] The experimental ground-state dipole moment value μ_{GS} from ref 15. [e] The experimental S_1 and T_1 excited-state dipole moment values are from refs 15 and 16. [f] Charges that reproduce a reasonable octapole moment.[32]

Table II. Molecular-Mechanical Geometrical Parameters

geom param[a]	force constant[b]	geom value
	Acetone	
bonds		
C–O	570.00	1.19
C*–O*	469.00	1.35
C–CT	317.00	1.51
C*–CT	317.00	1.51
H–CT	331.00	1.09
H*–CT	331.00	1.09
angles		
H–CT–H	35.00	108.78
H*–CT–H*	35.00	108.18
C–CT–H	35.00	110.15
C*–CT–H*	35.00	110.73
CT–C–O	80.00	121.77
CT–C*–O*	80.00	111.90
CT–C–CT	70.00	116.46
CT–C*–CT	70.00	120.69
specific dihedrals		
O–C–CT–H	0.82	0.00
O*–C*–CT–H*	1.60	180.00
planar constraint		
X–X–C–O	10.50	180.00
	Formaldehyde	
bonds		
C–O	570.00	1.18
C*–O*	469.00	1.34
H–C	340.00	1.09
H*–C*	340.00	1.08
angles		
H–C–O	40.00	122.14
H*–C*–O*	35.00	113.32
H–C–H	35.00	115.71
H*–C*–H*	35.00	118.65
planar constraint		
X–X–C–O	10.50	180.00
	Carbon Tetrachloride	
bonds		
C–Cl	222.00	1.77
Cl–Cl	222.00	2.89

[a] An asterisk designates a π^* excited state atom type. CT is acetone's tetrahedral methyl carbon. The force constant for C–Cl is interpolated from the Weiner et al. force field; the Cl–Cl force constant is arbitrarily high to ensure symmetric rigidity, although this was not critical since CCl_4 simulations used SHAKE. [b] Force constants are in kcal/mol, bond lengths in Å, angles in deg. The dihedral force constants are actually $V/2$ as input to AMBER.

Table III. Free Energy Perturbation Results

transition[a]	solvent	time$_{SG}$[b]	$\langle\Delta\Delta G_{SG}\rangle$[c]	time$_{1W}$	$\langle\Delta\Delta G_{1W}\rangle$	$\langle\Delta\Delta V_{1W}\rangle$[d]	expt[e]
A → A*(T_1)	H2O	42	4.52	100	4.0 ± 1.5	6.2 ± 1.7	
A*(T_1) → A	H2O	92	-4.64	138	-6.0 ± 1.8	-2.8 ± 1.9	
A → A*(S_1)	H2O	50	3.86	150	3.3 ± 1.1	4.8 ± 1.4	4.46 (5.78)
A(*S_1) → A	H2O	100	-3.80	100	-4.8 ± 1.4	-2.5 ± 1.7	
average$_{S0\rightarrow S1}$			3.83		4.0 ± 1.3		
A → A*(T_1)	MEOH	50	2.32	87	2.7 ± 0.9	5.0 ± 1.8	
A*(T_1) → A	MEOH	50	-3.34	60	-2.3 ± 0.9	-1.3 ± 1.1	
A → A*(S_1)	MEOH			250	1.9 ± 0.9	4.0 ± 1.9	2.40 (4.18)
A*(S_1) → A	MEOH			200	-3.7 ± 1.9	-1.2 ± 1.2	
average$_{S0\rightarrow T1,S1}$			2.83		2.7 ± 1.3		
A → A*(T_1)	CCl_4/STO-3G[f]			39	0.0 ± 0.2	0.0 ± 0.1	
A → A*(T_1)	CCl_4/6-31G*[f]			56	0.2 ± 0.2	0.3 ± 0.3	-1.43 (0.00)
A*(T_1) → A	CCl_4/6-31G*[f]			30	-0.2 ± 0.3	-0.1 ± 0.2	
A → A*(T_1)	CCl_4/Ω=15DA^2			22	0.6 ± 0.4	0.8 ± 0.6	
A*(T_1) → A	CCl_4/Ω=15DA^2			16	-0.5 ± 0.4	-0.2 ± 0.6	
average$_{S0\rightarrow T1}$	/6-31G*				0.2 ± 0.2		
F → F*(T_1)	H2O	71	2.93	166	2.7 ± 1.0	4.4 ± 1.3	
F*(T_1) → F	H2O	50	-3.20	111	-3.6 ± 1.4	-1.8 ± 0.6	
average$_{S0\rightarrow T1}$			3.04		3.2 ± 1.2		
F → F*(T_1)	CCl_4/STO-3G[f]	28	0.04	37	0.1 ± 1.0	0.1 ± 0.2	
F*(T_1) → F	CCl_4/STO-3G[f]	14	0.08				
average$_{S0\rightarrow T1}$			0.05		0.1 ± 0.1		

[a] A indicates acetone ground state, A* the excited state, and similarly F and F* for formaldehyde. [b] TIME$_{SG}$ and TIME$_{1W}$ are total times in picoseconds for slow growth and one-window molecular dynamics. [c] The absolute magnitudes of the forward and reverse $\Delta\Delta G$ values were averaged and are reported as positive. Slow growth $\Delta\Delta G_{SG}$ were averaged with use of time weighting since they are essentially equivalent ensembles. One-window runs $\Delta\Delta G_{1W}$ were simple averaged since they represent nonidentical ensembles and are reported with the standard deviations of their values as collected over the FEP-MD trajectories. [d] $\Delta\Delta V_{1W}$ are differences in potential energy of solvation for ground versus excited states. [e] EXPT is the spectroscopically measured frequency shift converted to energy (kcal/mol). Experimental blue shifts (without parentheses) were determined at λ_{max} for absorption, by Bayliss and McRae[10] for water and CCl_4, see ref 4 for methanol. Shift values in parentheses are from: Suzuki, H. *Electronic Absorption Spectra and Geometry of Organic Molecules*; Academic Press: New York, 1967; p 99. [f] Basis sets from which CCl_4 charges were derived, Ω=15DA^2 indicates charge model when octapole moment is fitted.

Table IV. Potential Energies of Interaction

system[a]	INTN BETW[b]	NUM CRDS[c]	NUM RES[d]	NUM HBOND[e]	HBOND ENERGY[f]	ELECTR ENERGY[g]	VDW ENERGY	MONO/BULK ENERGY
A/H2O	u-n	1600	135	1.90	-3.94	-12.02	-7.56	-19.58
A*-FC/W	u-n	1600	135	0.50	-2.72	-5.79	-7.56	-13.35
A*/H2O	u-n	1195	135	0.18	-2.79	-3.02	-7.72	-10.74
A-FCE/W	u-n	3195	135	1.39	-3.67	-5.82	-7.68	-13.50
A/H2O	n-n	2795	96	3.32	-4.40	-23.25	3.06	-20.19
A/MEOH	u-n	3300	79	1.13	-5.50	-10.13	-6.95	-17.08
A*-FC/M	u-n	3300	79	0.53	-3.04	-4.87	-6.95	-11.82
A*/MEOH	u-n	3000	80	0.01	-2.90	-1.71	-7.92	-9.63
A-FCE/M	u-n	3000	80	0.18	-3.02	-2.91	-7.92	-10.83
A/MEOH	n-n	6300	61	1.96	-5.73	-16.33	-1.52	-17.82
A/CCl_4	u-n	920	61	0.00		-0.47	-10.64	-11.11
F/H2O	u-n	550	122	1.54	-2.98	-8.16	-4.30	-12.46
F*/H2O	u-n	418	133	0.02	-2.35	-2.35	-3.81	-6.16
F/CCl_4	u-n	12	107	0.00		0.16	-6.36	-6.21

[a] All of these values are the result of analysis of configurations saved from the ensembles generated during the calculation of free energy values in Table I. For the system A*-FC/W, acetone's excited T_1 state molecular mechanical parameters are superposed onto the ground-state geometry in an ensemble generated by equilibrating a ground-state acetone in water; this is the Franck-Condon (F-C) excited state. A-FCE/W is the F-C emission state for acetone in water. Similarly for A*-FC/M and A-FCE/M in methanol. All other systems are thermally equilibrated solutes. The CCl_4 charges were derived from the 6-31G* basis set for A/CCl_4, from STO-3G/esp for F/CCl_4. [b] INTN BETW indicates whether the interaction examined is between solute-solvent (u-n), or solvent-solvent (n-n). The n-n data are the average of values of the two ensembles, with either A or A* present. [c] The number of saved coordinate sets over which the table quantities were averaged. The time increment between collection of ensemble configurations varied from every 20 to 50 steps or approximately every 0.05 to 0.10 ps. [d] NUM RES is the average number of solvent residues per configuration that are within the nonbond cutoff sphere and thus directly interacting with the monomer of interest. [e] NUM HBOND is the average number of hydrogen bonds formed between a monomer and the solvent (having dimer pair interaction energy ≤ -2.5 kcal/mol). [f] HBOND ENERGY is the average energy of H bonds. [g] ELECTR ENERGY is the electrostatic component, and VDW ENERGY is the van der Waals component, which sum to give MONO/BULK ENERGY, the average potential energy of interaction between a monomer and the bulk solvent.

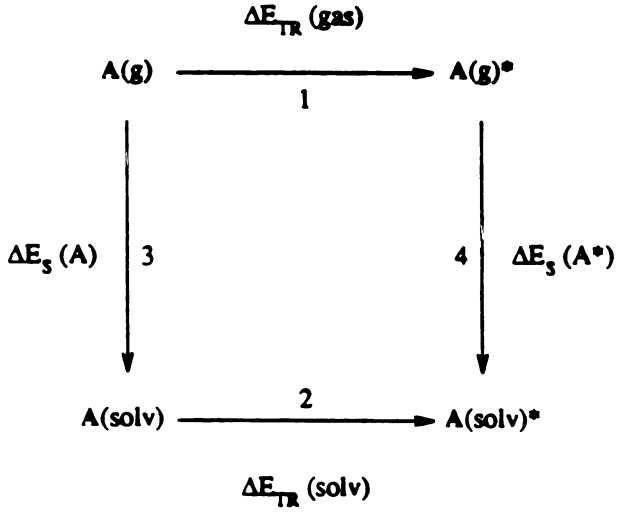


Figure 1. Thermodynamic cycle energy perturbation. ΔE_s is solvation energy. ΔE_{tr} is electronic transition energy.

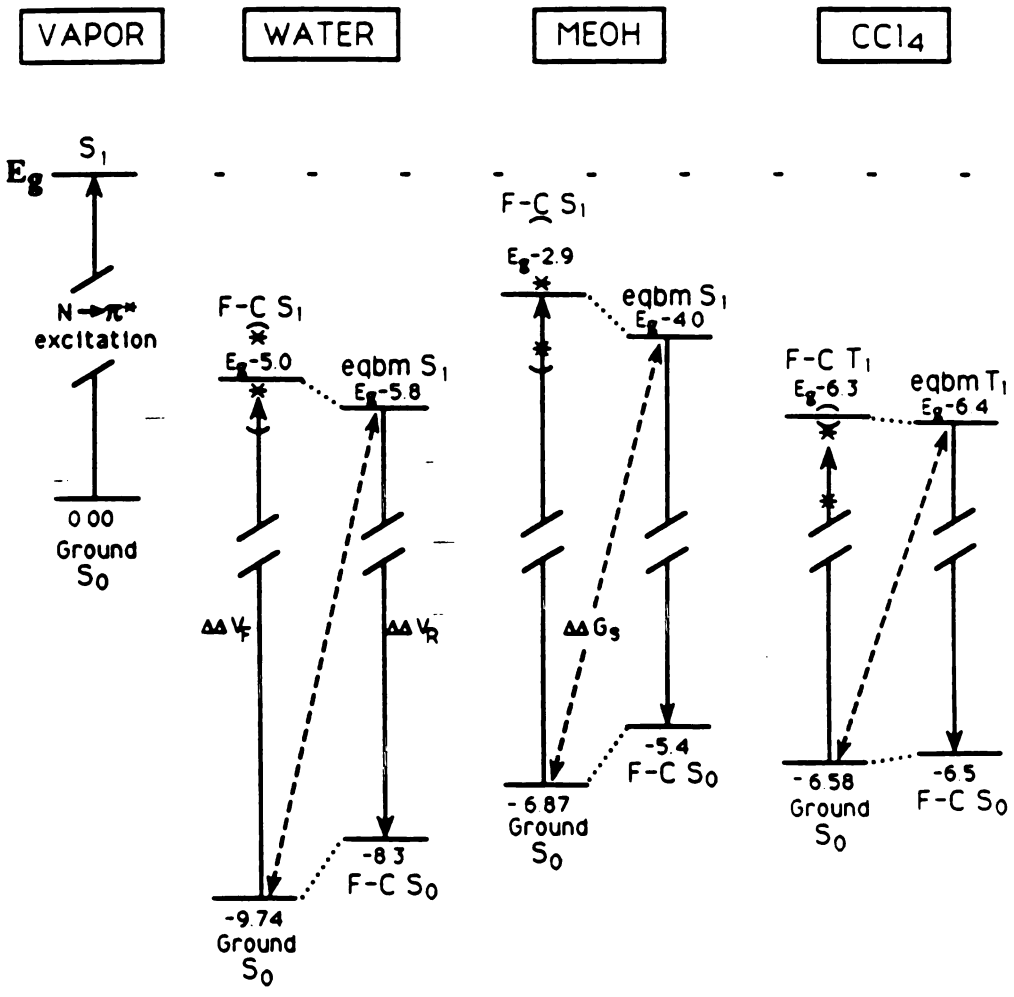


Figure 2. Differential solvation of acetone's ground vs n → π* excited state in three solvents. All values listed in the figure can be arrived at from data in Table III. E_g is the excitation energy in the gas phase. The S_0 ground state is taken as the 0.0 energy level. Energies lower than 0.0 are due to solvation of ground states; energies lower than E_g (but above 0.0) are due to solvation of excited states. The states are referred to as ground, F-C (Franck-Condon), or eqbm (solvent equilibrated). The symbols $\Delta\Delta V_F$ and $\Delta\Delta V_R$ refer to the forward and reverse (absorption and emission) shift energies and are placed beside the transition line for which the shift is calculated. The lowermost asterisks designate the (presumably) best spectroscopic data; upper asterisks designate alternative spectroscopic data. Rotated parentheses indicate ±1 standard deviation from calculated values. Also see text.

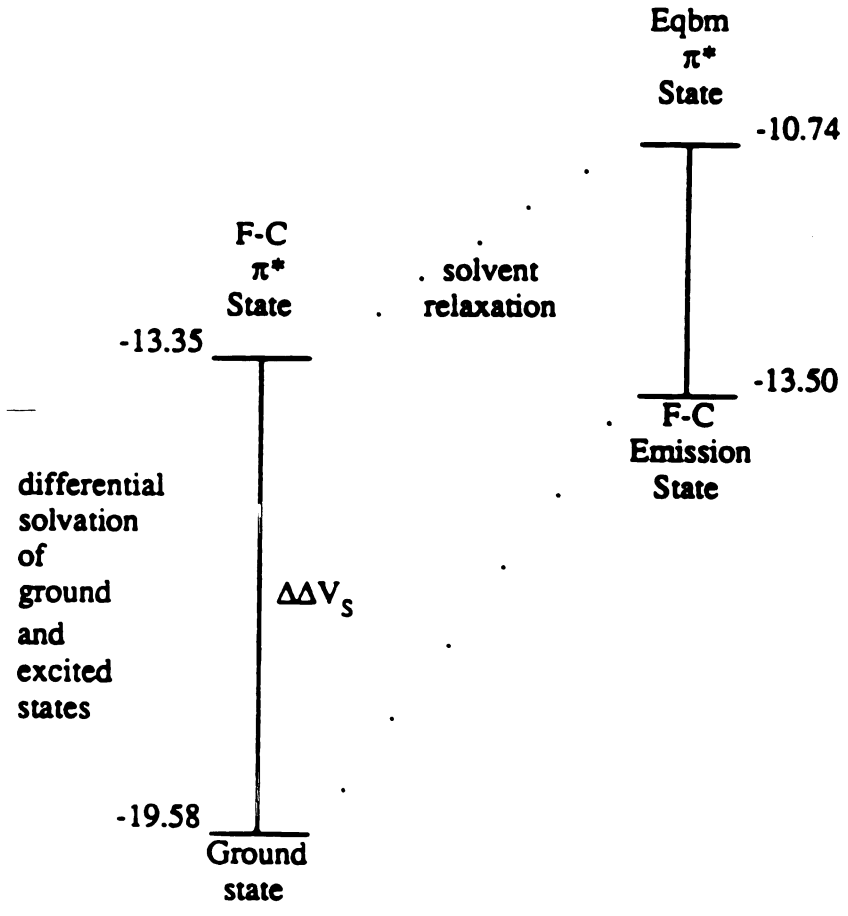


Figure 3. Solute–solvent interaction energies $\Delta\Delta V_s$ for acetone $S_0 \rightarrow T_1$ in TIP3P water.

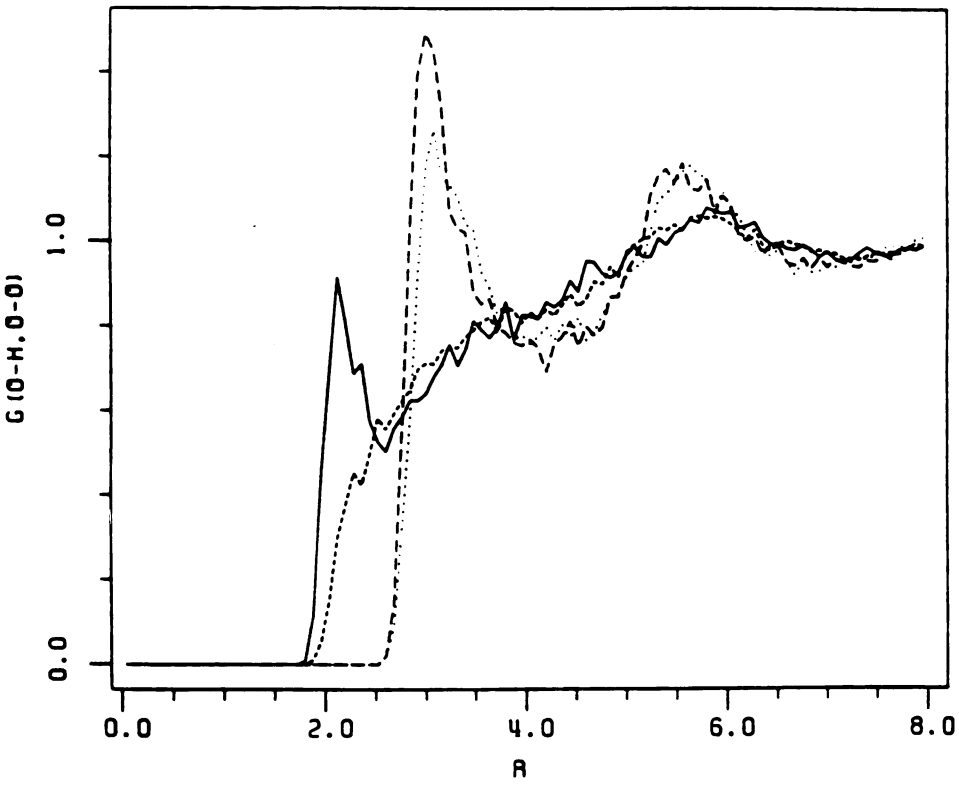


Figure 4. Acetone–water site–site radial distribution functions. Ground state: g(O–H), solid line; g(O–O), long dashes. $\pi^*(T_1)$ excited state: g(O–H), short dashes; g(O–O), dots. R is the distance in angstroms.

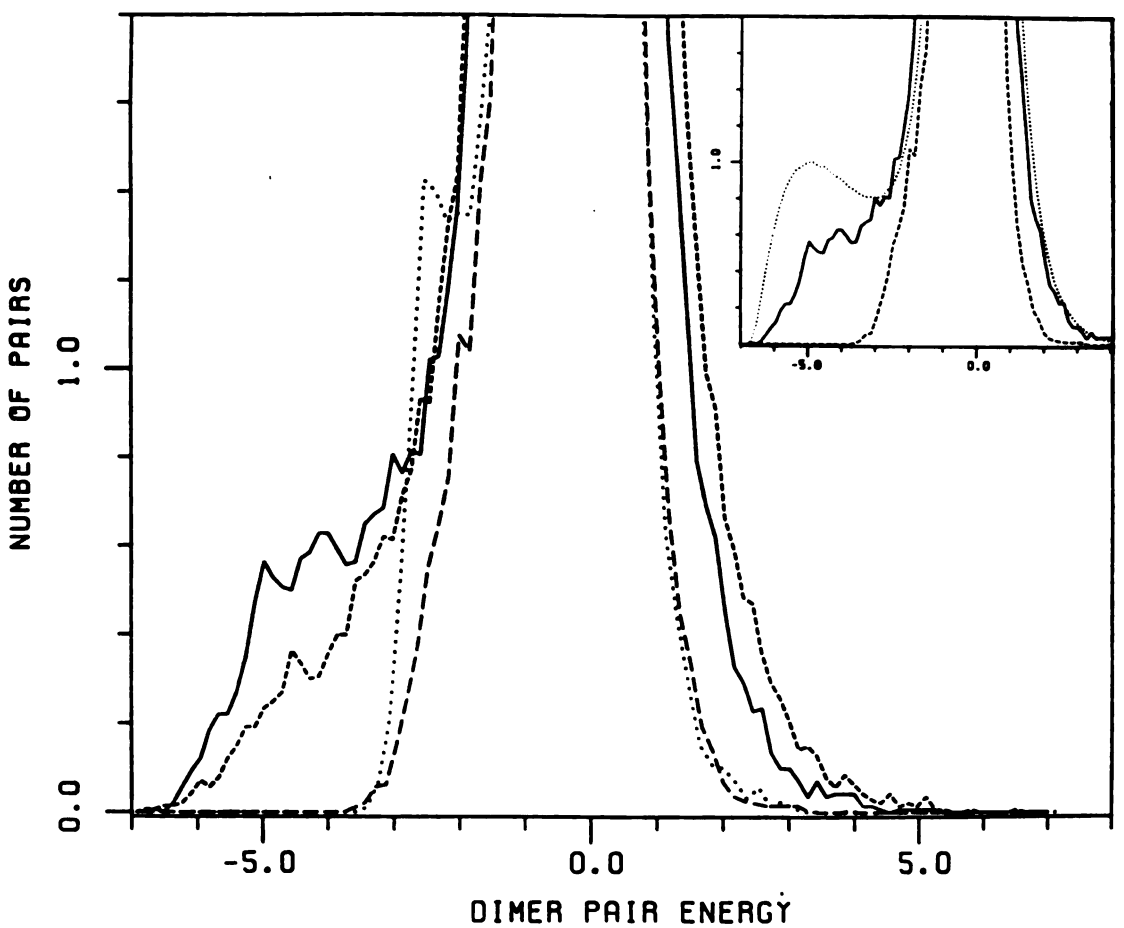


Figure 5. Acetone–water dimer energy distribution functions: equilibrated ground state, solid line; Franck-Condon T_1 excited state, dots; equilibrated T_1 excited state, longest dashes; Franck-Condon emission state, short dashes. Inset: water–water distribution function, light line; others are the same as above.

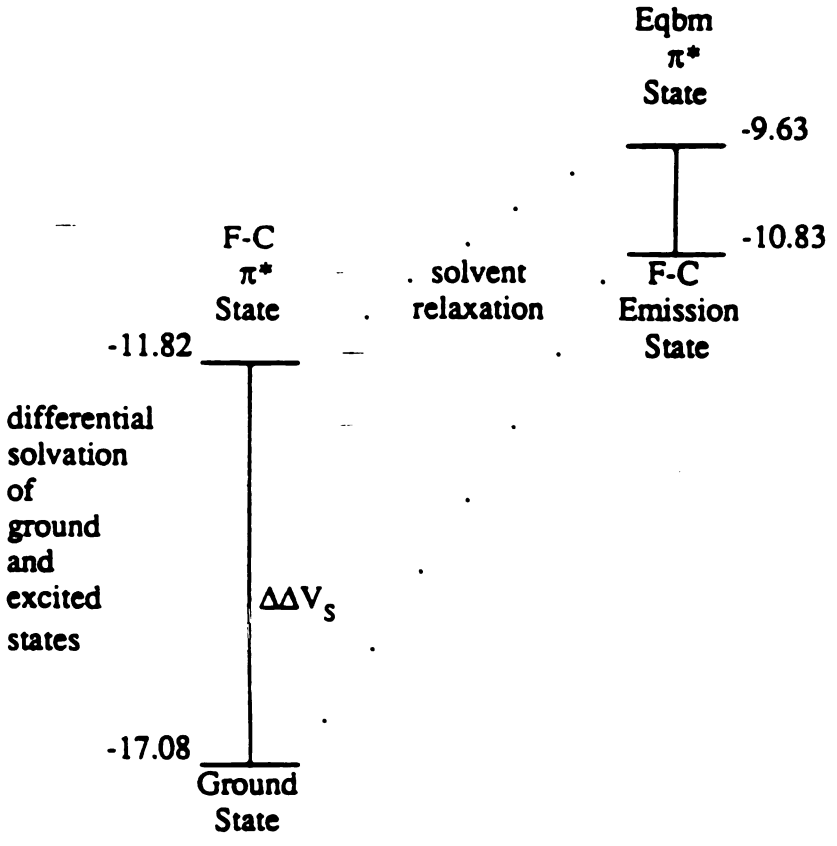


Figure 6. Solute–solvent interaction energies $\Delta\Delta V_s$ for acetone $S_0 \rightarrow T_1$ in modified-OPLS methanol.

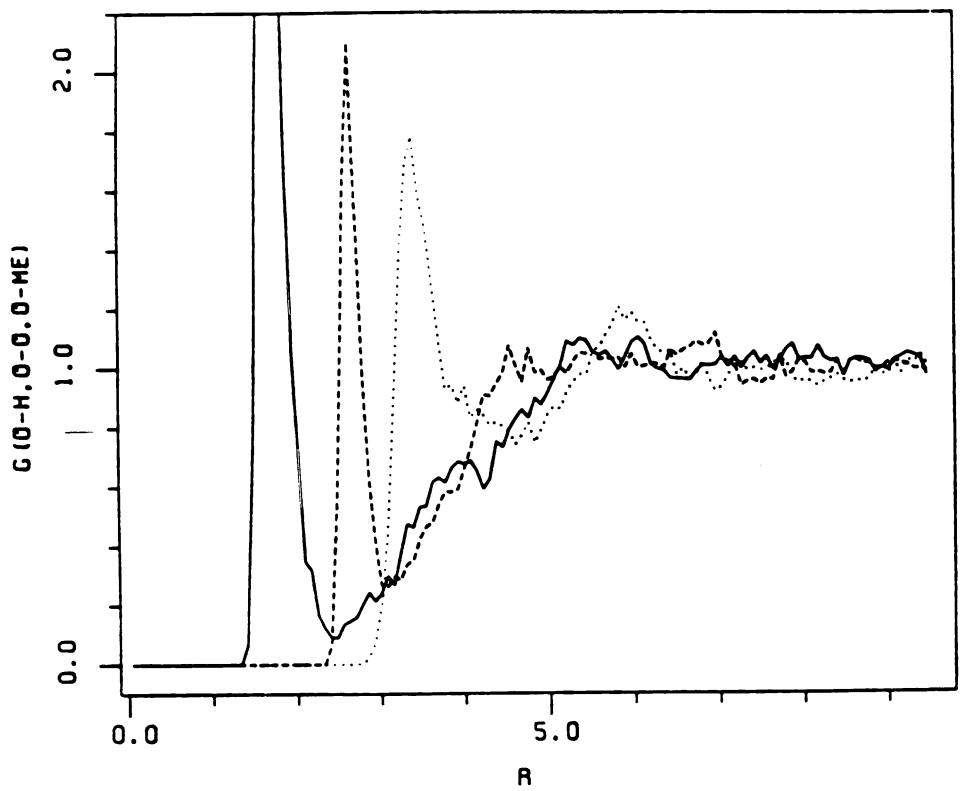


Figure 7. Ground-state acetone–methanol site–site radial distribution functions: g(O–H), solid line; g(O–O), dashes; g(O–Me), dots. R is in angstroms.

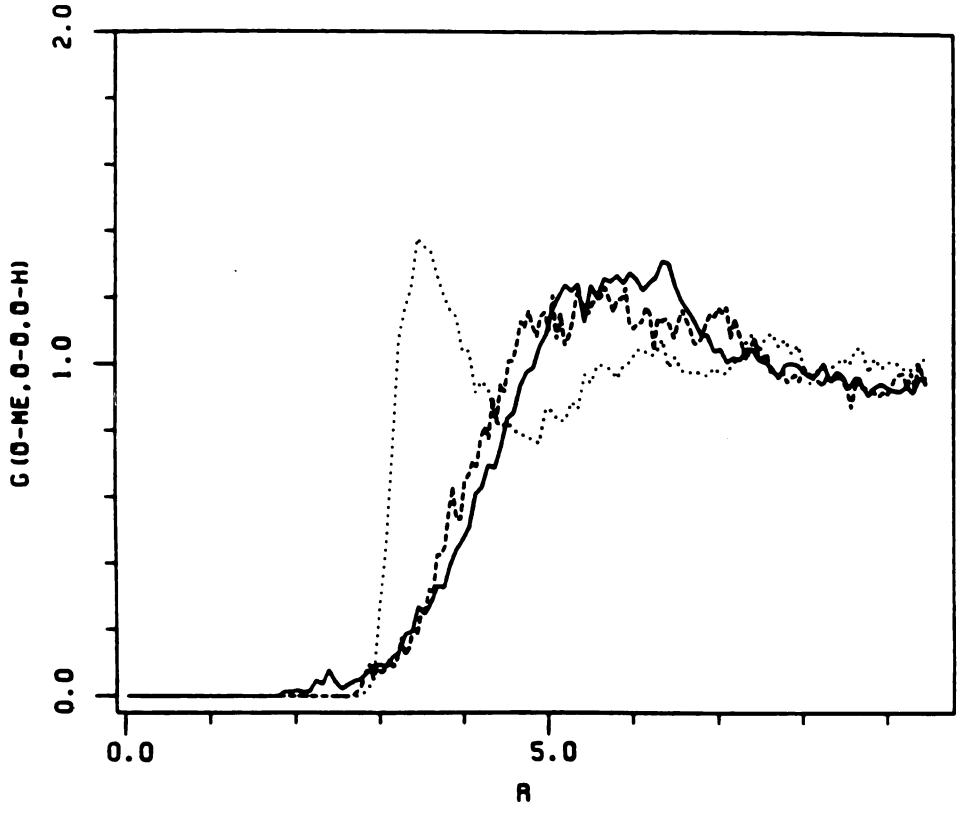


Figure 8. T_1 excited-state acetone–methanol site–site radial distribution functions. Line key is same as in Figure 6.

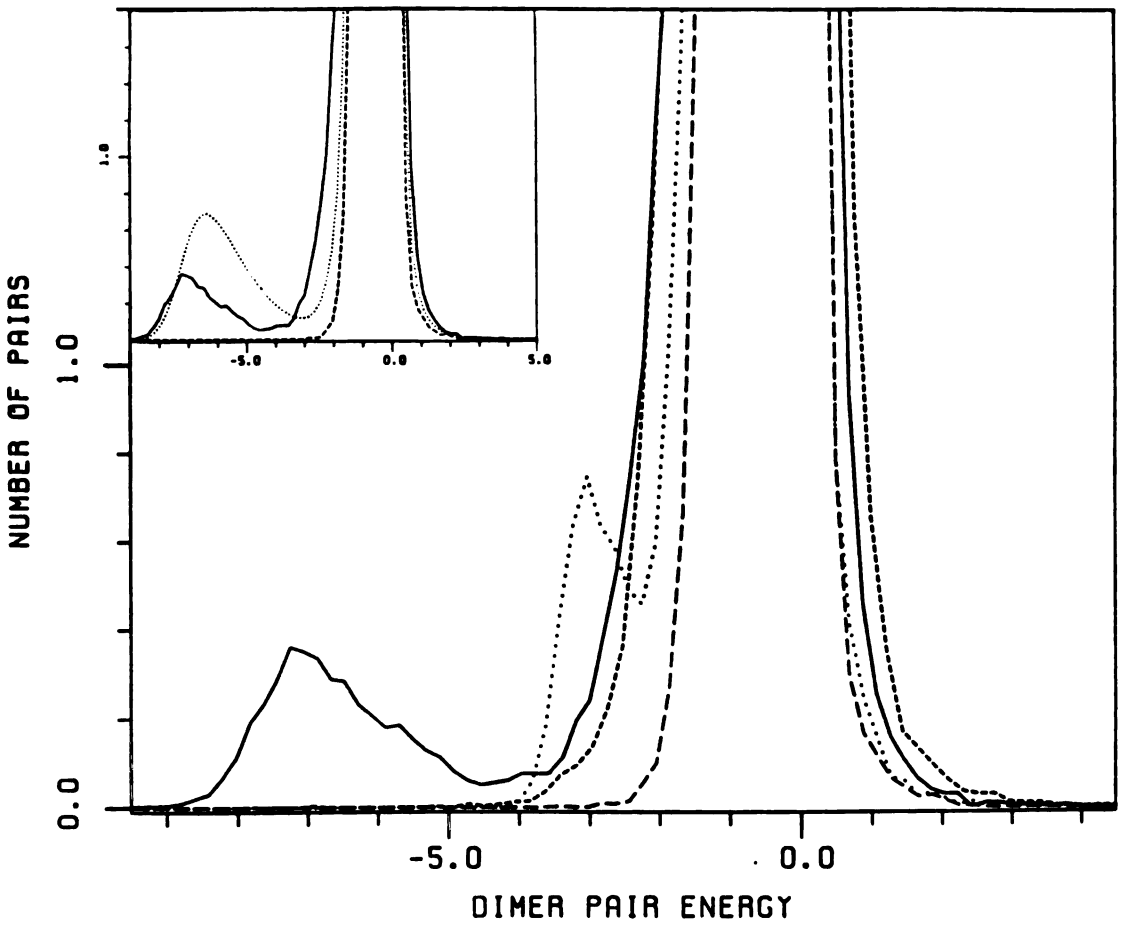


Figure 9. Acetone–methanol dimer pair energy distribution functions: equilibrated ground state, solid line; Franck-Condon T_1 excited state, dots; equilibrated T_1 excited state, longest dashes; Franck-Condon emission state, short dashes. Inset: methanol–methanol distribution function, light line; others are the same as above.

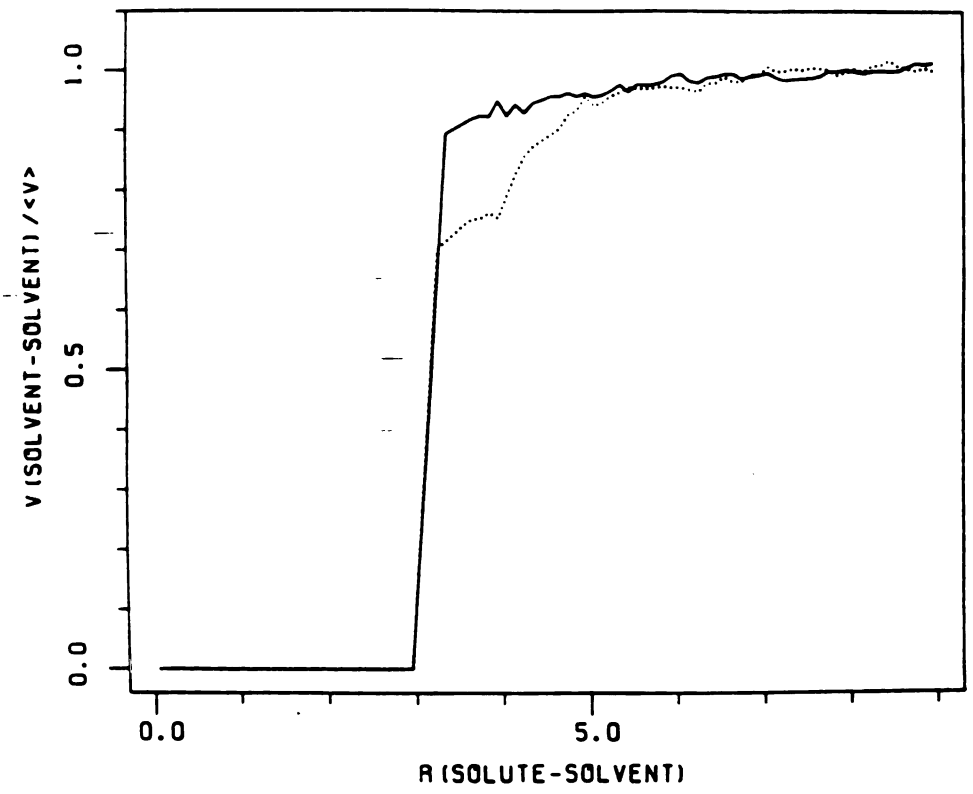


Figure 10. Acetone–methanol solvent–solvent potential energy of interaction (normalized) as a function of radial distance out from the center of mass of the solute: equilibrated ground state, dots (average over 3300 configurations); equilibrated T_1 excited state, solid line (average over 3000 configurations). *R* is in angstroms.

CHAPTER II

A Comprehensive Investigation of Structure and Dynamics in 1-Octanol and Water-Saturated 1-Octanol, Molecular Dynamics and Free-Energy Perturbation Studies.

CHAPTER II. LIQUID 1-OCTANOL AS A BIOPHASE ANALOG

A detailed understanding of the interactions of solutes such as drugs, metabolites, ionophores, and peptides with membraneous body tissues could help to direct rational drug design strategies. It has been shown that some common clinical drugs can interact very specifically with biomembranes. For example, Herbette, et al.[1] have shown that "structure-function activity relationships of drugs that bind to certain membrane-associated receptors must take into account the local membrane bilayer environment where the binding event occurs." A *membrane bilayer pathway* for the binding of lipophilic 1,4-dihydropyridine calcium channel blockers is discussed: "The drug first partitions into a well-defined, energetically favorable location, orientation, and conformation in the membrane bilayer before laterally diffusing to an intra-bilayer [protein] receptor binding site". Specific binding occurred (for propranolol) via such a pathway with an overall "on-rate" which is 1000 times faster than for the aqueous route[2]. Herbette is demonstrating the important role of specific interactions between certain drugs and the phospholipid components of the bilayer.

As experimentalists continue to gain further insight into the structure, dynamics, and specific interactions of drugs with biological membranes, it is becoming evident that there are issues which could be probed using theoretical methods. Computer simulations could provide a valuable source of molecular-level information concerning interaction specificities. But, even with current supercomputers it is infeasible to simulate dynamic membrane-bilayer/solute systems at the level of molecular detail. It

would be beneficial to first examine a simpler model system possessing molecular traits which are analogous to those of biomembranes and proteins.

Given the fundamental role that octanol plays in drug partitioning studies and the fact that it is probably the best simple model for an amphiphilic lipid, it is surprising that no definitive molecular modeling simulations have been carried out on this system. There have been several molecular-level simulations involving alkanes, alkanols, or fatty acid chains, or micelles [3-7], and at this point even several short simulations of small bilayer "patches" [8-12]. The primary goal of this study was to reproduce the structural and thermodynamic properties of liquid 1-octanol and to provide a complete molecular level characterization of its polymeric hydrogen-bonding, hydrophobic clustering, internal micro-structure, and dynamics. An appropriate atom-level model was developed for use in carrying out molecular dynamics simulations.

Internal Structure Of 1-Octanol

The octanol/water partitioning system has become the standard reference for investigating the compartmental distribution characteristics of pharmaceutical agents [13-15]. It is also sometimes used to predict pharmacokinetic characteristics of drug compounds in the biophases. Why does octanol serve as such a remarkably good experimental model for reproducing the partitioning behavior of drugs in body tissues? Octanol is commonly regarded as a simple isotropic bulk solvent, essentially hydrophobic or "oily". But, water/oil partitioning studies have shown that when a strict hydrocarbon (instead of say octanol or olive oil) is chosen as the "oil" phase, correlations

between partitioning behavior and biological activity breakdown[13,16,17]. 1-Octanol shares the traits of amphiphilicity and hydrogen-bonding capability with the phospholipids and proteins found in biological membranes. It is a structural analog for phospholipids in particular, with its polar "headgroup" and flexible nonpolar "tail".

In computer simulations it is important to accurately model these hydrocarbon chains as they are mediators of structural order and viscosity. The chains play an important role as the hydrophobic component in the anisotropic partitioning environment. The polar hydrophilic alkoxyl group also mediates structure and dynamics by virtue of its capability to donate, accept, and exchange temporal hydrogen bonds. 1-Octanol monomers associate to form a wide diversity of structures, certain of them with greater frequency than others. These MD simulations should help to clarify the means by which the amphiphilic nature of the octanol molecule gives rise to distinctive internal structuring in the media, enabling it to serve as a biophase model.

Dynamics And Dielectric Response

The internal dynamics of octanol media are also of interest as pharacokinetic parameters involving transport are sometimes predicted from octanol/water partitioning studies[18,19]. Dielectric response (microwave spectroscopy) and infrared studies first indicated the presence of hydrogen-bonded polymeric entities in octanol media. Experimentalists predicted the presence of long linear open chains of varying lengths and the possibility for multimeric closed-ring structures as well. Is there a link between octanol's local structural anisotropy and its dielectric response behavior?

Exploration of this question was facilitated by implementing the reaction-field (long-range electrostatic modeling) methodology[20] within AMBER's MD module, ensuring that simulated octanol qualitatively mimics the dielectric traits of the macroscopic medium. 1-Octanol has three (frequency) dispersion regions. With the aim of clarifying and possibly confirming the various mechanisms proposed by experimentalists to account for octanol's dielectric behavior, time-correlation analyses were performed on MD trajectory data. The results are appraised with respect to the various explanations offered by experimentalists for the multiple frequency dielectric responses observed in octanol. Some new light is shed on the molecular events responsible for the less well-understood lower-frequency dielectric response phenomena in these long-chain amphiphiles.

Solute Partitioning In Binary Octanol/Water Media

Molecular dynamics simulations of a water-saturated 1-octanol biophase-analog system were carried out. The physical properties of the media were well reproduced. These studies further reveal molecular-level details of the structural micro-hetergeneity and dynamics of this binary media. The important role played by saturating waters in octanol's internal structure was made clear on analysis of the MD trajectory.

The feasibility of reproducing octanol/water partition coefficients via the free energy perturbation (FEP) methodology was examined. A coarse-grain parallel FEP computing regime was successfully instituted and demonstrated excellent promise in this regard. The difference in octanol/water transfer free energies were calculated for

two small solutes. Agreement with the experimentally determined value was excellent.

What is the structure of octanol/water media surrounding amphiphilic solutes? Partitioning solutes were found to interact specifically with octanol molecules as well as with hydration waters. Solute-solvent interactions are mediated by regions of preferentially hydrophilic versus hydrophobic solvation, such as occurs when a solute passes across an aqueous/biomembrane interface. Detailed quantitative information is obtained regarding the specific molecular-level interactions responsible for differences in the partition coefficients of these molecules.

Statistical Convergence

It was a challenge to ensure convergence in octanol systems for both purely liquid properties and interactions with solutes. A pure octanol or octanol/water system converges much more slowly than say pure water because it is heterogeneous, viscous, and inherently more slowly reorientating. And to ensure the generation of sufficient statistics, a binary media must undergo several cycles of mixing, redistribution, and general turnover. Solvation phenomena are naturally complicated by the more complex media. Because of this concern regarding convergence, unusually long MD simulations, into the nanoseconds, were required. Simulations involving water-saturated octanol with small solute molecules should provide a reference point for future calculations involving more slowly relaxing membrane bilayer model systems.

Partitioning, Physicochemical Properties, And Drug Activity

At the heart of the rational drug design strategy is the relationship between chemical structure and biological activity. Often, a drug's biological activity can be understood as a summation of the effects of the individual physicochemical parameters of the molecule. Each property of a drug molecule, e.g. hydrophobicity, electron density, polarity, polarizability, steric characteristics, conformational freedom, hydrogen bonding capability, unique nature of substituents, etc, can be assigned a value which correlates with its contribution to the drug's activity. In the study of quantitative structure-activity relationships the substituent hydrophobicity constant Π is based on a drug's partition coefficient and is commonly correlated with overall biological activity. $\Pi = \log P_X - \log P_H$, where P is (often) the octanol/water partition coefficent, X is a variable substituent, and H indicates the hydrogen atom of a nonsubstituted form.

The partition coefficient P of a solute, or drug specifically, is defined as the ratio of drug concentrations in two phases, normally an organic and an aqueous phase:

$$P = K_P = \frac{[drug]_{organic}}{[drug]_{water}}$$

Because partition coefficients would be difficult to determine in living systems, they are usually measured *in vitro*, using an octanol organic phase (substituting for tissues or cell membranes) and water or a pH-buffered aqueous phase (substituting for blood or aqueous compartments). The Π molecular descriptor is most useful in comparing the relative differences in correlated pharmacokinetic properties between the molecules in a

cogener series.

Pharmacokinetic behavior is equally as important in understanding a drug's total pharmacological activity as are receptor-specific interactions. A drug's *in vivo* partitioning behavior is a determining factor in the pharmacokinetic processes of absorption, transport, distribution, and availability. Since the interaction of a drug with a receptor system can occur only after the drug reaches its site of action, the drug's potential pharmacological activity is thereby governed by its partitioning characteristics. It is not surprising that the partitioning behavior of a drug so strongly influences its biological activity, as often the drug must pass through many "layers" of cellular and organellar membranes in order to reach its site of action. Drugs which are essentially identical in terms of their specific receptor recognition interactions are by no means necessarily equivalent in terms of their bioavailability.

Methods for predicting or measuring drug-receptor binding and action do not foretell toxic, metabolic, or pharmacokinetic properties, all of which are paramount in the development of a therapeutically successful drug. Rational drug design pursuits may benefit from a consideration of the molecular-level physicochemical aspects of drug distribution. In the future, an understanding of these drug-biophase interactions might suggest possible molecular modifications for improving the overall efficacy of a drug, modifications which manipulate pharmacokinetic characteristics without attenuating receptor binding.

PARAMETERS AND METHODS FOR FLEXIBLE LIQUID 1-OCTANOL

At the outset of this study, the most successful simulations of molecular liquids using a potential energy (PE) function similar to AMBER's were those using Jorgensen's OPLS models. In particular, Jorgensen published several systematic studies tracing a development to the OPLS parameters for liquid short chain alkanes and alcohols[21,22]. Hydrocarbon molecules were modeled using the "united atom" approach, where alkyl hydrogens are included only implicitly. The presence of hydrocarbon hydrogens is accounted for by parameterizing methylene or methyl groups as oversized single center CH_n "united atom" units. The appropriate fourier torsion terms are added to model the hydrogen-to-hydrogen 1-4 interactions mediating the H-C-C-H dihedral rotations. In view of the OPLS model's economy in reducing the number of degrees of molecular freedom, of and its success in reproducing experimentally observable thermodynamic quantities, OPLS hydrocarbon and alkoxyl models were adapted for use in AMBER MM/MD. But some notable parameter modifications were required.

Transformation of OPLS-MC Parameters to AMBER-MD

Jorgensen used rigid bonds and angles in his derivation of potential energy function parameters. In the Monte Carlo (MC) simulations he used to perform the ensemble sampling, the molecular degrees of freedom were translational (3 axes), rotational (3 axes), and torsional rotations about dihedral angles. But AMBER also incorporates flexible bonds (optionally) and angles (always), whose "vibrations" are modeled as har-

monic oscillators, permitting significant intramolecular deformations in the presence of neighboring molecular forces. Also, the OPLS potential energy function differs from AMBER's in its use of *different* sets of nonbonded (Lennard-Jones type) parameters σ and ε for inter- versus intramolecular interactions (where intramolecular involves atoms separated by more than three bonds: 1-4, 1-5, ... 1-N interactions). AMBER's potential energy function uses the *same* van der Waals (VDW) parameters for *both* inter- and intramolecular nonbonded interactions, but provides a scaling factor which can be used to attenuate intramolecular 1-4 and higher 1-N interactions. Also, although both OPLS and AMBER potential energy functions can incorporate zero to five component fourier terms as torsional potential refinements, the fourier parameter values are naturally coupled to those of the intramolecular nonbonded terms. Also, the influence of intramolecular deformability must be considered in parameterizing AMBER/MD. In view of these differences in molecular models and functional representations, one would not expect OPLS/MC parameters to transfer directly to AMBER/MD without modification. Indeed, initial MD simulations using *unmodified* OPLS parameters did not reproduce the correct experimental liquid densities and heats of vaporization, contrary to the successful OPLS/MC results.

Fortunately, the intermolecular nonbonded and electrostatic interaction terms are the parameters governing molecular packing volumes and interaction energies. These terms are most responsible for producing correct liquid behavior, and are directly analogous in both force fields. They were adopted unchanged from the OPLS to the AMBER force field with a single exception. The nonbonded well-depth parameter ε for the methyl group was slightly decreased for long chain hydrocarbons. While σ is identical

for OPLS methylene and methyl groups (where $\sigma = 2^{1/6}R^*$), ε(methyl) = 0.175 and ε(methylene) = 0.118. Here, ε(methyl) = 0.150 gave a better conformer distribution and improved liquid density and heat of vaporization for hexane.

For the adoption of intermolecular parameters to succeed, it was necessary to fit AMBER's intramolecular terms to those of the OPLS potential energy function (which incidently were designed to faithfully reproduce the energies of 91 monomer conformational states by mimicking Allinger's MM2 force field, which very closely reproduces experimental data)[21,23]. For the organic liquids simulated in the current study (hexane, octane, 1-octanol, and 1-octanol/water), the OPLS geometries were adopted, and AMBER's bond stretching and angle bending force constants were retained. However, in the production simulations, bonds were constrained via the SHAKE algorithm[24]; angles remained free to bend. After systematic trial searches over many parameter combinations in many liquid MD simulations, the best values for the intramolecular VDW scale factor (1/7.4) and new torsional potentials were determined. The fitting criteria were the liquid density, heat of vaporization, and conformational distribution.

Fitting Criteria

The density ρ is available from a constant temperature, constant pressure (NPT ensemble) MD simulation as the molecular weight of N molecules divided by the overall average volume of the periodic box.

$$\rho = \frac{Mass \cdot N}{<Volume>_{box}} \qquad (1)$$

The heat of vaporization can be determined as (recall from thermodynamics that ΔH =

$\Delta E_i + P\Delta V$),

$$\Delta H_{vap} = \left[E_{inter}(g) - E_{inter}(l) \right] + \left[E_{intra}(g) - E_{intra}(l) \right] + RT \qquad (2)$$

where E_{inter} is the internal energy of the interacting liquid molecules (the intermolecular VDW and electrostatic terms in AMBER), E_{intra} the intramolecular energy (bonds, angles, torsions, and intramolecular VDW/electrostatic terms) for the liquid is subtracted from the gas value, R is the gas constant, and T the temperature. Assuming an ideal gas, $E_{inter}(g)$ for the gas phase is approximated as zero, and the RT term substitutes for $P\Delta V$, the work of expansion, in the constant pressure enthalpy change equation. $E_{intra}(g)$ for the gases were determined from an average of multiple extensive molecular dynamics simulations of the isolated monomers summing to 1 or 2 nanoseconds time span. After initial liquid relaxations, liquid-phase fitting criteria were evaluated by averaging over (multiple) MD runs of 100 picoseconds for hexane, 200 psec for octane, and 300 psec for octanol.

Molecular Systems

The SPC water model[25] was used for the internal waters in water-saturated octanol. The mole fraction of water X_{wat} in fully water-saturated octanol is temperature dependent. X_{wat} is 0.245 at 20°C, 0.255 at 25°C, 0.290 at 40°C, and 0.366 at 75°C [26]. Simulations of water-saturated octanol analysed for structural and dynamic properties were carried out at 40°C with X_{wat} equal to 0.250 (228 octanols, 76 waters). Another

simulation at 25°C contained 125 octanols and 42 SPC waters, with X_{wat} equal to 0.251. The systems were approximately fully water-saturated in each case. The octanol/water system was not used in parameter fitting.

Pure liquid hydrocarbon chains (hexane and octane) were parameterized first, avoiding the complication and interference of a hydroxyl group. A previous MD simulation[27] of rigid 3-point methanol demonstrated the transferability of OPLS hydroxyl parameters from MC to MD (see Chapter I in this dissertation). Initial configurations were prepared by placing molecules at evenly spaced lattice points within the appropriate volume space, then randomly displacing within a tolerance zone (0.5 Å) along all three axes, and randomly rotating (0 to 360 degrees) about each axis. Periodic boxes of 125 molecules (hexane, octane), or 250 molecules (octanol), or 304 molecules (water-saturated octanol), were used in the fitting simulations. The NTP (constant temperature (298K), pressure (1 bar), and composition) MD systems were minimized, then extensively relaxed. Temperature and pressure were held constant via the Berendsen algorithm for coupling to an external bath[28]. Relaxation proceeded until monotonic limiting values were reached for density, internal energy, temperature, and pressure. The same relaxation criteria were followed for 40°C and 75°C simulations, but the equilibrated 25°C system was used as their starting configuration. The higher temperature systems, one near human body temperature and one higher still, ensured rapid enough internal reorientations to facilitate the study of dynamical phenomena and media reconfiguration over a computationally accessible time span. Nonbonded cutoffs of 9 Å (hexane, octane) or 10 Å (1-octanol) were used.

For water saturated octanol, 76 waters from a pre-equilibrated system were surface-juxtaposed to the 228 molecule pre-equilibrated octanol system. After initial brief relaxation, the temperature was increased to 500 C (under constant volume conditions) until a uniform distribution of the water molecules was achieved. The process is illustrated in Figure 1. The temperature was then incrementally decreased to 25°C, annealing, over a period of 1.0 nanoseconds. Equilibration followed until limiting values were reached (2.0 nanoseconds), allowing configurational reorientation of water and octanols at 25°C prior to any data collection.

For the phenol-to-benzene perturbation, the aromatic solute is solvated in a periodic box of water-saturated octanol solvent (98 octanols, 26 SPC waters, obtained by "hacking" down and reequilibrating the 228:76 box), averaging approximately 30.2 Å per side. A 10.0 Å residue-based nonbonded interaction cutoff was used. The all-atom phenol and benzene parameters were those of Jorgensen[29]. The temperature was maintained at 310 Kelvin to encourage reorientation within the octanol media. For the FEP, λ_1 state refers to the solvated phenol, the λ_0 state to the solvated benzene. The phenol hydroxyl group is mutated into a benzene hydrogen. Each end-state was relaxed for 200 psec before beginning the perturbation calculations.

The aqueous (SPC water model) phenol to benzene perturbation was done as a fifty window FEP calculation, 100 psec (40/60 equilibrate/collect) in each of forward and reverse directions (after 20 psec end-state relaxations), at a constant temperature of 310K, using a 10 Å nonbonded cutoff, 2.0 fsec stepsize, updating the pairlist each 20 MD steps.

Breaking up long chains into multiple (bonded) residues ensures that overlapping interaction spheres (i.e. spheres of 10 Å radii for residue pairing criteria in octanol) extend smoothly through the media and terminate evenly at their peripheries. The pseudo-fragmentation of long chains was also necessary in order to avoid severe molecular collisions between molecules' tailward atoms. Due to the use of the "first atom" pairing criteria, long unfragmented chains have a finite probability of their tailends overlapping before the head (pairing criterion) atoms come within each other's 10 Å range and "see" each other. VDW overlaps result in sudden unrealistically large repulsive forces in systems where the residue size approaches the nonbonded cutoff length. For hexane, the molecule was broken down into two residues, atoms 1-3 and 4-6, where the first atom of each residue was used to determine residue pairing and thus atom pairing. For octane, two residues of four atoms apiece were used. 1-Octanol was partitioned into four residues of atoms 1-3, 4-5, 6-7, 8-10, where atom one is the hydroxyl hydrogen.

Computational Details

An MD stepsize of 2.0 femtoseconds was used for all liquid phase calculations. All molecular bonds were constrained to their equilibrium lengths with holonomic constraints via SHAKE, within a tolerance of 0.0001 Å or finer. Simulations were carried out with periodic boundary conditions, at constant pressure (1 bar) and temperature, with separate temperature coupling for the solute and solvent. Temperature was coupled to a heat bath with a relaxation-time τ_T of 0.2 psec. The pressure coupling τ_P was

also set to 0.2 psec. Nonbonded pair lists were updated every 30 MD steps for simulations involving octanol. Reaction field conditions were always employed to handle long-range interactions. For the calculation of bulk dielectric fluctuations the three-dimensional **M**(t) was computed at every MD step. For structural and dynamic analyses, coordinates were saved every 0.1 or 0.05 psecs. Each system was initially minimized and then thoroughly MD relaxed prior to collection of data.

Fitting To Conformer Distributions

Definitive experimental data regarding the conformational distribution of chain molecules is difficult to come by. Wong, Mantsch, and Snyder (WMS) used rotational isomeric state (RIS) models to assign raman spectroscopy bands to specific conformers of hexane[30]. While the use of RIS analysis (a rudimentary theoretical method for determining the distribution frequencies among a set of available conformational states by using their Boltzman probabilities[31]) is not "experimental" in the wet-lab sense, WMS were able to achieve good results in assigning experimentally determined raman signals to hexane's set of RIS conformational states. Their conclusions concerning the conformational distribution are dependent on the uncertain magnitude of the gauche/trans barrier height energy chosen as an RIS parameter. WMS present data for two potential barrier heights, 0.5 and 0.9 kcal/mol. In Table I their findings for the percentages of each conformer are compared to those generated by our best set of AMBER/MD liquid alkane parameters.

The conformational distribution produced by the MD simulation is in relatively close agreement with the WMS RIS results, most closely resembling those generated using the ΔE=0.9 G/T barrier. There is excellent agreement in the ordering of conformational preferences. The triply-trans conformation (ttt) is somewhat more frequent while singly-gauche (gtt) conformation is slightly less frequent in the MD than in the WMS results. These two conformations (ttt and gtt) are consistently the first and second most highly preferred in all three data sets. The bottom of Table I reports the average percentage of gauche or trans conformations in the outer two hexane torsional rotors versus the inner rotor. The agreement between the WMS(ΔE=0.9), OPLS/MC, and AMBER/MD is quite strong. Given the "experimental" error involved in matching the original raman spectra with the RIS-approximated conformational frequencies, the AMBER/MD hexane model sufficiently reproduces the "experimental" conformational distribution. Certainly this is true in the qualitative sense.

Reproduction Of Physical And Thermodynamic Properties

The final AMBER/MD long-chain alkane/alkanol parameters are given in Table II. The computed physical properties of the MD-modeled pure liquids hexane, octane, and 1-octanol, are reported in Table III. Jorgensen's OPLS/MC results for hexane are restated in Table III for purposes of comparison. Note that densities (in terms of volume per molecule) and molar enthalpies of vaporization from MD simulations are all acceptably within five percent deviation from experimentally determined physical properties (percent deviations are given in parentheses in Table III).

The density of water-saturated octanol can be calculated using the known value for the partial molar volume of water added to octanol. The volume of water at 298K is slightly less at saturation levels in octanol (17.92 cm^3 mol^{-1}) than in pure water (18.57 cm^3 mol^{-1})[32]. This produces a slight "contraction" of volume in the octanol/water media relative to what would be observed if pure octanol and pure water were simply volume additive. For the water-saturated octanol system containing 125 octanols and 42 waters, the experimental volume at 25°C is 34100 cubic Å/system, arrived at by using the experimental volume per molecule for octanol (262.8 cubic Å) and for saturation waters (29.76 cubic Å). Simple volume additivity would result in 34145 cubic Å. The AMBER/MD model closely reproduced the correct binary system density, 34014 cubic Å/system, differing from the experimental value by only 0.25 percent. The standard deviation associated with the MD-computed value is ± 402, due to normal volume fluctuations over the 1.0 nsec MD averaging period. The phenomena responsible for the volume "contraction" is apparently reproduced (and perhaps even slightly overemphasized) in the MD model.

The heat of vaporization of water-saturated octanol was not available for comparison. However, the heat of solution of water vapor in 1-octanol at infinite dilution is given as 9.72 kcal/mol (by subtracting the heat of solution of liquid water in 1-octanol at infinite dilution (0.80 kcal/mol) from the heat of vaporization of liquid water at 25°C (10.52 kcal/mol)[33]. The MD-computed ΔH_{vap} sums to 2049 ± 20 kcal for a 125-molecule pure octanol system (by multiplying the molar ΔH_{vap} in table Nprop for pure octanol by 125), and 2516 ± 36 for the mixed 125/42 octanol/water system. Potential

energies were -1783 kcal/system for octanol/water and -1341 kcal/system for pure octanol, which means added waters contribute -442 kcal, or about -10.5 kcal/mol. The computed number 10.5 is within 8% of the experimental (infinite dilution) number 9.72. However, this was *not* a simulation of the solvation of an isolated water (infinite dilution) in bulk octanol, where each of the water molecule's first-solvent-shell hydrogen-bonding sites may be occupied only by octanol molecules. Apparently, a water is more strongly bound in water-*saturated* octanol than at infinite dilution, stronger in proportion to the number of first-shell sites that are occupied by other waters of saturation rather than by octanol molecules. In fact, pure SPC water has a potential energy value of about 10.5 kcal/mol[25], suggesting that saturation waters are as tightly bound in hydrated 1-octanol as in pure water. Following this rationale, the computed value of 10.5 kcal/mol must be *closer* than 8% to the true heat of solution of an isolated water into water-*saturated* octanol.

Overall, simulations employing these AMBER/MD long-chain alkanol parameters have faithfully reproduced critical thermodynamic and physical liquid properties. This provides an excellent model system for looking into the structure and dynamics of octanol media as well as for use in solvation studies.

INTERNAL STRUCTURE OF PURE & WATER-SATURATED 1-OCTANOL

For many years now researchers have pursued an understanding of the microscopic character of liquid long-chain alcohols, including 1-octanol. Although numerous experimental techniques have been applied, and some insight has been gained, no definitive molecular-level picture of the media has resulted. Using experimental methods alone, it would be difficult to derive detailed information about the liquid-phase internal arrangements and interactions of even these seemingly uncomplex amphiphilic molecules. But, the micro-level interactions that determine the media's macroscopic properties can be described by statistical mechanical treatments. Here we pursue a direct understanding of structural and dynamical details of the media via molecular dynamics computer simulations. Assuming a "good" model can be proven by demonstrating that it correctly reproduces observable thermodynamic and macroscopic physical properties, a reasonable level of confidence in the model's predictive ability should follow. This gives the molecular modeler unique access, in the form of a time-sequential ensemble trajectory, to information regarding molecular-level details which are largely unavailable to experimentalists.

MD simulations establish that the internal structure of liquid 1-octanol media divides naturally into four levels of structural classification. Structural features at the levels of individual *molecules*, molecular *aggregates*, and *regions* of special character will be discussed. The static dielectric constant, a macroscopic property which can be viewed as an average of *system-wide* microscopic interactions, has been computed from

the MD trajectory and will be discussed in a later section along with frequency dependent dielectric behavior.

Features Of Individual Molecules

Features of interest at the individual molecule level include conformational distributions and hydrogen bonding characteristics. The orientational character of hydrogen bonds, their numbers, and types of bonding arrangements with direct neighbors, are examined and compared with available experimental data. Table IV reports the percentage of gauche/trans (G/T) conformations for each of 1-octanol's seven torsion angles, and the distribution of the specific G/T combinations for all possible conformers in the alkyl portion of the liquid phase molecules. The working definitions for gauche and trans conformations are 0 to 120 degrees G^+, 0 to -120 degrees G^-, and 120 to 240 degrees T. Although there is a 62% gauche, 38% trans, distribution for the H-O-C-C torsion, the trans conformation is actually slightly preferred over either of the individual G^+ or G^- conformations. The O-C-C-C torsion (48%G, 52%T) exhibits a net trans preference, but with significant gauche occupancies. The alkyl portion (24%G, 76%T) shows a strong preference for the more extended trans conformations. Using the same datasets as in Table IV, Figure 2 illustrates the average rotation-angle frequencies for 1-octanol's seven torsions. Although the G/T working definition used here has wide tolerances, it is clear that there are well defined occupancies of the traditional 60, -60, and 180 degree angle G/T positions. The solid line representing the H-O-C-C torsion illustrates the very slight preference for the trans hydroxyl conformation. The dashed

line representing the O-C-C-C torsion demonstrates the increasing trans preference. Skewing toward 70 and -70 degree gauche angles occurs along the hydrocarbon chain as the methyl tail is approached.

The "head group" torsions are the most flexible rotors. This trait is built into the model; it is mostly a function of the parameterization rather than of liquid interactions. Recall that there is no VDW parameter on the hydroxyl hydrogen, leaving the two-component fourier torsion potential to control flexibility in the realistically "loose" hydroxyl hydrogen. Although the fourier terms involving the oxygen O-C-C-C torsion are identical with the C-C-C-C terms, oxygen's VDW radius is smaller, resulting in lower 1-4 repulsion in the gauche conformation. The reduced conformational restriction of the hydroxyl atoms enhances the hydroxyl's ability to attain favorable hydrogen bonding orientations in spite of potential hindrance by the bulky alkyl group. This also affords greater opportunities for the formation of hydrogen bonds even when the alkyl tail is caged or entangled, and will influence the relaxation dynamics of any property depending on reorientation of the dipolar H-O-C group.

The lower portion of Table IV reports the average occupation frequencies for all possible conformational states in the eight-carbon alkyl chain portion of the liquid octanol molecules. These alkyl conformations read left to right, from torsions involving the alkoxyl methylene to those involving the tail-end methyl, i.e. gxxxt means the torsion containing the alkoxyl methylene is gauche, and the torsion containing the tail-end methyl is trans. Clearly, the single most preferred conformation is when five torsions simultaneously assume the fully extended all-trans configuration, accounting for

21.5% of all chains. There are about twice as many, 44.1%, singly-gauche chains. The gauche torsion is distributed fairly evenly among the five possible positions, although a gauche torsion involving the larger tail-end methyl group is the least preferred of this group. The doubly-gauche conformations account for 27.2% of chains, with tandem gauche torsions understandably contributing the least (because of van der Waals repulsions incurred from steric contact). Triply-gauche conformations contribute only 6.7% to the total, again with triply tandem conformations occupied least frequently. Only 0.5% of the chains are quadruply-gauche, and in a sample of 25000, no quintuply-gauche chains were found.

Hydrogen-Bonded Aggregates

The typical environment of an octanol hydroxyl group, and its orientation in to its nearest solvent shell neighbors can be determined by referring to its radial pair distribution function (RDF) as in Figure 3. Recall a RDF, denoted g(i,j), simply measures the local density of an atom *j* at a specified position R (a radial distance away from some chosen center *i*) relative to the bulk density of atom *j*. It is a measure of the probability of a specified atom being located within the volume of a spherical thin-shell (centered about R) relative to the random probability of its being located there. A working definition for a hydrogen bond was chosen as any two hydroxyl oxygens closer than 3.5 Å, having one of their hydrogens contained between, where the angle of $O^{\prime H}\cdots O$ is no greater than 30 degrees at the maximum O-O distance. This turns out to mean the bound hydrogen must never lay beyond 2.68 Å of the partner oxygen. As the RDF's

g(O,O) (solid line) and g(O,H) (broken line) indicate, most O-O and O-H hydrogen bond distances are well within these geometric limiting criteria. On average, an octanol oxygen O_c "sees" two other oxygens about 2.75 Å away, one hydrogen at 1.9 Å, and a second hydrogen about 3.25 Å away (see the integration line data between 0.0 and 3.5 Å for the coordination numbers). Octanol oxygen O_c is hydrogen bonded to two other oxygens; one (with its attached hydrogen at 1.9 Å) is a donor to O_c, and the other (with its hydrogen at 3.25) is an acceptor of O_c's hydrogen. By using the law of cosines, the average angle $O^{-H}\cdots O_c$ between the donating neighbor and O_C is 21 degrees, putting the hydrogen out of plane so the bond is not strictly linear. Some "bending" of hydrogen-bonded polymer chains may influence this. As a reference for comparison, the analogous angle between water dimers has been experimentally measured as 6 ± 10 degrees[34]. The average angle $O_c\cdots O^{-H}$ between O_c and the accepting neighbor is 114 degrees. There are two more hydroxyl oxygens within 5.0 Å of O_c as evidenced by the second peak at about 4.5 Å and the coordination plot. Even at this level of resolution, a low third peak appears at 7.0 Å. On average, hydroxyl oxygen O_c is apparently a member in a chain of hydrogen-bonded molecules, or some other complex of hydroxyl groups.

It is well established that some form of polymeric hydrogen-bond-linked aggregate is present in 1-octanol media. Several possible hydrogen bonded polymeric species have been hypothesized by experimentalists, although no definitive description has been given. In fact, various interpretations of conflicting experimental data have provided an abundant source of arguments regarding the nature of these aggregates. There

have been many questions. Are there long polymeric chains, or closed ring structures, or both? Are there various sizes and configurations or are there a limited number of species? Is there a rapid turnover of such structures? What is the influence of temperature on the internal structure and dynamics of the media?

This MD study finds that 1-octanol forms variable sized hydrogen-bond linked aggregates or "clusters". A cluster structure is identified by choosing an initial molecule, designating any molecules hydrogen bonding to it, then continuing to follow along the chain(s) of hydrogen-bond connectivity until there are no more. As described above, the great majority of molecules serve as donor-acceptor links in a hydrogen-bonded network. Figure 4 (inside octanol) is a snap shot from a pure 1-octanol MD trajectory, a typical configuration, emphasizing the orientations of hydroxyl groups. The tendency for molecules to form polymeric networks of hydrogen bonds is evident. Long chains can be seen coursing through the media alongside smaller closed-loop ring clusters.

Table V reports the numbers, types, and percentages of hydrogen-bond "types" formed at two temperatures in pure 1-octanol. The values in the upper section of Table V are the actual *numbers* of items found on average in a periodic system of 250 molecules. In the lower section, the values are recast as the *percentage* of all 250 molecules which appear as a particular given species. There are fewer clusters found per coordinate-set (24.1 versus 32.5) at 40°C than at 75°C, reflecting the greater number of molecules per cluster (10.4 versus 7.4) at the lower versus higher temperature. The division of this data into rings per set and chains per set indicates that an overall prefer-

ence for rings versus chains is reversed as temperature increases, as expected based on the influence of entropy. There is a marked increase in the number of monomers at higher temperature. It is important to establish the number of unpaired hydroxyl hydrogens, referred to as non-donors in Table V, as this will provide a comparison against available near-infrared data. Acceptor-only molecules are those whose oxygen acts as a hydrogen bond acceptor, but whose hydroxyl hydrogen is unbound. Non-donors are the number of acceptor-only plus the number of monomers. Donor-only molecules are those whose hydrogen acts as a hydrogen-bond donor, but whose hydroxyl oxygen is unbound. Non-acceptors are the number of donor-only plus monomers. The number of partially-unbound species increases with temperature as rings are broken and large clusters break down.

The percentage of free (unbound) hydroxyl hydrogens calculated from the MD simulation can be compared with the experimental near-infrared results of Grunwald and Pan (GP)[35]. GP determined the mole fraction of free hydroxyl hydrogens from the sharp band intensity at 1430 nm, which signals the first overtone region of the free OH bond-stretching vibration. A broad band spanning from about 1430 to beyond 1700 nm represents OH-stretching when the hydroxyl hydrogen is hydrogen bonded; GP attribute this to some generic polymeric species. For MD simulations we have designated the species causing these two stretching regions as *1)* the non-donors, and *2)* those which participate in ring clusters or polymeric chains by at least donating. There is very close agreement between experimental[35] and MD results regarding the percentage of free hydroxyl hydrogens present on average (see Table V, lower section). The 40°C MD

simulation predicts 5.2 percent of hydroxyl hydrogens are unbound; the experimentally determined value is 5.4 percent. The 75°C MD simulation predicts 11.1 percent of hydroxyl hydrogens are unbound; the experimentally determined value is 11.2 percent. This strongly indicates the MD model is a realistic representation of 1-octanol's internal structure, and lends credibility to further observations made here regarding the structural specifics of hydrogen-bonded species and cluster-size distributions.

In Figures 5 and 6, the histograms illustrate the frequency distribution of cluster sizes from MD simulations of pure 1-octanol at 40°C and 75°C respectively. The main display histograms report the percent of molecules (y-axis) which participate in a cluster of a given size (x-axis). The inset histograms report the actual number of clusters (y-axis) of a given size (x-axis) that were present on average in a system of 250 molecules. Although there is a significant frequency of very large sized aggregates, note the tendency to prefer the formation of 4-, 5-, and 6-, membered clusters (and 7-membered at 40°C), and the conspicuous unfavorability of 2- and 3-membered complexes. It should not be surprising that pentagonal and hexagonal motifs are present, as these are prevalent in ice water[36] where the hydrogen-bonds are more linear and energies are optimized. The factor of finite system size is probably not significantly influencing the distribution since the preferred cluster size is very small relative to the overall number of molecules.

Let us assume for the moment that complexes form linear open-ended chains, with each molecule's oxygen *accepting* one hydrogen bond and its hydrogen *donating* to form another. Then every chain would have one "uncompleted" hydrogen bond as a

result of the free oxygen acceptor and hydrogen donor on opposite chain ends. The shorter a chain is, the lower is the ratio N_B/N_M of the number of hydrogen bonds formed (N_B) relative to the number of molecules involved (N_M), i.e. 0/1 for monomers, 1/2 for dimers, 2/3 for trimers, 3/4 for a tetramer and so forth. As the size of a chain increases, the N_B/N_M ratio increases, optimizing the average enthalpy gained per molecule involved in an aggregate, causing larger aggregates to be more stable. But why are dimers and trimers so unstable, and why the sudden increased stability in going from a trimer to a tetramer?

When a four-membered hydroxyl chain bonds back onto its own terminus, closing to form a polymeric ring, N_B/N_M becomes 4/4 rather than 3/4. Another way to increase the ratio would be in forming more than two hydrogen bonds per hydroxyl group, branching, opening the door to the possibility of multicyclic rings. Ring structures of any size should be enthalpicly favored, except for 3-membered rings. There may be enough extra strain in forming cyclic trimers to overshadow much of the gain. Their $H\cdots:O^{\prime H}$ angles must be nearly 60° to form linear hydrogen bonds, far from the more optimal 110°-120° (because of the lone pair geometry). In a cyclic trimer, all three partially positively charged hydroxyl hydrogens are very proximally located, causing electrostatic repulsion. Also, there is less freedom in configuring a trimer's bulky alkyl tails; to avoid steric clashes they must fan out, extending into the media. Of course a dimer can not form a true cyclic ring, although bifurcated hydrogen bonds may be possible, subject to strain akin to that in the trimer.

If higher order aggregates are preferable from an enthalpic standpoint, why is there a clear trend of decreasing frequency as N_M grows beyond five or six? Why not form a single massive inter-hydrogen-bonded network? In fact this does occur, but with very low frequency due to the decrease in entropy as aggregate size increases. The entropy of association is net negative and recall $\Delta G = \Delta H - T\Delta S$. With an increase in temperature the entropy factor contributes to an increase in the free energy of the higher order aggregates, decreasing their frequency even more. In comparing Figure 5 with Figure 6, a decrease in the number of larger aggregates at higher temperature is evident, along with an increase in the number of smaller clusters.

Four, five, and six-membered rings are the favored species at both 40°C and 75°C. A planar five-membered ring would have linear hydrogen bonds with optimal 108° for angles $H\cdots:O^{\prime H}$ (thus centering the lone pair electrons between the donor H and acceptor O). Six-membered (planar) rings also have an excellent 120° linear hydrogen-bonding geometry. Four-membered rings formally make 90° angles, but as it is normal for the hydrogen to lay slightly off the O-O vector, this opens up the tetramer ring angles. If the hydrogens of a planar tetramer ring are 21 degrees out of plane, the $H\cdots:O^{\prime H}$ angle is 111 degrees. In the trade-off between enthalpic and entropic factors at 40°C, five-membered ring clusters are the more prevalent species; at 75°C the four-membered tetramer rings are slightly more favored.

It should be noted that "rings" are defined as polymeric species having at least one head-to-tail loop closure but they may also contain extended chain regions. Chains have no loop closures but may have branching. Thus a 6-membered ring may be a

closed loop of six hydrogen-bonded octanols, a closed loop of five octanols and one attached "tail-piece" octanol hydrogen bonded to one of the ring members, or possibly a closed loop of four octanols and a single dimeric tail-piece, or two monomeric tailpiece octanols bound to different positions on the 4-membered ring. Figure 7 shows several examples of isolated ring structures from an MD trajectory. As the number of molecules in a "ring" increases, the probability that the species is a simple closed loop decreases. However, *most* of the smaller ring structures, up to and including 6- or 7-membered rings, are simple closed loops. Figure 8 illustrates aggregate formations in pure liquid 1-octanol media. The oxygen VDW surfaces are in blue, hydroxyl hydrogen surfaces in red, and the alkyl tails are represented as white lines. The three-dimensionality of the illustration is enhanced when the viewer remembers that alkyl tails closer to the viewing eye are "thicker". Although some short to medium-length chains are present, the majority of clusters seen in this particular example are small ring clusters forming a central hydroxyl core from which their alkyl chains radiate generally outward.

In Figures 5 and 6 (main display), the upper solid line drawn through the histogram indicates the contribution from species containing rings, the lower solid line indicates the contribution from open chain species. Both rings and chains contribute significantly to the peak centered around 5-membered species. At 40°C the excess of four, five, and six-membered structures is due mainly to the formation rings. Aggregates (both rings and chains) of a size greater than say 20 molecules become less frequent with an increase in temperature. The equilibrium shifts toward the smaller species, including monomers, dimers, and trimers. As temperature increases the

number of chains in the peak area doubles, but the ring contribution essentially remains balanced; both large rings and large chains are converted to small chains.

These results support those experimentalists who have presented evidence for a distribution of cluster sizes rather than those who argue for a limited set of very few species. Several experimentalists have recently presented data in support of their hypothesis that 1-octanol consists almost exclusively of a monomer/tetramer equilibrium system, with perhaps a very low concentration of some other small species such as dimers[37,38]. Some of their data might now be usefully reinterpretable in light of this new information from MD simulations. Yes, tetramer (ring) structures are present in high concentration, but rather than argue the case for the presence of exclusively two species (monomers and tetramers), their findings might also be interpreted as supporting the presence of two less specific hydroxyl species, namely H-free hydroxyls and H-bound hydroxyls. This would bring them into agreement with these MD studies as well as with other experimentalists who have found evidence for the presence of several species[39,40]. These studies indicate the presence of monomers, small multi-mers, closed rings, and large hydrogen-bonded polymeric aggregates, and also demonstrate temperature dependent changes in species' concentrations which parallel the MD results.

Self-Diffusion Coefficients

The three-dimensional self-diffusion coefficients were computed at 40°C and 75° for pure 1-octanol by monitoring the displacement of a molecule as a function of time using the Einstein relation[41]

$$6\,D\Delta t \;=\; \lim_{\Delta t \to \infty} < |\,\mathbf{r}_i(t+\Delta t) - \mathbf{r}_i(t)\,|^2 > \quad , \tag{3}$$

where D is the center of mass self-diffusion coefficient, $\mathbf{r}_i(t)$ are the molecular positions at time t, and Δt is the time interval between the initial position and the position at time t. An average $<(\Delta \mathbf{r})^2>$ was computed for each of the 250 octanol molecules in a simulation, over a series of progressively longer time intervals, Δt, and all choices of time origins taken from a 600 psec segment of MD trajectory. The results of the 250 separate diffusional trajectories were then averaged together to improve statistical accuracy.

Figure 9 is a plot of the average $<(\Delta \mathbf{r})^2>$ versus Δt, shown for 40°C and 75°C, for Δt's up to 30 psec. The best least-squares-fit straight lines to each diffusion plot are also shown; all points contributed equally in the fitting. The slopes of the plots are equal to $6D$. While the shorter time intervals (represented by points nearer t=0 in Figure 9) have the greatest accuracy, the Einstein relation is most applicable in the limit of an infinitely long time interval. The plots in Figure 9 have reached a linear regime as evidenced by their unbiased random wavering about the fitted lines. The experimentally determined self-diffusion coefficients for pure octanol is 0.325×10^{-5} cm^2s^{-1} at 40°C, and is 0.85×10^{-5} cm^2s^{-1} at 75°C The two pure 1-octanol MD simulations (lowermost and upper-most curves) give $0.68\pm0.0l\times10^{-5}$ cm^2s^{-1} and $1.55\pm0.01\times10^{-5}$ cm^2s^{-1}, respectively, from the slopes of the fitted lines in Figure 9. The modeled 1-octanol is about a factor of two times more mobile than real 1-octanol. Modeled 1-octanol is more rapid for several reasons, most importantly the lack of explicit hydrocarbon hydrogens, thereby reducing intermolecular friction and entanglement. Also, the

torsional potentials, taken along with the distortional flexibility due to freely bending angles, were parameterized to err on the fast side if at all. So long as it remains qualitatively useful, a model which reorients "too rapidly", rather than accurately reproducing octanol's realistically slow dynamics, is probably desirable in most cases for MD studies.

The lower of the two inner curves (Figure 9) is for the diffusion of 1-octanol molecules in the presence of saturating waters, taken at 40°C. The waters apparently "loosen up" the media, causing the octanols to diffuse slightly more rapidly ($0.70\pm0.1\times10^{-5}$ cm^2s^{-1}) than in the pure liquid. The upper inner curve represents the diffusion of the saturating waters, which are quite mobile ($1.10\pm0.1\times10^{-5}$ cm^2s^{-1}) relative to the more sluggish octanols. As a point of reference, the experimental self-diffusion coefficient of pure liquid water[42] at 25°C is 2.30×10^{-5} cm^2s^{-1}, and the calculated diffusion coefficient for model SPC water[25] at 27°C is 3.6×10^{-5} cm^2s^{-1}. The octanol-saturating waters are certainly slowed by the presence of the long-chains and the restricting hydrophobic regions.

Straight line fits to the first 10 psec interval give 0.88±0.01, 2.14±0.02, 1.05±0.01, and $1.42\pm0.02\times10^{-5}$ cm^2s^{-1}, for 40°C and 75°C pure 1-octanol, 40°C water-saturated octanols, and saturation waters, respectively. One can see from the plots that as shorter time intervals are considered, the diffusional mobility (slope) increases. There is very good statistical accuracy in the first 5 picosecond interval; the concave downward curve is a reliable signature. In the short term, diffusion is more rapid, with individual octanol molecules rattling around in a "cage" of surrounding solvents. Over the long

term a molecule may "pass" through the slowly distorting cage, diffusing with no memory of a previous location. But the molecules' short-term memories of their positions with respect to their cage-neighbors confers elasticity to the media.

What is the mechanism of this short-term configurational memory? Hydrogen-bonded molecules may wag their alkyl tails if space permits, thus changing their center of mass over the short term, but remaining in place until the hydrogen bond eventually breaks. Also, a broken hydrogen bond may reform or exchange to a nearby site while the VDW-packing interactions of the hydrocarbon tail anchor the molecule in place. The polymeric hydrogen-bond network gradually reconfigures over the long term, but undergoes elastic distortion over the short term. As the viscous media distorts, molecules change their center of mass positions but remain in essentially the same configuration with respect to their neighbor solvents.

Preferential Solvation In Water-Saturated Octanol

The third level of structure in 1-octanol and water-saturated 1-octanol was designated as *regions of special character*. More specifically, there are regions containing relatively higher or lower concentrations of polar groups. Polar hydrophilic groups prefer to directly associate with or reside in the nearby vicinity of other polar hydrophilic groups; consequently, hydrophobic groups also tend to congregate. Amphiphilic-chain solvents, such as octanol, facilitate optimal allignment and association of "like" groups by forming clusters resembling inverted micelles. A *region* can be further characterized as to its average "size", its degree of definition, its relationship to

adjacent regions, and the effect of added saturating waters.

Firstly, the average number of hydrogen bonds formed per molecule can be determined from the coordination number plots presented in Figure 10 for water-saturated octanol at 40°C. The number of hydrogen bonded partners that a given hydroxyl group forms to each other species in its first solvent shell (0.0 to 3.5 Å between heavy atoms) is tabulated in Table VI. Each octanol forms 2.21, and each water forms 3.80 hydrogen bonds on average. Hydrogen bonding sites are not fully saturated, but nearly so, especially for water, which is very tightly bound. In Table VI, the numbers in parentheses are the number of hydrogen bonds which would be formed if hydroxyl pairing were random among the available hydrogen-bonding sites. For example, since there are 98 × 3 = 294 sites available for hydrogen bonds to octanols, and a total of (98×3) + (26×4) = 398 possible sites, the number of "unbiased" octanol-to-octanol hydrogen-bond pairs would be 2.21 × 294/398 = 1.63. But on average, only 1.48 octanol-to-octanol hydrogen bonds form per octanol molecule; there is a 0.15 bond bias favoring octanol-to-water hydroxyl pairing. The octanol hydroxyl is slightly more stable when bound to a water than when bound to another octanol. For water there is a 0.62 hydrogen-bond bias for binding to other waters.

Marcus examined the incidence of preferential solvation in water-saturated 1-octanol using thermodynamics and quasi-lattice theory[33]. He predicted that the saturating waters would have a local excess mole fraction of neighboring waters in their first solvent shell as compared to the bulk mole fraction. At a saturation level where the bulk mole fraction of water X_{wat} is 0.250, Marcus predicted the local mole fraction of

waters surrounding a given water would be 0.446, assuming water has four binding sites. This MD study (also with X_{wat}= 0.250) finds the average number of bound sites to be 3.80, and a given water binds to 1.61 other waters. The resulting local mole fraction is 1.61/3.80 = 0.424, in close agreement. Marcus also predicted a deficiency in the local mole fraction of waters around octanol molecules, 0.185. The MD simulation gives 0.73/4.0 = 0.183, if we assume a total of four "binding" sites for octanol as Marcus does. In this perspective, waters most prefer to associate with other waters. and octanols preferentially associate with other octanol molecules. This demonstrates the preferential self-solvation behavior of "like" groups, a conclusion which is not really surprising; hydrophilic and hydrophobic groups prefer not to mix.

Inverted Micellar Bodies In Pure 1-Octanol

The tendency to form small inverted micellar-like regions prevails throughout both pure 1-octanol and water-saturated 1-octanol. These aggregates might be called mini-micelles. It is clear from Figure 3 than a given hydroxyl oxygen (located at R = 0) is immediately surrounded by a polar region of high definition, as indicated by peak region extending out to about R = 5.5 Å. Figure 11 is a closeup of the curves first presented in Figure 3. It focuses in on the dip below g=1.0 in the hydroxyl atoms' RDFs (solid line O and dashed H) from about 5.5 Å out to about 14.75 Å. This dip indicates a hydrophobic area bereft of hydroxyl groups. The hydrocarbon tails reside here, as confirmed by g(O,Me), the dotted curve, peaking at R = 10.25 Å. The (solid) curve g(O,O) then raises above g=1.0 at 14.75 Å until it crests at 17.75 Å. This rise

above g=1.0 at so long a distance from the central hydroxyl oxygen (at R = 0) is not a statistical anomaly, it is the appearance of the next hydrophilic region, the core of adjacent mini-micelle(s). Notice that when hydroxyl groups *are* present in the nonpolar dip area, they reside at very specific locations (at about 7.0 Å and 9.0 Å). These intermittent hydroxyl peaks in the hydrocarbon trough of the RDF may be interpreted as sequentially ordered hydroxyls in a hydrogen-bond-linked polymeric chain stretching through or alongside a nearby hydrophobic region. This "average" structure also results from the fact that most hydroxyls are located at the micellar polar/nonpolar boundaries, with neighboring hydroxyls packed to one side, and the hydrophobic milieu off to the other direction. If the hydrophilic cores of the mini-micelles are thought of as essentially spherical, then most "core" hydroxyls lay nearer to the outer radius of a core than to its exact center.

Figure 12 shows three pure 1-octanol RDFs, two involving hydrocarbon-to-hydrocarbon pairs, g(Me,Me) and g(Me,ML), and also g(Me,O) which is equivalent to g(O,Me). The corresponding coordination number plots are also shown. We can declare the first "hydrophobic" solvation shell to have a center-to-center radius of 4.38 Å since united atom methyls or methylenes have individual optimal van der Waals contact radii of 2.19 (see Table II) and are uncharged. On average, there are 1.0 methyl, 0.3 oxygens, and 3.0 methylenes within the 4.38 Å contact distance of a given methyl group. It was not determined which of any of these neighbors belong to a common molecule, although it would seem likely that some do.

The oscillating contours of g(Me,Me) (solid line) demonstrate well-defined structuring of the hydrophobic milieu about the methyl group. There is one neighboring methyl group in the near-shell (R = 4.25), followed by shells which are alternatively rich in methyls (peaks at 9.0, and beyond 16.0 Å), or depleted of methyl groups (troughs at 7.25, 11.5 Å). Turning to the g(Me,O) curve (long dashes), the coordination number reaches 1.0 oxygen at R = 5.35 Å. The region from 8.25 to 16 Å, where g(Me,O) > 1.0 is rich in polar oxygen groups, as evidenced by the peak in g(Me,O), while depleted of both methyl and methylene groups. We thus see a partitioning of the media into polar and nonpolar regions. At about R = 16 Å, the local concentration of methyl groups increases, as the concentration of oxygens decreases again. There must be a significantly enhanced presence or deficit of these groups at this radial distance for the appearance of a signal at other than g=1.0.

The approximate size of a mini-micelle and the average number of molecules composing it can be estimated from the distances between regions indicated by the RDF. Beginning with g(O,O) (Figure 11, solid), an octanol oxygen positioned at R = 0 sees (on average) a polar-enriched region of length 5.5 Å. This should not be interpreted as a spherical region of radius 5.5 Å, as most hydroxyls lay near the periphery of a hydrophilic core. This polar zone results more from oxygens "looking across" the polar core than it does from centralized oxygens which look out over an encapsulating polar sphere. We can conclude that a pure 1-octanol polar core region has a total diameter of at least 5.5 Å, and somewhat more, but less than 11.0 Å (if 5.5 were a radius). If the second enriched polar region (representing adjacent polar cores) in the g(O,O) plot is symmetric about 17.75, it would measure about 6.0 Å. But the system was not of

sufficient size to confirm R = 17.75 as the true crest; an RDF trails off when R reaches ½ the length of the simulation box (which would be 20.35 Å here). In g(Me,O) (Figure 12, long dashes) a complete enriched-polar-region signal is available from R = 8.25 to R = 16.0, at the boundary where polar and nonpolar regions abut. Note the polar region's signal is not symetric about its peak at 10.25 Å.

Let us piece together these several clues regarding the diameter of the polar-hydrophilic core. The "near" polar-enriched region of g(O,O) indicates that the diameter is greater than 5.5 Å, but less than 11.0 Å, the incomplete "far" polar peak indicates 6.0 Å or greater, and the polar-enriched region of g(Me,O) indicates a core diameter of 16.0 – 8.25 = 7.75 Å. All of these indicators are mutually consistent; let us take 7.75 Å as an estimate of the average diameter of the hydrophilic core region formed when 1-octanol aggregates into inverted mini-micelles.

Apparently, the hydrophobic area sandwiched between micellar clusters fills the approximately 9.25 Å span between the boundaries of the hydrophilic regions at R = 5.5 and R = 14.75 Å in g(O,O) (Figure 11). Half of the hydrocarbon groups in the hydrophobic area are contributed by the micellar cluster with a core near R = 0.0, and half by a cluster with a core centered near 17.75 Å. If the overall hydrophobic span is divided into halves and one half is assigned to each opposite side of a given hydrophilic core, then a micellar body has an overall diameter of 7.75 + 9.25 = 17.0 Å on average in pure 1-octanol. Using a radius of 7.75/2 + 9.25/2 = 8.50 makes the volume of a micellar region 2572 cubic Å. Dividing this by the volume per molecule in the simulation (see Table III), 265.2 cubic Å, yields 9.7 molecules per mini-micelle. This compares

well with the compilation directly from the simulation (Table V) indicating the average number of members in a cluster is 10.4 (at 40°C). Not all of the molecules aggregate to form micellar clusters. The presence of some extended hydrogen-bonded (non-micellar) chains may increase the average number of molecules per cluster.

The hydrocarbon chain portion of an octanol molecule can extend about 9.0 Å, and a significant number of the chains are fully extended (Table IV). Since there is only about one extended chain's length between the hydrophilic regions, the chains from adjacent mini-micelles must either overlap or avoid each other. Figure 12 indicates that a terminal methyl group makes a van der Waals contact with 0.30 oxygens (integration line, long dashes). In other words, about one of every three methyl groups reaches far enough to contact one of the hydroxyl oxygens on the periphery of an adjacent mini-micelle's hydrophilic core. Some of the hydrocarbon tails from octanol molecules in neighboring aggregates must be close-packed adjacent to facilitate this. This is visually evident in color Figure 8 where the clustered octanol hydroxyl oxygens are represented as blue van der Waals surfaces and the nearby hydrocarbon tails as white chains. At this point we have a clear and detailed description of the internal structure of pure liquid 1-octanol.

The Effect Of Saturating Waters On Structure

Do saturating waters disrupt or significantly alter 1-octanol's internal structure? In the binary octanol/water media, the strongly attractive water molecules cause all the other hydroxyl groups to gather around them, with a special bias for other water

hydroxyls. Therefore, when water partitions into octanol it should be expected to lodge in and perhaps promote or stabilize hydrophilic cores, causing larger and more distinctive polar regions than in pure 1-octanol.

There are several notable features in the cluster-size histogram for water-saturated 1-octanol presented in Figure 13 Note the flatness of the curve compared to pure 1-octanol, and the presence of many clusters larger than even 80 molecules. Water molecules pack into "tight" spaces in the media (where a bulky octanol molecule could not fit), promoting a much more extensive hydrogen-bonding network and facilitating interconnections or "bridging" of large aggregates. On average there is a scarcity of long linear hydrogen-bonded polymer chains. In pure 1-octanol the plot of chain species trails off with seventy-mers at 40°C (Figure 5), and with fourty-mers at 75°C (Figure 6). The long-chain curve trails off with twenty-mers in hydrated 1-octanol at 40°C (Figure 13). Long extended polymeric hydrogen-bonded chains are rare in the presence of "loop-closing" waters. Notice also that even with the noiser statistics, four and five-membered species are apparently still slightly favored while dimers and trimers are not.

The RDFs presented for hydrated 1-octanol in Figure 10 are very similar in general contour to those seen for pure 1-octanol (Figure 3). We see a more highly structured polar region in the three hydroxyl-oxygen RDFs, g(Ow,Ow), g(Ow,Oh), and g(Oh,Oh). In Figure 10, the first peak of g(Ow,Ow) (solid line) rises to g=26.6 at R = 2.7 Å (lopped off in Figure 10). Even with the relatively few waters in the system (76 versus 228 octanols), the probability of finding the waters close to other waters, in the

hydrophilic core region, is very high. Where the pure 1-octanol RDF (Figure 3) indicates *two* polar-enriched solvent shells (g>1.0) near a given alkoxyl group, the water-saturated curves indicate *three* polar-enriched solvent shells adjacent to a given water or alkoxyl group. The hydrated-octanol curves do not dip below g=1.0 until about 9.0 Å, whereas in pure 1-octanol the dip begins at 5.5 Å. Figure 14 (a closeup of Figure 10) shows the g(Ow,Me) curve (dots) crossing over the g(Ow,Oh) curve (long dashes) at about 9.0 Å, while the g(Oh,Oh) and g(Ow,Ow) curves cross the g=1.0 boundary to either side of this. The g(Ow,Me) line rises above g=1.0 at 9.5 Å, indicating the prominence of hydrophobic groups, just as the g(Ow,Ow) curve finally falls below 1.0. In water-saturated 1-octanol, the hydrophilic core region is filled with waters and overall larger than in pure 1-octanol, as expected. Following the rationale used in analysing pure 1-octanol, we can conclude the core region has a diameter greater than 9.5 Å, but less than 19.0 Å (if 9.5 were a radius). Notice that Figure 14 just barely extends far enough (R = 20.0 Å) to show the appearance of the neighboring hydrophilic core(s).

The diameter of the polar-hydrophilic core can be more closely estimated from the RDFs in Figure 15. A methyl group at R = 0.0 sees an enriched polar area (dashed lines) beginning at R = 9.0 and ending at about R = 19.0 Å, where methyl enrichment begins again. This means the hydrophilic core has a 10.0 Å diameter, consistent with the indications from Figure 14. It is also clear that water molecules reside not only in the central part of a hydrophilic core; they are present throughout the entire core region and packed in among the octanol hydroxyls along the perimeter of the core. We can estimate the span of the hydrophobic region between the polar cores from the g(Ow,Me) curve in Figure 14. Because g(Ow,Me) rises above g=1.0 at 9.5 Å and falls

back down below g=1.0 at about 19.5 Å, the hydrophobic span must be 19.5 – 9.5 = 10.0 Å. The overall mini-micelle diameter is thus estimated to be 20.0 Å. This is 3.0 Å longer than in pure 1-octanol; here the polar core diameter is 2.25 Å longer, and the hydrophobic diameter is 0.75 Å longer. Apparently the hydrocarbon chains overlap to a slightly lesser extent when water is present.

Figure 16 illustrates how water-saturated 1-octanol media divides into preferentially polar-hydrophilic or nonpolar-hydrophobic regions, forming aggregates again reminiscent of inverted micelles. The yellow van der Waals surfaces represent the oxygens of water molecules, the red VDW surfaces are the hydroxyl hydrogens of octanol molecules, the hydrocarbon tails are represented as black chains. These water-containing aggregates have a more well-defined micellar structure than that seen in the pure 1-octanol clusters (Figure 8). There is a notable absence of long hydrogen-bonded chains in this snapshot. Waters pack in among the octanol hydroxyls, creating enhanced polar areas. Other octanol hydroxyls are then drawn to any unbound water sites, forming stable inverted micellar clusters. The long polymeric hydrogen-bonded chains common in pure 1-octanol are thus less favored under hydrated conditions. An interesting alternative hydrophilic core motif can be seen in Figure 17, which illustrates "water-columns", elongated micellar-bodies forming cylindrical regions of hydrophile with radially extending hydrocarbon tails. Here, the water oxygen VDW surfaces are blue and the octanol hydroxyl hydrogens are red. The water-columns appear to be two or more micellar-bodies with joined hydroxyl regions.

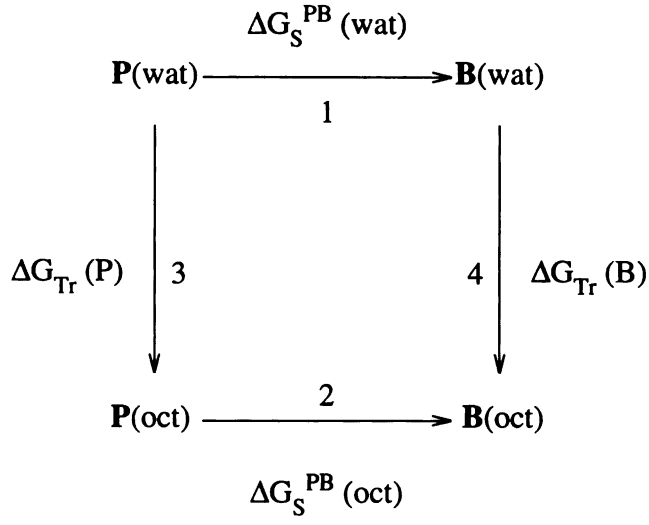
Partitioning Of Small Solutes

1-Octanol is the most frequently used solvent in partitioning studies. It serves as a simple biophase analog media for determining drug or metabolite distribution among the compartments of the body. Sometimes, pharmacokinetic behavior and/or drug activity can also be predicted on the basis of the molecule's interaction with water-saturated octanol. Why is octanol such an exceptionally good analog for the biophases?

An understanding of the internal structural features of water-saturated octanol media helps to clarify the mechanism behind this capability. By virture of its amphiphilicity and the arrangement of its internal structural, water-saturated octanol can mimic the compartmental boundaries of the body, the membranes. It does this by establishing preferentially aqueous or aqueous-like regions, the hydrophilic cores of inverted-micelles. The alkyl tails, regions reminiscent of lipidic tissue, congregate between these hydrophilic cores. The "interfaces" between these regions mimic the membrane boundaries separating the hydrophilic/hydrophobic compartments in the body. A drug can partition between these opposing regions, selectively migrate to a preferred locale, and exhibit a characteristic residence time, much as it would within the compartmentalized body tissues. It would be of interest to know what influence solutes have over the structure of the octanol media surrounding them, whether they can essentially promote an optimally arranged solvent cage. One would expect hydrophilic solutes or their hydrophilic portions to preferentially take up residence in or near the core of an inverted mini-micellar region. Similarly, hydrophobic solutes or hydrophobic solute portions might be expected to migrate preferentially to the hydrocarbon rich

alkyl "tail" regions. An amphiphilic solute (most drugs and metabolites) would be expected to prefer some compromise, perhaps by positioning across regional boundaries.

The difference in the logarithm of the octanol/water partition coefficients $\Delta Log(K_{O/W})$ was determined for the aromatic solutes phenol and benzene. This is the first time the octanol solvation contribution to an octanol/water partition coefficient has been calculated from data generated by high-resolution MD/FEP simulations of a solute bathed in hydrated octanol media. Previously, investigators have reported $\Delta Log(K_{O/W})$ studies for small solutes in carbon-tetrachloride[43] and chloroform[44]. Further studies described here consider the solute/solvent interactions, the orientational relationship of the solutes to the solvent, and the differential positioning of the solute in regions of preferentially hydrophilic or hydrophobic character. The thermodynamic and structural aspects of small molecule partitioning between water and water-saturated octanol were examined by carrying out a molecular dynamics free energy perturbation study of benzene and phenol at 37°C. The following thermodynamic cycle was used:



The difference in the aqueous solvation free energies of of phenol (**P**) versus benzene (**B**) was computed using FEP via path 1. In brief, the Hamiltonian of the system is smoothly perturbed so that the molecular parameters describing phenol are gradually mutated into parameters describing benzene, i.e. the hydroxyl group of the phenol is converted into a benzene ring-hydrogen. (For a derivation of the FEP equation, see the Dissertation Introduction and see chapters I and IV for details on the application of the method). The difference in the free energy of solvation in water-saturated octanol (phenol versus benzene) was computed via path 2. The free energy of transfer of phenol or benzene from water to hydrated octanol (paths 3 and 4) would be too complex to compute directly, requiring the juxtaposition of the two bulk medias, an interface, and extended sampling over multiple solute transfer vectors. Instead, the following relationship was used:

$$\Delta G_S^{PB}(oct) - \Delta G_S^{PB}(wat) = \Delta G_{Tr}(\mathbf{B}) - \Delta G_{Tr}(\mathbf{P}) = \Delta\Delta G_{Tr}^{PB} \quad . \qquad (4)$$

The FEP computational result in pure SPC water (path 1) was ΔG_{aq}(phenol→benzene) = 4.89±0.20 kcal/mol. This compares well with the value obtained by Jorgensen for the similar TIP3P water model, 5.2±0.2 [29]. Subtracting the FEP-computed value in water-saturated octanol, $\Delta G_{oct/wat}$(phenol→benzene) = 3.80±0.35, gave 1.09±0.55 kcal/mol for the difference in solvation free energies of phenol and benzene in hydrated octanol versus pure water. This is equivalent to the difference in their transfer free energies by equation 4 above. Converting this number through the formula relating free energies of solvation to octanol/water partition coefficients,

$$\Delta\Delta G^{PB}_{transfer}(oct/wat) = -2.303\, R\, T\, \Delta Log(K_{o/w}) \quad , \qquad (5)$$

gives a computed value of ΔLog($K_{o/w}$) = 0.76±0.39 for benzene versus phenol. The experimentally determined Log($K_{o/w}$(benzene) is 2.13, and (Log($K_{o/w}$(phenol) is 1.46 [45]. By experimental measurement then, ΔLog($K_{o/w}$) for benzene versus phenol is 0.67 kcal/mol at 25°C.

Although the experimental and computed octanol/water partition coefficients were determined at slightly different temperatures, there is very close agreement between the experimental and calculated values. The experimental partition coefficient is placed well within the standard deviation range resulting from the FEP calculations. Issues of sampling and convergence in these systems are discussed in detail in chapter IV on parallel FEP methods in this dissertation. Extensive sampling was required, so the phenol and benzene solutes had the opportunity to optimally locate and orient within

the heterogenous binary media. Two-thousand coordinate sets were saved from of the λ=1 and λ=0 FEP end-state trajectories (i.e. "pure" phenol and "pure" benzene in hydrated 1-octanol). From these trajectories the solute-to-solvent RDFs were generated. Solute-solvent RDFs are less statistically accurate than solvent-solvent RDFs as only a single solute molecule is present to provide data.

Three solute-to-solvent RDFs are reported in Figures 18, 19, and 20. In all of these plots, the solid line refers to an RDF where the phenol hydroxyl oxygen Op is central (at R = 0), and the dashed line refers to cases where the benzene ring-hydrogen Hb (which replaces the phenol oxygen) is central. In Figure 18, the interaction of these solute groups with octanol's hydrating-water oxygens is represented. Clearly, the phenol hydroxyl is positioned near a polar zone in the media. From the (solid) integration curve we see that 0.75 near-shell hydrogen bonds are made with a hydrating water. The first peak in g(Op,Ow) reaches 18.6 at R = 2.75. It was also determined, by integrating g(Hp,Ow), an RDF (not shown) involving the phenol hydroxyl hydrogen, that 0.36 of this Op-Ow bond has the phenol hydroxyl donating to the water oxygen. So the phenol oxygen donates to water about half and accepts from water for the other half of the 0.75 bond. The (0.75) binding water is hydrogen bonded to a second water (peak at R = 5.0), and there is a third water bound to the second (peak at R = 7.5). Past that, a hydrophilic depleted region is indicated.

Figure 19 (solid line) shows 0.48 hydrogen bond between the phenol Op and Oh octanol hydroxyl oxygen, 0.30 of which is donated by Op (from g(Hp,Oh) not shown). The binding octanol hydroxyl is in the vicinity of other hydroxyls and also the hydrat-

ing waters which were pointed out in g(Op,Ow). Past R = 9.5, a hydrophilic depleted region is indicated. From what was learned about the internal structure of water-saturated 1-octanol, namely that inverted micellar aggregates with hydrophilic cores of 10.0 Å diameter, it can be concluded that on average phenol's hydroxyl group lodges at the periphery of such a polar core. Figure 20 (solid line) shows 0.38 methyl groups within VDW contact distance (3.91 Å) of Op. One would expect this is from a nearby alkyl tail which packs in around the benzene ring.

Figure 21 illustrates the phenol molecule (red) immersed in hydrated octanol, where hydroxyl hydrogens are represented as VDW surfaces (water-H = black and octanol H = cyan) and the positions of the alkyl chains are indicated by gray chains. Notice the phenol hydroxyl binding at the periphery of a polar region, in this case donating to an octanol oxygen (black surface), and the hydrophobic benzyl portion buried in a large nonpolar zone. Also of interest, note that several hydrophilic core regions appear to be connected by short polymeric octanol chains (3-D overlap somewhat obscures the separation of the inverted mini-micelles which are present).

Turning our attention to the benzene ring-hydrogen Hb in Figures 18-20, no polar-enriched zone is closer than 9 Å (broken RDF lines in Figures 18, 19), although a single octanol hydroxyl resides about 6 Å away (broken line coordination plot in Figure 19). There is a region enriched in methyl groups beginning at the Hb VDW contact distance and extending to about 9 Å (Figure 20 broken). This is very interesting in light of our having demonstrated the hydrophobic region in hydrated 1-octanol spans 10 Å on average. If benzene were positioned along the "edge" of a hydrophobic zone, it would

also be near the abutting polar region. But on average Hb does not encounter a hydroxyl oxygen within a radius of 6 Å, and since all six of benzene's Hb atoms are equivalent, none of them are any nearer to a polar region. Benzene must be centralized among octanol's hydrocarbon tails in a wider than average hydrophobic region, one having a diameter at least as wide as 12 Å plus the size of benzene itself.

Figure 22 shows the benzene molecule (red, in center) encircled by many alkyl chains, but with polar groups far removed. Dashed orange lines are shown connecting several benzene hydrogens Hb to the nearest oxygens in polar clusters; the distances are listed in the upper right-hand-corner of Figure 22. Many methyl groups can be seen projecting toward the benzene molecule. It is somewhat surprising that even a solute as small as a benzene molecule may perturb the average solvent configuration by virtue of its hydrophobicity. The hydrated-octanol media accomodates a lipid-loving solute by configuring an optimally lipidic solvation shell around it. Hydrated octanol could undoubtedly form various kinds of solvent shells depending on the nature of the molecule being solvated. It should be interesting to follow this initial study with simulations involving more complex molecules to examine issues of solute conformational preferences, ion binding and other forms of complexation, correlation to pharmacokinetic variables, and the behavior of drugs in a cogener series.

DIELECTRIC PHENOMENA IN PURE 1-OCTANOL

Background Material

The macroscopic static dielectric constant ε_0, or zero frequency relative permittivity, is a measure of a substance's ability to act as an electrical insulator. In the absence of an external field, the molecules in a liquid undergo chaotic motion; there is no net polarization in the media. The bulk dipole moment $< \mathbf{M} >_{time} = < \Sigma_i \boldsymbol{\mu}_i >_{time}$, taken over all molecular dipoles, equals zero. Imposition of a static electrical field $\mathbf{E_0}$ promotes the directional allignment of all polar and polarizable groups in the media, giving rise to collective reorientation. This produces a net polarization, $< \mathbf{M} >_{time} \neq 0$, an induced "bulk" moment, counter to the applied field. This dielectric response occurs as a time-dependent dynamic process during which molecular rearrangements occur until an equilibrium "induced field" condition is reached. If the external field is then removed, the net polarized equilibrium state decays back to the unpolarized equilibrium state where $< \mathbf{M} >_{time} = 0$. Using microwave spectroscopy, this depolarization process is measured as the exponential decay with time of the polarization in a dielectric when the externally applied field is removed, i.e. the rate at which the polarization $\mathbf{M}(0)$ is reduced to 1/e of $\mathbf{M}(0)$. Each substance exhibits its own characteristic rate of dielectric relaxation, depending on the kinetics of the molecular events occurring in the media. This is usually expressed as a relaxation time $\tau_r = 1/k_{rate}$. Even in the absence of external fields, a local region of the dielectric media will temporarily attain a polarized condition $\mathbf{M}_{local} \neq 0$ as it responds to fleeting electrical field perturbations created by the

chaotic motions in the surrounding media. Because spontaneous (bulk or local) fluctuations in **M** can be in the positive or negative direction, they average to zero over time, i.e. $< \delta \mathbf{M}(t) >_{time} = 0$.

The three sources of polarizability are electronic, distortional, and orientational. An electronic or atomic polarization response occurs even in the presence of a very high frequency oscillating field $\mathbf{E}(\omega) = \mathbf{E}_0 \cos(\omega t)$, giving rise to the optical dielectric constant ε_∞. In this study, no explicit electronic polarizability is simulated; the contribution of electronically polarizable groups to the bulk dielectric value ε_0 is ignored. While ignoring the effect of atomic polarizability probably causes the calculated dielectric constant to be slightly underestimated, the effect is small. At very high frequencies the relative permittivity of octanol is about $\varepsilon_\infty = 2.15$ [46]. Another minimal contribution results from distortional polarizability, which arises from the stretching and bending of bonds connecting partially charged atoms. This contribution is partially accounted for as angle bending is included in the molecular model. The orientational polarization resulting from the rotational allignment of polar molecules and rotatable polar groups is the most substantial contributor to ε_0. The strongest polar group in octanol is the H-O-C alkoxyl "head group". The gas phase dipole for this group is about 1.7 Debye; water has a dipole of about 1.85 D in the gas phase. The liquid phase "apparent" dipole of octanol is 2.88 D (20°C) and a range from 2.35 to 2.82 D is generally accepted for liquid water[48-50]. The H-C bonds of the hydrocarbon portion of 1-octanol contribute weak permanent bond dipoles to **M** in reality, but their contribution is considered as "averaged to zero" in this united-atom hydrocarbon model.

The frequency dependent dielectric value $\varepsilon^*(\omega)$ registers how rapidly the media can reorient its set of dipole vectors (its bulk moment) with respect to an externally imposed oscillating electrical field $\mathbf{E}(\omega)$. It is a complex quantity usually given as $\varepsilon^*(\omega) = \varepsilon'(\omega) + \varepsilon''(\omega)$, where ε' and ε'' represent the in-phase (dispersion) and out-of-phase (absorbance or loss) components of the dielectric response in the presence of an oscillating field. Figure 23 shows plots of ε' and ε'' versus $\log(\omega)$ for 1-octanol at 20°C, indicating typical dispersion behavior with increasing frequency[46]. The plot "steps" down at specific frequencies when the external field oscillations become more rapid than the molecular motions responsible for generating the field induced polarization. Notice that 1-octanol has three dispersion regions, and thus three internal dielectric response mechanisms; one occurring at 1 to 22×10^{10} s^{-1} is the major contributor, and two at 1.7 to 5×10^{11} and 1.4 to 4×10^{12} s^{-1} are lesser contributors.

Let us consider the slowest response mechanism. As the field oscillations increase to a frequency beyond the molecular response capability, the orientationally rearranging dipolar molecules can no longer "keep up" with the field, but lag behind it, their motions becoming progressively more out of phase with the field. Energy is thermally dissipated or "lost" as these molecules torque against the field, giving rise to the ε'' absorbance signal. Upon further field frequency increases, the "slow" molecules can no longer sense the shifting polarization, but see a time averaged electrical field of $\mathbf{E}(\omega) = 0$. Thus the absorbance trails off for this particular frequency-dependent response mechanism. Only the more rapid molecular or electronic motions are now capable of responding to the higher frequency field oscillations.

Each dielectric response mechanism corresponds to a dynamic molecular motion, and exhibits a characteristic decay or relaxation time, τ_r. Fortunately, it is possible to calculate these τ_r, via MD simulations, without simulating an explicitly applied electrical field. The τ_r can be determined by examining the time-correlation of the spontaneous fluctuations, $\delta\mathbf{M}$, at equilibrium in the absence of an external field. The fluctuation-dissipation theorem states, "In a system close to equilibrium, we can not distinguish between spontaneous fluctuations and deviations from equilibrium that are externally prepared"[51]. This means that the relaxation of a macroscopic non-equilibrium disturbance (such as the displacement from equilibrium by application of a field) is governed by the same laws and exhibits the same decay characteristics as the regression of spontaneous fluctuations in an equilibrium system. This is the fundamental statistical mechanical connection between the microscopic events accessible via MD simulations and non-equilibrium macroscopic physical processes which are measurable in the lab. Thus, the rate of decay of a macroscopic non-equilibrium process (such as the depolarization that occurs on removal of an electrical field) may be calculated from the correlation of spontaneous fluctuations of an associated microscropic observable ($\delta\mathbf{M}(t)$ in this case) in an equilibrium system.

Computing The Static Dielectric Using An MD Ensemble

In 1936 Lars Onsager formulated a theory of orientational polarization which related ε_0 to molecular dipole moments and atomic polarizabilities[50]. Onsager's model considered the polarization induced in a macroscopic dielectric sphere filled with

polarizable molecules (with its attendant bulk dipole) surrounded by an external continuum of (its own) uniform dielectric. Kirkwood extended the theory to include a consideration of the explicit spatial configuration of charges in the spherical region of dielectric. An orientational "coupling" term was incorporated to account for the influence of interacting molecular dipoles[52]. Frohlich then expressed the theory in terms of statistical mechanical relations involving dipole-moment fluctuations due to the changing position vectors of the charges in time[53]. By statistical reasoning, Frohlich obtained this generalization of Kirkwood's extension of Onsager's equation:

$$\frac{4\pi}{3}\frac{\langle \mathbf{M}^2 \rangle}{3Vk_BT} = \frac{\varepsilon_0 - \varepsilon_\infty}{\varepsilon_0 + \varepsilon_\infty}\left[1 - \frac{\varepsilon_0 - \varepsilon_\infty}{\varepsilon_0 + \varepsilon_\infty}\frac{\varepsilon_{RF} - 1}{\varepsilon_{RF} + 1}\right]^{-1} \tag{6}$$

where V represents the volume of dielectric, k_B is the Boltzman constant, T is temperature, and $< \mathbf{M}^2 >$ is equal to $< \Sigma_{ij} \mu_i \cdot \mu_j >$, accounting for the interaction of molecular dipoles. Note that the fluctuation term $< \mathbf{M}^2 > - < \mathbf{M} >^2$ reduces to $< \mathbf{M}^2 >$ since $< \mathbf{M} >^2$ equals zero when observed over time in a field-absent equilibrium state. The reaction field fluctuation term for handling long-range electrostatic interactions has been included here in square brackets; ε_{RF} is the dielectric of the continuum surrounding the central sphere of explicitly modeled molecular media. The Kirkwood-Frohlich relations, coupled with an explicit consideration of surrounding media effects (via the reaction field model in this case) can be implemented within a molecular dynamics regime[54,55]. Several different boundary conditions may be used. When the sphere is immersed in a continuum of its own dielectric, substituting ε_0 for ε_{RF} in equation 6

gives

$$\frac{4\pi}{3}\frac{\langle \mathbf{M}^2 \rangle}{3Vk_BT} = \frac{(2\varepsilon_0 + \varepsilon_\infty)(\varepsilon_0 - \varepsilon_\infty)}{9\varepsilon_0} \quad . \tag{7}$$

Or, when the sphere is immersed in a continuum with $\varepsilon_{RF} = \infty$, an infinitely conducting dielectric, substituting ∞ for ε_{RF} in equation 6 gives

$$\frac{4\pi}{3}\frac{\langle \mathbf{M}^2 \rangle}{3Vk_BT} = \frac{\varepsilon_0 - \varepsilon_\infty}{3\varepsilon_\infty} \tag{8}$$

which, when rearranged to solve for ε_0, gives

$$\varepsilon_0 = \varepsilon_\infty + \frac{4\pi\varepsilon_\infty \langle \mathbf{M}^2 \rangle}{3Vk_BT} \quad . \tag{9}$$

A value of $\varepsilon_\infty = 1.0$ (for an electronically nonpolarizable media) is used when distortional and electronic polarization effects are not explicitly treated in the simulation. Frohlich (1958) pointed out that "[electronic or] distortional polarization may be dealt with by assuming that the dipoles contributing to **M** are embedded in a dielectric continuum of relative permittivity ε_∞; the only fluctuations are then those arising from changes in orientational coordinates."[53,56]. The same approach is taken in the derivations presented by Neumann[54,55]. However, Steinhauser reminds that "the inclusion of the actual optical dielectric constant ε_∞ [on the right-hand-side of equations 6-9] becomes important if ε_0 and ε_∞ are of comparable magnitude."[57]. For a dielectric such as water with $\varepsilon_0 \approx 80.0$, substituting ε_∞ with 1 changes the left hand side of equation 7

by only a few percent, but for 1-octanol, with ε_0 nearer 8.0, the change is 33 percent; the change is 17 percent in equation 8. For MD simulations where electronic polarization is not treated, the permanent partial charges are enhanced to mimic the effect in an "average" way, resulting in what are commonly referred to as effective potentials, and thus effective dipole moments[58]. It is difficult to ascertain the degree to which these approximations involving the treatment of multibody effects influence the final calculated results for a principally nonpolar media such as octanol. Because simple additivity of the effective molecular dipoles has been assumed for these studies, the appropriate formula to use with toroidal (periodic) boundary conditions, with $\varepsilon_{RF} = \infty$, and with $\varepsilon_\infty = 1.0$ was

$$\varepsilon_0 = 1 + \frac{4\pi <\mathbf{M}^2>}{3Vk_BT} \quad . \tag{10}$$

The Reaction Field Method was employed to treat long-range electrostatics. The field on a dipole in the simulation consists of two parts, a short-range contribution for the explicit molecules situated within the interaction cutoff sphere and a long-range contribution arising from implicit molecules surrounding the interaction sphere. These implicit molecules are considered to form a dielectric continuum with dielectric constant (ε_{RF})[20,50]. When the continuum is polarized by the molecules inside the cavity, this in turn produces a reaction field within the cavity. The size of the reaction field acting on molecule *i* is proportional to the moment of the sphere surrounding *i*. The interaction between these two regions approximates the dielectric friction which would exist between local regions in a macroscopic media. Neumann showed that it is not

possible to accurately compute a dielectric constant via MD if this long-range effect is ignored[59]. The following term is added to the molecular mechanical potential energy equation to correct for the additional torques arising in the presence of a reaction field:

$$\upsilon_{RF} = \frac{2(\varepsilon_{RF}-1)}{2\varepsilon_{RF}+1}\frac{1}{r_c{}^3}\sum_{a,b}^{sphere}\mu_a\cdot\mu_b = \frac{(\varepsilon_{RF}-1)}{2\varepsilon_{RF}+1}\sum_{i,j}^{sphere} q_i q_j \frac{r_{ij}{}^2}{r_c{}^3} \qquad (11)$$

where ε_{RF} is the frequency-independent dielectric constant of the background continuum, r_c is the spherical cutoff radius, and μ_a and μ_b are interacting molecular dipoles. In the site-site form of the equation, q is the atomic partial charge, and r_{ij} is the distance between atoms i and j.

Using the above formulations it is straightforward to calculate octanol's static dielectric constant, ε_0, from the mean-square of the bulk moment by sampling thoroughly over all possible configurations. Molecular dynamics followed over a sufficient time trajectory is used to generate a representative sample of accessible states.

Computing The Frequency-Dependent Dielectric Relaxation

There is also a straightforward method to calculate 1-octanol's frequency dependent relaxation time(s) τ_r. Because of the fluctuation-dissipation theorem, the dielectric relaxation time(s) τ_r can be determined by fitting the decay of the time autocorrelation $C(\delta\mathbf{M}(t))$ of the spontaneous fluctuations of the bulk moment $\mathbf{M}(t)$ to an exponential function (or functions for multiple τ_r).

$$C(\delta\mathbf{M}(t)) = \frac{<\delta\mathbf{M}(0) \cdot \delta\mathbf{M}(t)>}{<\delta\mathbf{M}^2>} = \sum_{r=1}^{N_r} A_r \, e^{-\frac{t}{\tau_r}} \quad . \tag{12}$$

Liquids which are Debye-like in their dielectric relaxation behavior follow an exponential decay in their depolarization; 1-octanol has been shown to be Debye-like[46]. One value of **M** was saved at every MD time step as the vector components M_X, M_Y, and M_Z.

Computational Methods and Results

1-Octanol is relatively viscous and therefore a slowly rearranging media. In gathering sufficient statistics to compute relaxation times, the length of the MD calculation should be 100 to 1000 times longer than the slowest decay. 1-Octanol's slowest relaxation is on the order of hundreds of picoseconds to over a nanosecond, depending on temperature. An MD simulation of 2500 atoms over tens to hundreds of nanoseconds is unfeasible, even with today's supercomputers. Insufficient sampling of an ensemble of accessible states could result in incomplete convergence of properties such as ε_0 and τ_r.

Although these simulations covered several nanoseconds, it is unlikely that the computed *system-wide* description of the media is as accurate as the description of the internal structure at the three lower levels. One factor to consider in this regard is the imposition of finite (non-macroscopic) system size. The correlation length across the 250 molecule octanol box is about 6.3 molecule's length. This means that due to the periodic boundary conditions, molecules may interact "across the box" with molecules

which are actually only several neighbor distances away "in the box". This approximation may have a restrictive effect on dipolar allignment freedoms, and specifically on the achievable contours and orientations of aggregated molecules in relation to neighboring aggregates. For instance, there is a finite probability that at some point in time twenty molecules would form a linear-hydrogen bonded chain, with all dipoles simultaneously alligned in the same general direction, making a very large contribution to the bulk dipole moment of the media. In a cubic box of 250 molecules, a linearly extended twenty-mer will never occur, its contribution will be missed, and the resultant calculated bulk dipole moment may be that much lower.

ε_0 was calculated for modeled pure 1-octanol at 40°C and 75°C. The reaction field method described above is applicable only for homogeneous isotropic systems (of uniform composition) and therefore would not be appropriate for computing dielectric quantities in a small finite sample of binary media like water-saturated octanol. Figure 24 is a plot of the convergence of ε_0 versus time. Apparently convergence is being approached near ε_0 = 5.1 at 40° and 4.2 at 75°C, although longer simulations would be required for confirmation. The computed static dielectric constants agree with the experimentally observed values, 8.7 (40°) and 5.7 (75°), within a few permittivity "units" at both temperatures. Only one MD calculation of water has come within as close as 3 permittivity "units" of 80.0, the experimental value of ε_0(300K)[60]. MD calculations of water's static dielectric constant have given ε_0(292K) = 35 for MCY water[49], ε_0(293K) = 53 for TIP4P water[61], ε_0(298K) = 72 for SPC water[62], ε_0(298K) = 70.7 for SPC/E water[63], ε_0(300K) = 82.5 for flexible SPC water[48], and ε_0(293K) ranged

from 21 to 97 for TIP4P water depending on the smoothing function used at the spherical interaction cutoff boundary[64]. Those who have modeled dielectric properties of pure water regard an ε_0 value anywhere in the range of 50 to 80 as acceptably representative of the highly polar media.

The computed ε_0 values are a bit lower than the experimentally observed ε_0 due to insufficient sampling, the limitations of microscopic system size, or to approximations in the octanol model or its parameterization. Also, the effect of neglecting electronic and distortional polarization may figure significantly in octanol. Recall that only the reorientational contribution to ε_0 has been computed. One might make a strictly nonrigorous case for simply adding in the experimentally known contribution of the optical component, an extra $2.15 - 1.0 \approx 1.1$ to each calculated ε_0, giving computed values of 6.2 and 5.3 for 40° and 75°C. This would bring the results into closer agreement with experiment. Either way, the simulations produce a qualitatively correct low dielectric media, yet one which is not totally nonpolar. Also, the correct temperature dependence is produced, a trend toward lower ε_0 as temperature increases.

Frequency dependent dielectric behavior provides information about the reorientational dynamics of a liquid media. Each of the three characteristic relaxations of long chain n-alkanols is associated with some dynamic motion or rearrangment event in the media. Although experimentalists have reached a general concensus regarding the molecular motions responsible for the faster relaxations τ_2 and τ_3 in n-alkanols, the nature of the molecular level event responsible for the slowest relaxation time has stood long in question. Can our new found knowledge of octanol's internal structural details

help to suggest a plausible mechanism for this slow relaxation?

The dielectric relaxation behavior exhibited by the modeled octanol was characterized by computing the time-autocorrelations of the bulk dipole moment fluctuations $C(\delta\mathbf{M}(t))$ using the algorithm described by Allen and Tildesley[65]. Correlation lags (time-multiples of 0.04 psec) of lengths up to 400 psec were taken along the 2.0 psec (40°) or 3.3 psec (75°) trajectory stretches. Lags were followed from each MD step. All values for lags of the same length were then averaged for each temperature. Also, for short-time response, lags (time-multiples of 0.006 psec), of lengths up to 50.0 psec were computed. Figure 25 shows the decay over 3.3 psec of the 75° trajectory. Clearly correlations decay within about 100 psec, after which point the curve oscillates about the zero line. Figure 26 display plots of $C(\delta\mathbf{M}(t))$ at 40° (upper solid line) and 75°C (lower solid line), shown fitted with single exponentials (broken lines) over 400 psec by using the Levenberg-Marquardt non-linear least squares fitting method[66]. Values for τ_1 and the pre-exponential coefficient A_1 (see equation 12 above) are reported in Table VII. The error in $C(\delta\mathbf{M}(t))$ progressively increases in the direction of increasing time as data becomes more sparce. Because there are many more short-time lags to average over, the statistical accuracy over short times is much greater. The shorter time τ_2 and τ_3 relaxations (and an alternative value for τ_1) were obtained from triple-exponential fits to 30 psec segments within the 50 psec maximum lag period analyses. The short-sections (solid lines) and their fits (broken lines) are plotted in Figure 27. Table VII lists the τ_r and A_r values obtained.

The single exponential fit is most influenced by the slowest relaxation mechanism. The value of τ_1 obtained from this fitting can be regarded as qualitatively representing the decay of long-term fluctuations in **M** and thus the long-term reorientation dynamics of the media. The root mean square of deviations (RMSD) between the 400 psec decay plots and their fitted curves (in Figure 26) are 12.8 (40°C) and 13.3 (75°C). In the 400 psec autocorrelation plots (Figure 26) the faster relaxations are masked due to statistical uncertainty. The degree of error in the autocorrelations can be estimated by a visual appraisal of the waviness seen in the 400 psec curves, and the magnitude of the oscillations about the zero line in Figure 26. This is mainly attributable to the sparsity of statistics. It is also possible that the limited system size and a relatively short correlation length across the periodic box could cause reduced periodic boundary-frictions or disturbance in the border region, attenuating the configurational memory. An accumulation of such effects over a long time span could influence the computed value of τ_1.

The result of the triple-exponential fitting of the short-term decay is encouraging in that the relative relationships between τ_1, τ_2, and τ_3, are reproduced by the simulation. In Figure 27, the fit curves (broken lines) essentially superimpose onto the decay plots (solid lines). The RMSD between the decay plots and fitted curves are 0.002 (40°C) and 0.012 (75°C). The three relaxation times each have about an order of magnitude between them (see Table VII). It is unnecessary to reproduce correct *absolute* time-dependent behavior that is identical with experiment. The simulation is required to exhibit the proper *qualitative trends* in dynamic behavior. The purpose here is to examine the molecular details associated with the dynamic events responsible for the

dielectric relaxation, primarily τ_1. In the absolute sense, the simulation dynamics affecting dielectric relaxation are 3 to 6 times more rapid (40°C), or 2 to 3 times more rapid (75°C, more converged), than what is measured by experiment. This is partially a result of torsional parameterization and partially due to the united-atom approximation. Because no explicit hydrogens are attached to the carbons, intermolecular friction/entanglement is expected to be significantly reduced.

The relative contributions of each molecular relaxation mechanism to the overall dielectric response are determined from experimental data[46,67]; they are presented in Table VII as %CONTRIB. These values are comparable with the values of the coefficients A_r in equation 12 above and are reported under COEFF in Table VII. The percent contributions of each relaxation mechanism are very well reproduced in the pre-exponential coefficients A_r. The model media also produces the correct trend with respect to temperature; the faster relaxation mechanisms make a greater relative contribution to the decay as the barrier to τ_1 is overcome at increased temperatures.

The τ_3 Relaxation Mechanism

Investigators agree that the picosecond time scale relaxation τ_3 is due to the rotational motion of free hydroxyl groups. The low $\Delta H^{\ddagger}$ barrier height associated with the process, about 1.0 kcal/mol, is identical with the barrier height hindering hydroxyl rotation (from microwave spectroscopy)[68]. The dipolar H-O-C-C group's rotation about the O-C bond at the rate of about 10^{12} sec^{-1} is about an order of magnitude slower than

a slow molecular vibration. This is apparently the only molecular motion in the system which could account for the τ_3 relaxation. Because only 5 to 6 percent (on average) of the hydroxyls are *free* the overall contribution should be minimal. Notice the agreement between experiment and simulation in Table VII regarding the percentage contribution made by τ_3 to the overall dielectric relaxation; headings %CONTRIB and COEFF are near 5%. The computed (fit) τ_3 is significantly faster than experimental observations. When combined with freely bending three-atom angles, the H-O-C-C torsional terms may give rotational barriers which are too low, permitting overly facile conformational adjustments. Yet, at about ½ picosecond, the computed τ_3 is of the correct order of magnitude.

The τ_2 Relaxation Mechanism

The second most rapid dielectric relaxation mechanism, τ_2 occurs in ten to several ten's of picoseconds, on a time scale corresponding to the rotation of an entire molecule or perhaps the rotation of a small complex of 2 or 3 molecules[46,67,69-71]. Consistent with this explanation, single molecules (or small complexes) rotate, change their conformational shape, and generally contort themselves to obtain new orientations within their temporary surrounding media cages in the slowly deforming viscous media. Experimentalists believe the molecular event responsible for this relaxation does not require the breakage of hydrogen bonds, since it occurs on a more rapid time scale than that required on average for hydrogen bond turnover (which will be discussed in detail

later). Considering the hydrogen-bonded polymeric clusters in 1-octanol, it is possible that τ_2 could also involve twisting and rotating of the end-of-chain species. For this group of molecules bound by only a single hydrogen bond, dipoles could change direction and magnitude without hydrogen bond breakage, thus making a low-percentage contribution to $\delta\mathbf{M}$.

If the event responsible for τ_2 occurs within the time scale of monomer turnover, then it may well be due to monomer rotation. However, if the τ_2 relaxation happens over a time scale that is longer than that required for the turnover of unpaired monomers, yet shorter than the average hydrogen-bond lifetime, then it would not be completely accounted for simply on the basis of monomer rotations. In that case, τ_2 could be due to the reorientation of rapidly reorienting small complexes such as dimers or trimers or to the reorientation (without breakage) of singly-hydrogen-bonded end-group species. This is essentially the concensus rationale presented in the literature up to this point in time.

The decay time for the survival of completely unpaired species, namely monomers. was computed from MD trajectory data. At some initial time t_0 there is a set of free monomers $\mathbf{F}_i$ which can be identified. As time progresses, each member of this set of monomers will gradually become hydrogen bonded, forming some complexed species. In one relaxation time for monomer decay, about half the monomers do not survive; they become bound. The time autocorrelation of unpaired monomers follows the same formalism as is explained (later) in equation 13 for hydrogen bond lifetimes; just replace $\mathbf{B}_{ij}(t)$ in equation 13 with $\mathbf{F}_i$, the identified free monomers, and τ_B with τ_F,

the free monomer turnover time.

The decay curves (not shown) for pure 1-octanol monomer turnover were fitted to single exponentials giving τ_F(40°C) = 4.1 psec and τ_F(75°C) = 4.7 psec, with RMSDs of 0.20 and 0.31 respectively. But the computed dielectric relaxation times are τ_2(40°C) = 6.6 psec and τ_2(75°C) = 9.6 psec (as reported in Table VII). Monomers are "annhilated" extremely rapidly, and *inside* the time required for the τ_2 relaxation mechanism to occur. Of course, new (reorientable) monomers must be created at exactly the same rate. But, we shall see in the next section that hydrogen-bond turnover takes an order of magnitude longer than this on average. Apparently there is a small population of itinerant molecules (5-10%) which exchange hydrogen bonds at a much more rapid rate than the average. These would be the peripheral molecules, those that are breaking away from or binding onto the ends of chains, or those that wander at the outskirts of an otherwise stable structure such as a micellar aggregate. Prior to these MD simulations, this phenomena has not been described. It may be that the rapid exchange and concomitant reorientation of these itinerant molecules plays a major role in the τ_2 relaxation mechanism.

The τ_1 Relaxation Mechanism

Experimentalists have proposed various hypotheses to explain the τ_1 relaxation of several hundred's of picoseconds. Several explanations have been forwarded, but τ_1 has so far eluded description at the molecular level. Experimentals generally offer a

mechanism involving intermolecular cooperativity, some relatively obscure collective motion. Investigators disagree as to whether τ_1 is governed by hydrogen-bond breakage, although most believe this can not be the rate limiting step[36,67]; the free energy barrier for the τ_1 mechanism is a bit higher than that for hydrogen bond breakage. However, examination of a set of octanol isomers with the hydroxyl placed at various locations along the chain clearly shows that τ_1 is dependent on the placement of the hydrogen bonding group[36,47,72]. In octanol isomers where the hydroxyl group is stericly hindered by neighboring hydrocarbon groups, no τ_1 relaxation was observed. Further, a study of a series of n-alkanols of increasingly longer alkyl chains shows that τ_1 is also at least partially a function of chain length; τ_1 increases with chain size[72]. τ_1 must be due to a complex orientational rearrangement of dipolar groups which depends on the entire molecule, both the hydroxyl and the chain, and also likely involves interactions between multiple molecules. Knowledge of 1-octanol's several tiers of internal structure suggests that to understand the media's internal dynamics one might look to the next highest category of structural complexity beyond individual molecules and their nearest adjacent hydrogen-bonding neighbors. One might consider the reorientational dynamics at the level of molecular aggregates.

Let us propose that τ_1 is the result of the overall restructuring of the media through the simultaneous breakdown and reformation of the various aggregate species. In a static electric field, the final structure would be driven to optimize the allignment of the individual polarizable elements in the field, at once maximizing the bulk dipole moment and the free energy of the media. What happens when the field is then removed?

Thermally driven motion causes a complex reorientational event; hydrogen-bonded chains and ring clusters forming the alligned aggregates "turnover". The turnover rate for aggregates is governed by their ability to break and exchange hydrogen bonds, untangle their chains, reconform, shift position, rotate, undergo diffusion, and reform micelles, until the media is overall structurally reconfigured to give a net moment $< \mathbf{M} >_{time} = 0$. Is the rate of *aggregate turnover* consistent with the slow relaxation time τ_1?

Hydrogen Bond Lifetimes

Before treating that question, we must first rule out simple hydrogen-bond turnover as the mechanism responsible for τ_1. Over time, an initial set of hydrogen bonds formed at *t(0)* will break due to collisions and exchanges. Simultaneous with the destruction of the *t(0)* hydrogen-bond set, a new set is forming. At some time *t* in the future, none of the original hydrogen bonds will have survived; the new set will be totally uncorrelated with that at *t(0)*. Progressive changes in the hydrogen-bonding configuration will be reflected in the fluctuations of the bulk dipole moment **M**. Time-autocorrelation decay curves for hydrogen-bond endurance can be generated via the method described by Rapaport[73].

A quantity $B_{ij}(t)$ is defined to have a value of unity if molecules *i* and *j* are hydrogen bonded at time *t* and a value of zero if no bond is present. The set of values $\mathbf{B}_{ij}(t)$ for all pairs describes the hydrogen bonding configuration at a given instant *t*. Each initial hydrogen-bond pair $B_{ij}(t=0)$ is monitored through the time course of the MD trajec-

tory until it breaks at time t; at this point $B_{ij}(t)$ makes the transition from unity to zero. The time-autocorrelation function takes the form

$$C(\mathbf{B}_{ij}(t)) = \frac{\sum_{ij} B_{ij}(t_0)\, B_{ij}(t_0 + t)}{\sum_{ij} B_{ij}(t_0)} = A\, e^{-\frac{t}{\tau_B}} \qquad (13)$$

where the decay time constant τ_B, an indicator of continuous hydrogen-bond lifetime, is defined assuming the decay of $C(\mathbf{B}_{ij}(t))$ is exponential. The set $\mathbf{B}_{ij}(t)$ at the beginning of each of twenty-five 80 psec trajectories (40°C), or thirty-nine 100 psec trajectories (75°C), were taken as $t(0)$. The multiple decay samples, each containing 250 molecules, were then averaged to give $C(\mathbf{B}_{ij}(t))$. There are several possible definitions of hydrogen bond breakage involving either energetic or geometric criteria. A geometric definition was used here; when the heavy atom O-O distance exceeds 3.7 Å, the hydrogen bond is broken. Also, a bond may be severed but not considered permanently broken unless it *remains* severed for a specified period; bonds reformed within 0.5 picoseconds are retained as unbroken.

Figure 28 shows time-autocorrelation decays for hydrogen-bond lifetimes or "duration" times (solid lines) along with their double-exponential-fit functions (broken lines). The fit is excellent with RMSDs of 0.015 and 0.013 for 40°C (upper lines) and 75°C (lower lines) respectively. Two contributions τ_a and τ_b to the decay of hydrogen bonds were extracted. The pre-exponential coefficients are taken as percentage contributions. At 40°C, 82 percent of the decay is due to τ_a = 54.5 psec; another 11 percent is due to the minor contributor τ_b = 6.6 psec. At 75°C, 76 percent of the decay is due to τ_a

= 21.8 psec; another 16 percent is due to the minor contributor τ_b = 4.9 psec. The τ_a turnover rate for hydrogen bonds at 40°C and at 75°C is significantly faster than the dielectric relaxation time computed from the model ($\tau_1 \approx$ 176 to 207 psec at 40°C, and $\tau_1 \approx$ 40 to 55 psec at 75°C). Simple hydrogen-bond turnover can not be the rate limiting event that is ultimately responsible for the slow τ_1 dielectric response.

Note also the magnitude of the secondary mechanism contributing to hydrogen-bond decay. Hydrogen bonds are breaking via the τ_b mechanism at about the same rate as was computed earlier for the annihilation of free monomers. τ_b must reflect monomer (or *few*-mer) creation. It is reassuring that a process making a minor contribution to the calculated hydrogen-bond breakage time is consistent with the hydrogen-bond formation time computed by observing monomer annihilation. This lends further credence to our earlier proposal of a small set of itinerant molecules which rapidly (τ_2) exchange hydrogen bonds, reorienting in the process, and thereby influence the bulk moment fluctuations.

Figure 29 reports hydrogen-bond decays in water-saturated 1-octanol at 40°C. The upper-most line (long dashes) is for pure 1-octanol at 40°C, shown for comparison. The second-highest curve (short dashes) is for octanol-to-octanol hydrogen bond breakages. They turnover more rapidly in the presence of saturating waters. The third-highest curve (dotted) is for octanol-to-water bond breakage, and the bottom curve (solid) is for water-to-water hydrogen bond breakage. Hydrogen bonds involving waters are exchanged most rapidly. However, notice the long-term behavior of some water-to-water hydrogen bonds. While some water-water bonds are the most rapidly

exchanging of all, another set of water-water bonds are the most stable of all in the system. One might speculate that these enduring water-water hydrogen bonds are between the relatively few molecules lodged tightly within the core of inverted micellar aggregates.

Microstructure Turnover

Are the long term fluctuations of the bulk dipole moment directly correlated to the evolution of the micro-structure configuration? Let us determine whether the two processes share the same long-term rate of decay. Just as the internal structure can be characterized on one level by identifying distinct micro-structures (clusters, and mini-micellar aggregates) as described here earlier, the turnover rate of these structures can be determined by computing the time-correlation of their annihilation, as was done for hydrogen bonds. At any time t, say $t = 0$, there exists a uniquely configured set of identifiable micro-structures (clusters) $\mathbf{S}_{CL}(0)$. At some future time the ensemble of micro-structures composing $\mathbf{S}_{CL}(t)$ will be entirely un-correlated with those at $\mathbf{S}_{CL}(0)$. Although individual molecular positions may have changed only nominally, the formation of a renewed set of micro-structures may grossly alter the overall hydrogen-bonding configuration, giving rise to a new instantaneous direction and magnitude for $\mathbf{M}$. If the turnover rate for micro-structural configurations is equivalent to the rate of decay of long-term fluctuations in $\mathbf{M}$, then we might associate this molecular level process with the macro-level τ_J dynamic relaxation.

While correlation does not prove cause and effect, the requirement for consistency would be satisfied, and the explanation would be physically reasonable in the light of our newly enhanced understanding of octanol's internal structure. The time-autocorrelation function for the decay of aggregates takes a form very similar to hydrogen-bond lifetime decay. Just substitute $\mathbf{S}_{CL}$ for $\mathbf{B}_{ij}$ and τ_{CL} for τ_B in equation 13, where the decay-time constant τ_{CL}, an indicator of continuous aggregate lifetime, is defined assuming the decay of $C(\mathbf{S}_{CL}(t))$ is exponential. The calculation was carried out exactly as described above for hydrogen-bond lifetimes, except that clusters rather than bonds are the micro-structural entities "turning over". An aggregate cluster of molecules is considered annihilated when the hydrogen bonds originally connecting it together have *all* undergone breakage.

Figure 30 shows the decay of $C(\mathbf{S}_{CL}(t))$ at 40°C (upper solid line) and 75°C (lower solid line) along with single-exponentially fit functions (broken lines). The computed decay times are τ_{CL}(40°C) = 138.2 psec, and τ_{CL}(75°C) = 37.3 psec. Although the RMSDs were 0.20 and 0.71 respectively, it is not unlikely that the value of 138.2 in particular would change somewhat if more extended data were available. The pre-exponential coefficients are both 1.0; no fit-function improvement was possible by using multiple exponentials. In the most convergent simulation, the 75°C study (lower line in Figure 30), the cluster turnover time is almost identical to the computed τ_1 relaxation time (37.3 versus 38.0). In the 40°C simulation, the cluster turnover time accounts for about 80 percent (138.2 versus 176.0) of τ_1.

Figure 31 simultaneously displays four different time-correlation functions, all for pure 1-octanol at 75°C: $C(\mathbf{M}(t))$, $C(\mathbf{S}_{CL}(t))$, $C(\mathbf{B}_{ij}(t))$, and $C(\mathbf{F}_i(t))$. Between 0.0 and 50 psec, the bulk polarization (solid line) seems to extrapolate between the decay curves for clusters (upper long dashes) and shorter-term processes, namely individual hydrogen-bond decay (short dashes). But notice how the correlation of bulk moment fluctuations overlaps the turnover of aggregate-clusters in the long-term. These results strongly support our proposal that the τ_1 relaxation time is probably due to the overall restructuring of the media via the simultaneous breakdown and reformation of the various aggregate species.

Direct Dependence Of δM(t) On Aggregate Turnover

While out results show correlation of $\delta\mathbf{M}$ with microstructural restructuring events by virtue of their shared time-scales, we have not directly demonstrated cause and effect. We have not *proven* that cluster turnover is the mechanism responsible for τ_1. But, the most reasonable alternative mechanism has been ruled out. In order for any process to make a significant contribution to the dielectric relaxation, it must involve a drastic rearrangement of the polarizable moieties, namely the alkoxyl groups. Other possible explanations for the molecular mechanism responsible for the τ_1 relaxation would require arguing for an intra-aggregate or inter-aggregate reorientational mechanism which does *not* involve hydrogen-bond breakage. Because the cluster decay-time (which requires the annihilation of all intra-aggregate hydrogen bonds) occurs on approximately the same time scale as the τ_1 relaxation, no such arguement could be

made. The clusters do not survive long enough to accommodate such a process. Therefore, it is likely that the turnover and concomitant reorientation of the molecules constituting the aggregate micro-structures in 1-octanol is the very mechanism responsible for the τ_1 relaxation.

Some Intriguing Information

Figure 32 shows the actual raw fluctuations in the bulk moment $\delta\mathbf{M}(t)^2$ as a function of time between 400 and 2000 picoseconds at 40°C. The points represent individual values of $\delta\mathbf{M}(t)^2$, reported here every 50 MD steps, every 0.1 psec that is. While 1600 psec is a relatively short time span relative to $\tau_1 \approx 200$ psec, the data set from which this data was gleaned contained 800,000 values of the vector **M** and the accompanying 16,000 complete coordinate sets, each of 2500 1-octanol atoms. The fluctuations are overlayed with their 100 psec moving average. (Average a point *p* with all of the neighboring points between *p-50* and *p+50* psec to either side). One can see how a low frequency wave (note the alternating crests and troughs), with oscillations of *about* 200 psec might be extracted from this data. There are several regions over which $\delta\mathbf{M}(t)^2$ maintains a value that is "high" or "low" relative to the average over a significant stretch of time. There are high $\delta\mathbf{M}(t)^2$ regions between 400-425, 775-850, and 1375-1700 psec. There are low $\delta\mathbf{M}(t)^2$ regions between 950-1075, 1275-1350, and 1725-1850 psec. Cluster distribution analyses were performed over the coordinate sets corresponding to these isolated regions.

Figure 35 is analogous to the inset histograms in Figure 5 discussed above. But rather than present the data as histograms, lines were drawn between the points which would correspond the tops of the vertical bars in a histogram. The focus is on the peak regions between 1 and 10 molecules per cluster. Solid lines refer to the high $\delta\mathbf{M}(t)^2$ states, broken lines refer to low $\delta\mathbf{M}(t)^2$ states. The upper-most lines (peaked at X = 5 and X = 6) report the frequency of closed-ring clusters of the size indicated along the X-axis. The lower lines (after X = 3) report the frequency of open chains. Notice the well-defined shift in the distribution from six-membered rings preferred at high $\delta\mathbf{M}(t)^2$ to five- and seven-membered rings at low $\delta\mathbf{M}(t)^2$. This was a bit of a surprise; there were no other cluster sizes where such a strong difference signal was detectable.

The entire 400-2000 psec data set was then searched for five- and six-membered clusters, the largest difference signals. The findings are expressed in Figures 33 and 34 as 100 psec moving averages of 5- and 6-membered species frequencies superimposed onto the moving average of $\delta\mathbf{M}(t)^2$. The six-membered ring frequency oscillations are essentially in phase with the major signals in $\delta\mathbf{M}(t)^2$, showing strong overlap. The five-membered ring oscillations are out of phase, showing peaks where $\delta\mathbf{M}(t)^2$ has troughs and vice versa. Could the dynamic equilibrium between six-membered and five/seven-membered rings be responsible for the τ_j relaxation? How could a six-membered ring accentuate the bulk polar moment? One would expect a six-membered ring to have an aggregate dipole of low magnitude.

Upon examination of the individual six- and five-membered aggregates in the afore-mentioned high and low $\delta\mathbf{M}(t)^2$ regions, it was found that they did not possess

anomalously high or low dipoles. The average six-membered ring cluster had an aggregate dipole of 3.3 ± 1.6, and the five-membered rings were comparable. It is apparent that these clusters are stable and enduring. Perhaps their presence in the media promotes some organization in neighboring molecules, which in turn causes an increase or decrease in the bulk moment. On the other hand, with an extension of the data we might see that this allignment between specific aggregate sizes and the bulk moment is merely coincidental. While this particular part of the study is inconclusive, it is certainly intriguing. This method of inquiry could be used on a much longer data set to further probe the role of specific cluster types in the τ_1 dielectric response mechanism.

CHAPTER II SUMMARY

A thorough examination of the internal structure of 1-octanol and water-saturated 1-octanol has been done using molecular modeling. We have described the various micro-structures present in 1-octanol and demonstrated their influence upon the dynamic motions in the media which manifest as the macroscopic dielectric response. New insights have been gained into the molecular-level mechanisms underlying 1-octanol's three dielectric relaxations. The difference in the octanol/water partition coefficients of two small solutes (phenol and benzene) have been reproduced via the free energy perturbation method. It was found that water-saturated 1-octanol's capability to serve as a biophase substitute is due to the formation of preferentially hydrophilic versus hydrophobic regions in the media, mimicking the compartmental boundaries of the body. Hydrated 1-octanol can accommodate either polar or nonpolar substituent

REFERENCES: CHAPTER II

(1) R. P. Mason, D. G. Rhodes, L. G. Herbette, "Reevaluating Equilibrium and Kinetic Binding Parameters for Lipophilic Drugs Based on a Structural Model for Drug Interaction with Biological Membranes", *J. Med. Chem.*, **34**, 867-77 (1991).

(2) L. G. Herbette, M. Trumbore, D. W. Chester, A. M. Katz, "Possible Molecular Basis for the Pharmacokinetics and Pharmacodynamics of the Three Membrane-Active Drugs Propranolol, Nimodipine, and Amiodamine", *J. Molecular and Cellular Cardiology*, **20**, 373-378 (1988).

(3) J. J. Wendoloski, S. J. Kimatian, C. E. Schutt, F. R. Salemme, *Science*, **243**, 636 (1989).

(4) S. Karaborni, J. P. O'Connell, *J. Phys. Chem.*, **94**, 2624 (1990).

(5) K. Watanabe, M. L. Klein, *J. Phys. Chem.*, **93**, 6897 (1989).

(6) J. Shelley, K. Watanabe, M. L. Klein, *International Journal of Quantum Chemistry: Quantum Biology Symposium*, **17**, 103-117 (1990).

(7) B. Jonsson, O. Edholm, O. Teleman, *J. Chem. Phys.*, **85**, 2259 (1986).

(8) K. V. Damodaran, K. M. Merz, *Langmuir*, **9**, 1179-83 (1993); and K. V. Damodaran, K. M. Merz, B. P. Gaber, *Biochemistry*, **31**, 7656 (1992).

(9) S. J. McKinnon, S. L. Whittenburg, B. R. Brooks, *J. Phys. Chem.*, **96**, 10497 (1992).

(10) E. Egberts, H. J. C. Berendsen, *J. Chem. Phys.*, **89**, 3718 (1988).

(11) O. Edholm, J. Johansson, *Eur. Biophys. J.*, **14**, 203 (1987).

(12) P. van der Ploeg, H. J. C. Berendsen, *J. Chem. Phys.*, **76**, 3271 (1982); and *Mol. Phys.*, **49**, 233 (1983).

(13) R. N. Smith, C. Hansch, M. M. Ames, "Selection of a Reference Partitioning System for Drug Design Work", *J. Pharm. Sci.*, **64**, 599 (1975).

(14) A. Leo, C. Hansch, D. Elkins, "Partition Coefficients and Their Uses", *Chem. Revs.*, **71**, 525 (1971).

(15) C. Hansch and A.J. Leo, "Substituent Constants for Correlation Analysis in Chemistry and Biology", Wiley, NY (1979); and Hansch C. and Dunn W.J., *J. Pharm. Sci.*, **61** p1 (1972).

(16) G. L. Flynn, *J. Pharm. Sci.*, **60**, 345 (1971).

(17) D. E. Burton, K. Clarke, G. W. Gray, *J. Chem. Soc.*, 1314 (1964).

(18) G. V. Betageri, J. A. Rogers, "Correlation of Partitioning of Nitroimidazoles in the n-Octanol/Saline and Liposome System with Pharmacokinetic Parameters and Quantitative Structure-Activity Relationships (QSAR)", *Pharmaceut. Revs.*, **6**, 399 (1989).

(19) F. Gaspari, M. Bonati, "Correlation Between n-Octanol/Water Partition Coefficients and Liquid Chromatography Retention for Caffeine and its Metabolites, and Some Structure-Pharmacokinetic Considerations", *J. Pharm. Pharmacol*, **39**, 252 (1987).

(20) M. P. Allen, D. J. Tildesley, "Computer Simulation of Liquids", Chapt. 5, 162, Clarenden Press, Oxford, (1987).

(21) W. L. Jorgensen, J. D. Madura, C. J. Swenson, "Optimized Intermolecular Potential Functions for Liquid Hydrocarbons", *J. Amer. Chem. Soc.*, **106**, 6638 (1984).

(22) W. L. Jorgensen, "Optimized Intermolecular Potential Functions for Liquid Alcohols", *J. Amer. Chem. Soc.*, **90**, 1276 (1986).

(23) U. Burkert, N. Allinger, "Molecular Mechanics", American Chemical Society, Washington D. C. (1982).

(24) W.F. van Gunsteren, H.J.C. Berendsen, *Mol. Phys.*, **34**, 1311 (1977). J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, *J. Comput. Phys.*, **34**, 327 (1977). G. Ciccotti, M. Ferrario, J.P. Ryckaert, *Mol. Phys.*, **47**, 1253 (1982).

(25) H.J.C. Berendsen, J.P.M. Postma, W. van Gunsteren, J. Hermans, in *Intermolecular Forces*, 331-342; B. Pullman, ed., Reidel: Dordrecht, The Netherlands (1981).

(26) L. von Erichsen, *Brennst. Chem.*, **33**, 166-172 (1952).

(27) S. E. DeBolt, P. A. Kollman, *J. Amer. Chem. Soc.*, **112**, 7515 (1990).

(28) H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.*, **81**, 3684 (1984).

(29) W. L. Jorgensen, T. B. Nguyen, *J. Comp. Chem.*, **14**, 195 (1993).

(30) P. T. T. Wong, H. H. Mantsch, R. G. Snyder, "Effects of Pressure on Conformer Equilibrium in Liquid n-Hexane", *J. Chem. Phys.*, **79**, 2369 (1983).

(31) C. R. Cantor, P. R. Schimmel, "Biophysical Chemistry" III. The Behavior of Biological Molecules", 194, W. H. Freeman and Company, New York (1971).

(32) P. Berti, S. Cabani, V. Mollica, *Fluid Phase Equilibria*, **32**, 195-203 (1987).

(33) Y. Marcus, "Structural aspects of Water in 1-Octanol", *J. Solution Chem.*, **19**, 507 (1990).

(34) T.R. Dyke, K.M. Mack, J.S. Muenter, *J. Chem. Phys.*, **66**, 498 (1977).

(35) E. Grunwald, K. C. Pan, *J. Phys. Chem.*, **80**, 2929 (1976).

(36) W. Saenger, *Ann. Rev. Biophys. Biophys. Chem.*, **16**, 93-114 (1987) "Structure and Dynamics of Water Surrounding Biomolecules".

(37) A. N. Fletcher, C. A. Heller, *J. Phys. Chem.*, **71**, 3742 (1967).

(38) M. Iwahashi, Y. Hayashi, N. Hachiya, H. Matsuzawa, A. Kabayashi, *J. Chem. Soc. Faraday Trans.*, **89**, 707 (1993).

(39) W. Dannhauser, *J. Chem. Phys.*, **48**, 1918 (1968).

(40) T. Shinomiya, *Bull. Chem. Soc. Jpn.*, **62**, 3643 (1989); and **62**, 2258 (1989).

(41) M. P. Allen, D. J. Tildesley, 60 and 204, Clarenden Press, Oxford, (1987).

(42) R. Mills, *J. Phys. Chem.*,**77**, 685 (1973).

(43) J. W. Essex, C. A. Reynolds, W. G. Richards, "Relative Partition Coefficients from Partition Functions: A Theoretical Approach to Drug Transport", *J. Chem. Soc. Commun.*, 1152-1154 (1989).

(44) W. L. Jorgensen, J. M. Briggs, M. L. Contreras, "Relative Partition Coefficients for Organic Solutes from Fluid Simulations", *J. Phys. Chem.*, **94**, 1683 (1990).

(45) T. Fujita, J. Iwasa, C. Hansch, *J. Amer. chem. Soc.*, **86**, 5175 (1964).

(46) S. K. Garg, C. P. Smyth, *J. Phys. Chem.*, **69**, 1294 (1965).

(47) J. Crossley, L. Glasser, C. P. Smyth, *J. Chem. Phys.*, **55**, 2197 (1971).

(48) J. Anderson, J. J. Ullo, S. Yip, *J. Chem. Phys.*, **87**, 1726 (1987).

(49) M. Neuman, *J. Chem. Phys.*, **82**, 5663 (1985).

(50) D. Eisenberg, W. Kauzmann, "The Structure and Properties of Water", Oxford at the Clarendon Press, p.189-94 (1969); and G. Oster, J. G. Kirkwood, *J. Chem. Phys.*, **11**, 175 (1943); and J. A. Pople, *Proc. R. Soc.* A, **205**, 163 (1951); and L. Onsager, "Electric Moments of Molecules in Liquids", eqn. 30, *J. Amer. Chem. Soc.*, **58**, 1486, (1936).

(51) D. Chandler, "Introduction to Modern Statistical Mechanics", Chapt. 8, 241, Oxford University Press, New York, (1987).

(52) J. G. Kirkwood, "The Polarization of Polar Liquids", *J. Chem. Phys.*, **7**, 911-919 (1939).

(53) H. Frolich, "General Theory of the Static Dielectric Constant", *Trans. Faraday Soc.*, **44**, 238-44 (1948); and H. Frolich, "Theory of Dielectrics", Clarendon Press, Oxford (1958).

(54) M. Neumann, *Mol. Phys.*, "Dipole Moment Fluctuation Formulas in Computer Simulations of Polar Systems", **50**, 841 (1983).

(55) M. Neumann, O. Steinhauser, "On the Calculation of the Frequency-Dependent Dielectric Constant in Computer Simulations", *Chem. Phys. Letts.*, **102**, 508 (1983).

(56) B. K. P. Scaife, "Principles of Dielectrics", Clarendon Press, Oxford, 203 (1989).

(57) O. Steinhauser, "On the orientational structure and dielectric properties of water, a comparison of ST2 and MCY potential", *Ber. Bunsenges. Phys. Chem.*, **87**, 128-142 (1983).

(58) H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, "The Missing Term in Effective Pair Potentials", *J. Phys. Chem.*, **91**, 6269 (1987).

(59) M. Neumann, O. Steinhauser, G. S. Pawley, *Mol. Phys.*, **52**, 97-113 (1984).

(60) D. Bertolini, M. Cassettari, G. Salvetti, *J. Chem. Phys.*, **76**, 3285 (1982).

(61) M. Neumann, *J. Chem. Phys.*, **85**, 1567 (1986).

(62) K. Watanabe, M. L. Klein, *Chem. Phys.*, **131**, 157 (1989).

(63) M. R. Reddy, M. Berkowitz, *Chem. Phys. Letts.*, **155**, 173 (1989).

(64) H. E. Alper, R. M. Levy, *J. Chem. Phys.*, **91**, 1242 (1989).

(65) M. P. Allen, D. J. Tildesley, 186, Clarenden Press, Oxford, (1987).

(66) W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, "Numerical Recipes, The Art of Scientific Computing", Chapt. 14, Cambridge University Press, Cambridge (1989).

(67) B. Gestblom, J. Sjöblom, *Acta Chemica Scandinavica A*, **38**, 47 (1984).

(68) S. Mizushima, "Structure of Molecules and Internal Rotation", Academic Press, Inc. New York (1954).

(69) H. A. Semulon, *Proc. I. R. E.*, **39**, 175 (1951).

(70) W. Dannhauser, A. F. Flueckinger, *Phys. Chem. Liquids*, **2**, 37 (1970).

(71) L. Glasser, J. Crossley, C. P. Smyth, *J. Chem. Phys.*, **57**, 3977 (1972).

(72) T. Shinomiya, *Bull. Chem. Soc. Jpn.*, **62**, 908; and **62**, 3636 (1989).

(73) D. C. Rapaport, *Mol. Phys.*, 1151-62 (1983).

(74) S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A Case, *J. Comp. Chem.*, **7**, 230 (1986). S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, CC. Ghio, G. Alagona, S. Profeta, Jr., P. Weiner, *J. Amer. Chem. Soc.*, **106**, 765 (1984).

Table I. **HEXANE CONFORMATIONAL DATA**

confn	WMS RIS[a] ΔE=0.5	WMS RIS ΔE=0.9	AMBER MD
ttt	18.9	36.1	39.2
gtt	32.7	31.9	24.2
tgt	16.4	16.0	13.7
gtg	14.2	7.0	13.6
ggt	15.6	7.9	7.5
ggg	3.3	1.0	1.9

	Outer angles[b]	Inner angle
Exptl (WMS ΔE=0.5)		
Gauche	41.7	35.3
Trans	59.5	65.8
Exptl (WMS ΔE=0.9)		
Gauche	27.9	24.9
Trans	72.0	75.0
OPLS_MC		
Gauche	30.9	24.3
Trans	69.1	75.7
AMBER_MD[c]		
Gauche	31.3	23.1
Trans	68.7	76.9

[a] RIS: Rotational Isomeric States modeling of conformational distributions correlated with the raman spectra results of Wong, Mantsch, and Snyder[30]. ΔE is the energy difference between gauche and trans energy wells in kcal/mol. OPLS/MC results are from reference 21. AMBER/MD results are from this study. Values in columns are the percentage (on average) of molecules found in the listed conformations.

[b] Average of the two outer torsion angles of hexane in gauche or trans configurations, given as a percent.

Table II. Molecular-Mechanical Interaction Parameters For Hydrocarbon, Hydroxyl, And SPC Water [a]

VDW_TYPE	R^*	ε	Q	DESCRIPTION
OW	1.7765	0.15545	-0.810	-O- SPC water
HW	0.0000	0.00000	0.420	-H- SPC water
HL	0.0000	0.00000	0.435	-H- 1-alkanol H-O hydrogen
Oh	1.7230	0.17000	-0.700	-O- 1-alkanol O-H oxygen
Mo	2.1915	0.11800	0.265	-CH2- 1-alkanol C-O carbon
ML	2.1915	0.11800	0.000	-CH2- n-alkane chains
Me	2.1915	0.15000	0.000	-CH3 alkane end methyl

BOND	FC	R_{eq}
HL-Oh	553.000	0.945
Oh-Mo	386.000	1.430
Mo-ML	260.000	1.530
ML-ML	260.000	1.530
ML-Me	260.000	1.530
OW-HW	553.000	1.000
HW-HW	553.000	1.633

ANGLE	FC	τ_{eq}
HL-Oh-Mo	55.000	108.500
Oh-Mo-ML	80.000	112.000
Mo-ML-ML	63.000	112.000
ML-ML-ML	63.000	112.000
ML-ML-Me	63.000	112.000
HW-OW-HW	100.000	109.470
OW-HW-HW	100.000	35.265

TORSION	V_N	SHIFT	PERIOD
HL-Oh-Mo-ML	0.40	0.000	3
HL-Oh-Mo-ML	0.36	0.000	1
Oh-Mo-ML-ML	1.42	0.000	3
Oh-Mo-ML-ML	-0.08	180.000	2
Mo-ML-ML-ML	1.42	0.000	3
Mo-ML-ML-ML	-0.08	180.000	2
ML-ML-ML-ML	1.42	0.000	3
ML-ML-ML-ML	-0.08	180.000	2
Me-ML-ML-ML	1.42	0.000	3
Me-ML-ML-ML	-0.06	180.000	2
Me-ML-ML-Me	1.42	0.000	3
Me-ML-ML-Me	-0.06	180.000	2

[a] Force constants FC are in kcal/mol, bond lengths R_{eq} in angstroms, angles τ_{eq} in degrees. Geometries are OPLS[21,22], force constants are from the Weiner, et al. force field[74]. Bond force constant's are not critical since simulations used SHAKE. The two-term dihedral barrier heights are actually $V_N/2$ as input to AMBER. SHIFT is the phase shift, and PERIOD is the numerical periodicity for the torsion angle. R^* and ε are the radius and well depth for Lennard-Jones VDW-type nonbonded atomic interactions, Q is the partial atomic charge.

Table III. Physical Properties of Pure Liquids[a]

	ΔH_{vap}	Volume/molec
HEXANE:		
EXPT[b]	7.54	218.5
OPLS_MC[c]	8.07 (7.0%)	211.5 (3.3%)
AMBER_MD[d]	7.25 (3.8%)	215.9 (1.2%)
OCTANE:		
EXPT[b]	9.91	272.1
AMBER_MD	9.72 (1.9%)	273.9 (0.7%)
OCTANOL:		
EXPT[b]	15.60	262.8
AMBER_MD	16.39 (5.0%)	265.2 (0.9%)

[a] Heat Of Vaporization ΔH_{vap} and Volume per molecule data for simulations of 125 (hexane, octane) or 250 (octanol) molecules. Volume in $Å^3$, ΔH_{vap} in kcal/mol. The percent deviation of computed values from experimentally determined values are given in parentheses. The computed standard deviations are less than 1.0 $Å^3$ for volumes, and slightly less than 0.1 kcal/mol for ΔH_{vap} values. Temperature was 298K for all simulations and for experimental values. Results are after averaging over several MD liquid simulations summing to 100-300 psec after extensive system relaxation, and 1.0-2.0 nsec for individual gas phase molecule MD simulations (details in text).

[b] Selected values of Properties of Chemical Compounds, Manufacturing Chemist's Assn. Physical and Chemical Reference Data, V.2 (1973).

[c] As per Monte Carlo simulations using OPLS parameters by W. L. Jorgensen in reference 21.

[d] AMBER_MD parameters as per Table I, scaled by 1/7.4 for intramolecular. Used 9.0 Å (hexane, octane) or 10.0 Å (octanol) spherical nonbonded cutoff.

Table IV. Torsion Angle Distributions In Pure Liquid 1-Octanol[a]

TORSION	%GAUCHE	%TRANS
(1) HOCC	61.8	38.2
(2) OCCC	47.9	52.1
(3) CCCC	23.5	76.5
(4) CCCC	25.7	74.3
(5) CCCC	23.5	76.5
(6) CCCC	24.7	75.3
(7) CCCC	23.3	76.7
ALKYL AVG	24.1	75.9

CONFORMER	%	total
all trans:		
tttt	21.47	21.5
one gauche:		
tttgt	10.17	
tgttt	10.11	
ttgtt	8.78	
gtttt	8.22	
ttttg	6.85	44.1
two gauche:		
tgtgt	4.46	
tgttg	3.85	
gtgtt	3.82	
gttgt	3.35	
ttgtg	3.04	
gtttg	3.01	
tggtt	1.80	
ttggt	1.73	
tttgg	1.29	
ggttt	0.86	27.2
three gauche:		
gtgtg	1.20	
tgtgg	1.02	
tggtg	0.84	
ggttg	0.70	
gggtt	0.69	
gttgg	0.56	
ggtgt	0.48	
gtggt	0.44	
ttggg	0.42	
tgggt	0.34	6.7
four gauche:		
tgggg	0.30	
gggtg	0.13	
ggtgg	0.06	
ggggt	0.02	
gtggg	0.01	0.5
all gauche:		
ggggg	0.00	0.0

[a]Averages taken over 100 coordinate sets of 250 molecules, 1.0 psec apart, at 25°C.

Table V. Pure Liquid 1-Octanol Hydrogen Bonding and Cluster Analysis In 250 Molecule Box[a]

AVERAGE QUANTITIES	40 degC	75 degC
CLUSTERS per set	24.1	32.5
MOLECS per cluster	10.4	7.4
RINGS per set	16.6	15.7
CHAINS per set	9.2	18.4
MONOMERS per set	4.5	11.9
NON-DONORS (free H)	13.0	27.8
ACCEPTORS-ONLY	8.5	15.9
NON-ACCEPTORS (free 0)	25.9	40.0
DONOR-ONLY	21.4	28.1

SPECIES	% @40deg	% @75deg
MONOMERS	1.8	4.8
DONOR-ONLY	8.6	11.2
FREE OH OXYGEN	10.4	16.0
ACCEPTOR-ONLY	3.4	6.4
FREE OH HYDROGEN	5.2	11.1
FREE OH HYD EXPT[b]	5.4	11.2

[a]Averages taken over 62400 (40°C) or 78000 (75°C) coordinate sets, each set composed of 250 molecules. Notice that the number of clusters multiplied by the number of molecules per cluster, plus the number of monomers, gives approximately 250, the number of molecules in the simulation. Also note that the number of rings plus chains plus monomers does not exactly equal the average number of clusters. Discrepancies are due the formation of multiple rings within some large clusters and roundoff within the analysis program.

[b]Experimental numbers from Grunwald and Pan[35].

Table VI. Number Of Inter-Species Hydrogen Bonds In Water-Saturated 1-Octanol[a]

1st-SHELL NEIGHBORS	CENTER O_{OCT}	CENTER O_{WAT}
O_{OCT}	1.48	2.19
	(1.63)	(2.81)
O_{WAT}	0.73	1.61
	(0.58)	(0.99)
TOTALS	2.21	3.80

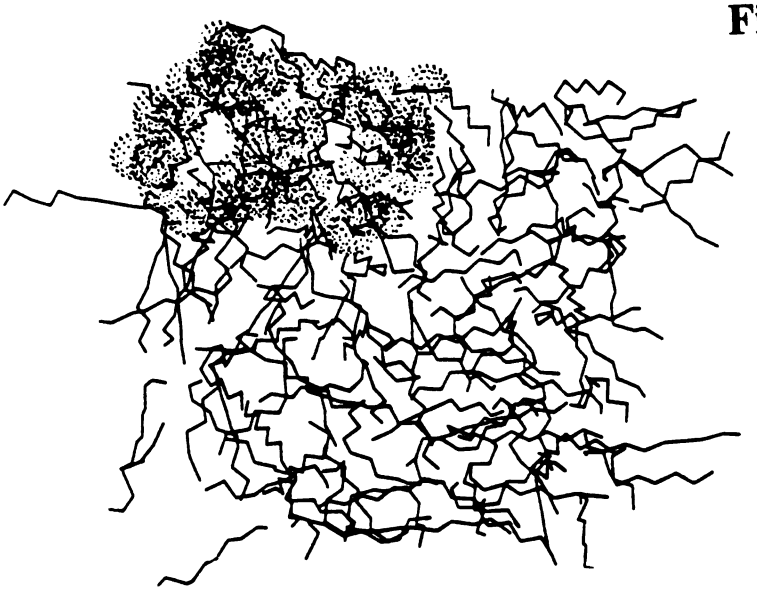
[a]Values arrived at by integrating the appropriate RDF from 0.0 to 3.5 Å. System of 228 octanol molecules and 76 waters (40°C). Parentheses indicate the number of hydrogen bonds which would result if hydroxyl pairing were random among all available hydrogen-bonding sites. Data gathered over 2.5 nsec (50000 coordinate sets).

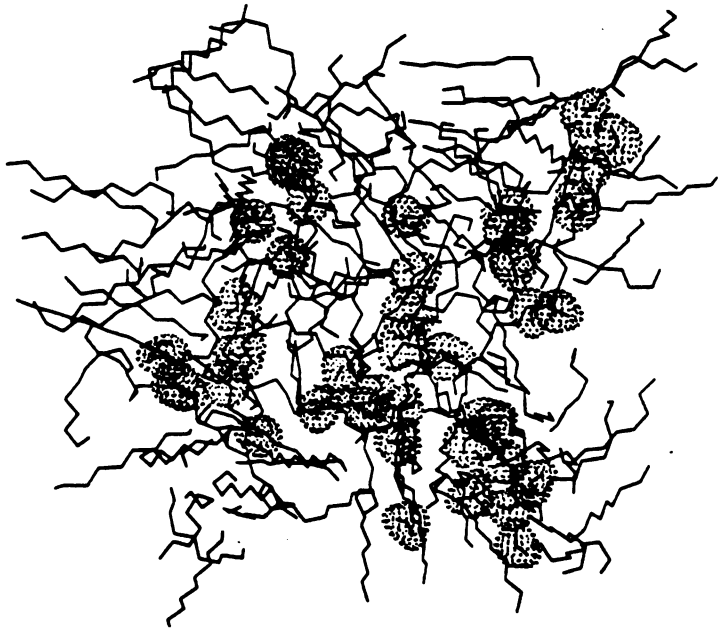
Table VII. Pure Liquid 1-Octanol Static Dielectric and Relaxation Time Analysis[a]

	40 degrees C				75 degrees C			
PROPERTY	EXPT	%CONTRIB	COMPUT	COEFF	EXPT	%CONTRIB	COMPUT	COEFF
τ_1 400ps fit			176.0	(89.)			38.0	(87.)
τ_1 3-exp fit	664.0	(85.3)	207.0	(83.)	162.0	(75.)	55.0	(70.)
τ_2 3-exp fit	29.5	(8.8)	6.6	(11.)	15.8	(15.)	9.6	(19.)
τ_3 3-exp fit	2.7	(5.9)	0.4	(5.)	1.9	(10.)	0.7	(9.)
ε_0	8.7		5.1		6.1		4.2	

[a]Experimental (EXPT) and computed (COMPUT) relaxation times τ_r and static dielectric constant ε_0 are reported at two temperatures. The percent contributions that each τ_r makes to the overall dielectric relaxation are given in parentheses (%CONTRIB and COEFF). Decay curves were computed over 2.0 nsec (40°C) or 3.3 nsec (75°C) trajectories. The single-exponential function was fit to a 400 picosecond segment of $C(M^2(t))$ decay curve. The triple-exponential fit was to a 30.0 psec segment of the decay curve. ε_0 was computed over 2.0 nsec (40°C) or 3.3 nsec (75°C). Experimental ε_0 and τ values were taken as the average of those from references 46 and 67. The 75°C "experimental" ε_0 values were linearly extrapolated from actual experimental data at 20°, 40°, and 60°C (correlation coefficient r = 0.999). Experimental %CONTRIB values are computed using experimental values for ε_{01}, $\varepsilon_{\infty 1}$, and $\varepsilon_{\infty 2}$, from reference 67, and use of $\varepsilon_{\infty 3} \approx 2.1$ from reference 46. The values of τ at 75°C were computed by first converting τ at 20°C, 40°C, and 60°C into activation free energies $\Delta F^{\ddagger}$ through Eyring's linearizing formula $\tau = h/(k_B T)\exp[-\Delta F^{\ddagger}/RT]$, extrapolating $\Delta F^{\ddagger}$ to 75°C (r = 0.998), then computing back through the formula to get τ(75°C).

Figure 1.





One Nanosecond MD Diffusion of Water into Octanol

The waters are represented by their oxygen van der Waals sphere, octanols as simple chains. The appropriate number of waters were loaded onto a pre-"equilibrated" octanol media, such that the final solution would have the octanol/water ratio that is found in water-saturated octanol.

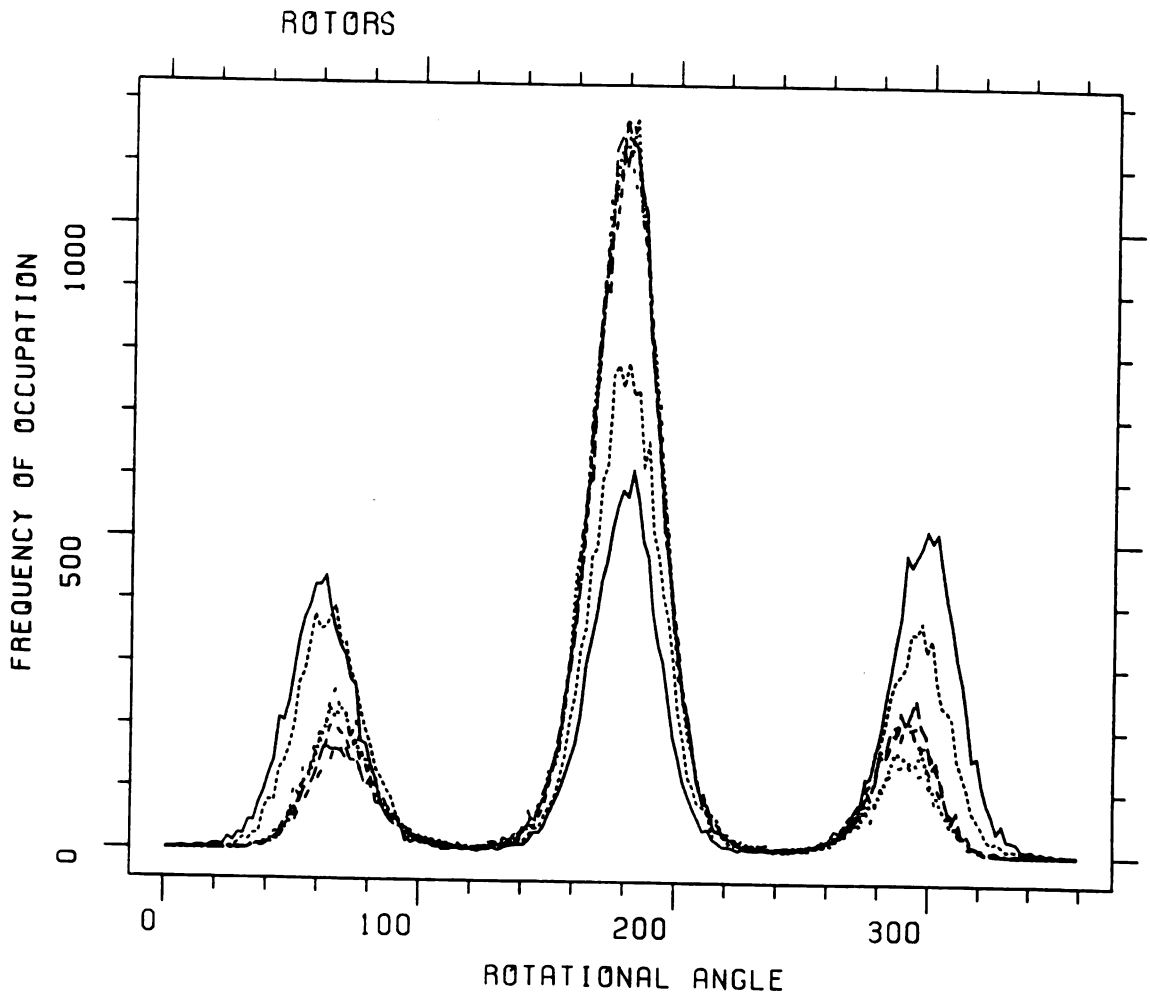


Figure 2. Rotation-angle frequencies for 1-octanol's seven torsions in the liquid phase at 25°C. Averages were taken over 100 psec in pre-equilibrated system of 250 molecules. From top-most to bottom-most lines in right-hand peak: HOCC angle (solid line), OCCC angle (broken line, short dashes), CCCC angles progress from remaining upper to lower lines as the end-methyl group is approached.

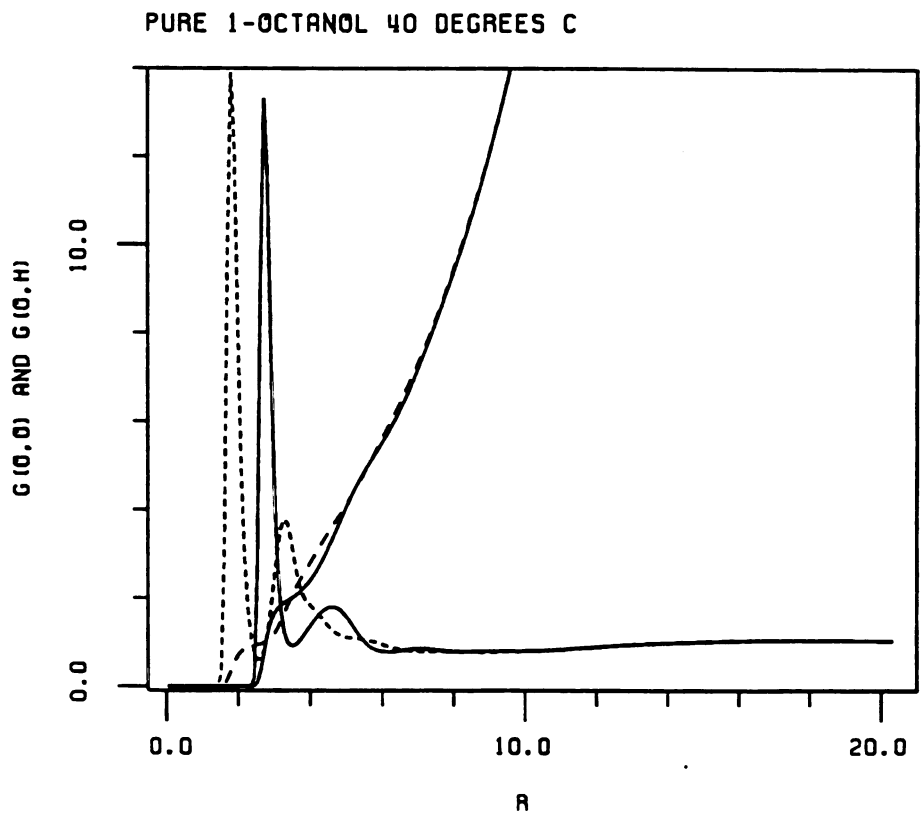


Figure 3. Radial pair distribution functions in pure 1-octanol. Plots are g(O,O) between hydroxyl oxygens (solid line), and g(O,H) between the oxygen and surrounding hydroxyl hydrogens (broken line). Integrating the RDF's gives the coordination numbers c(O,O) (solid line), and c(O,H) (broken line). Long-dash line at g=1 from R=0 to R=20 is for reference. Data from averaging 250 molecules of pure 1-octanol at 40°C, over more than 30000 coordinate sets, about 2 nsec.

Figure 4.

Hydrogen Bonded "Hydroxyl Polymers" in Pure Octanol

The dot-surface hemispheres represent one-half the van der Waals radius of a hydroxyl hydrogen. The straight lines represent the bond vector from the hydrogen to the hydroxyl oxygen. The direction of H-bonding is clear. Depending on the 3-D orientation of this sample, many chains and ring structures can be distinguished. Note the 5-membered ring in the upper-central region.

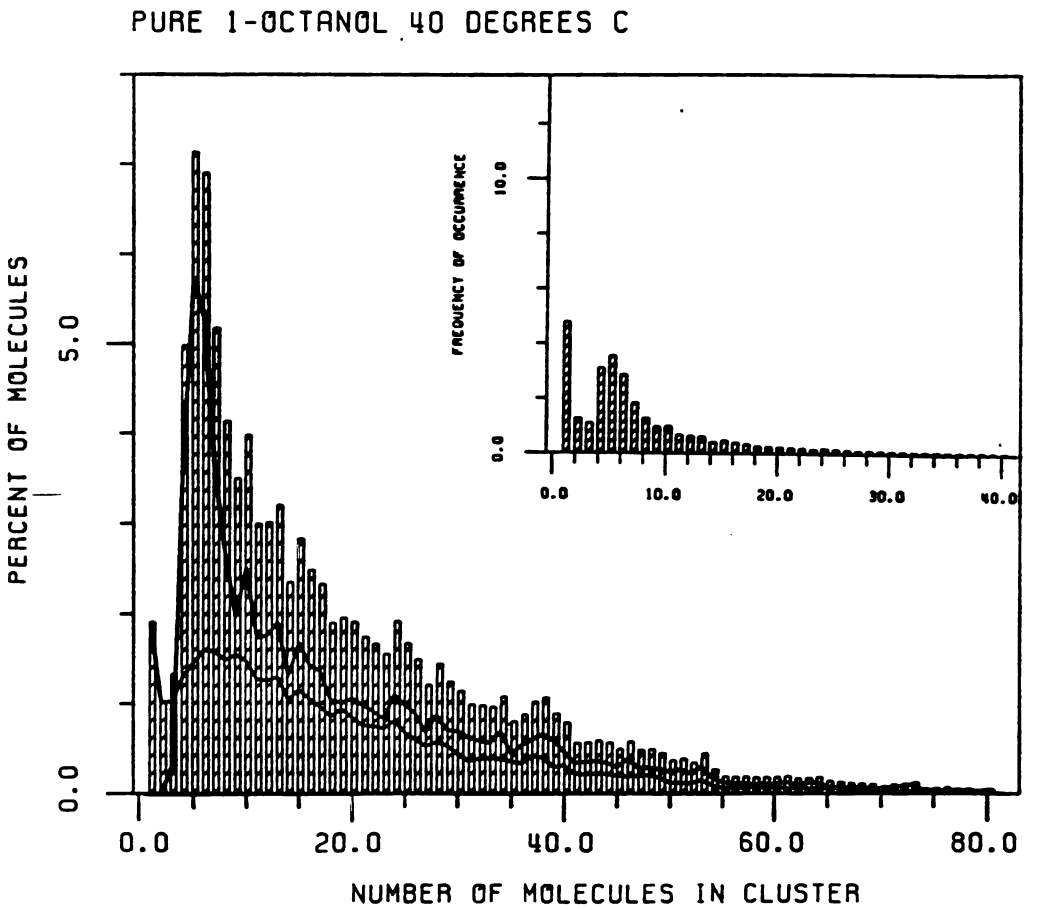


Figure 5. Frequency distribution of pure 1-octanol clusters of various sizes at 40°C. The main display reports the percent of molecules which participate in a cluster of a given size. The inset reports the actual number of clusters of a given size found in a periodic box of 250 molecules. Data collected as in Table V.

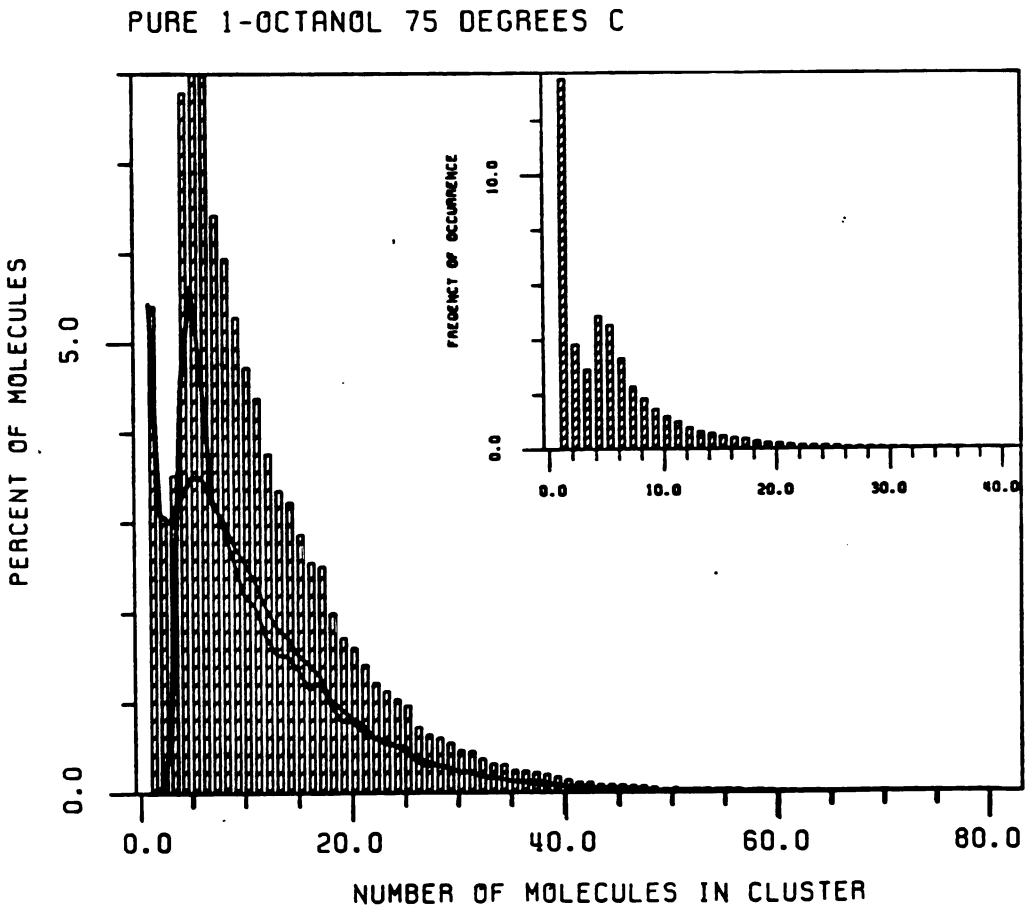


Figure 6. Frequency distribution of pure 1-octanol clusters of various sizes at 75°C, otherwise as in Figure 6.

Figure 7.

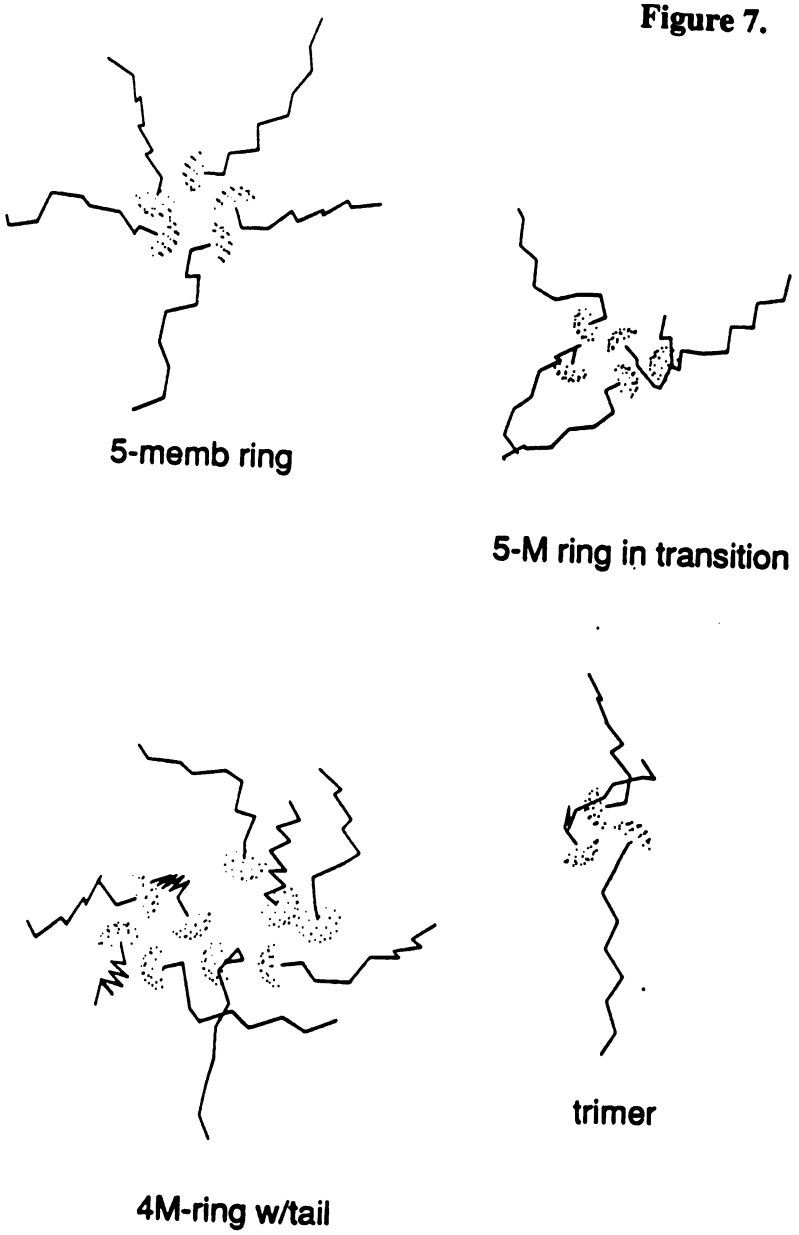
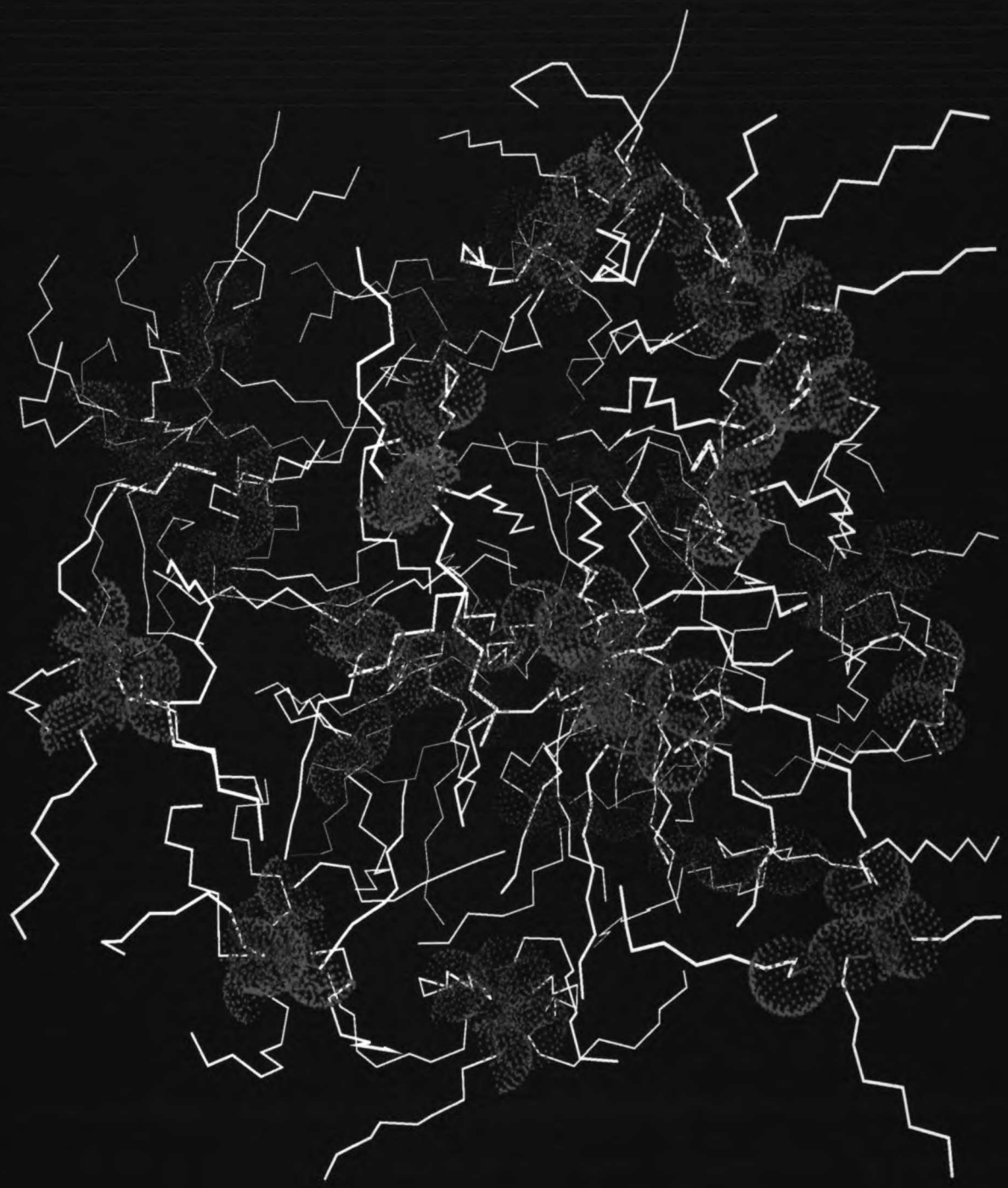


Figure 7. Hydrogen-bonded ring structures in 1-octanol.

Figure 8 p.172



pure 1-octanol clusters UCSF MidasPlus

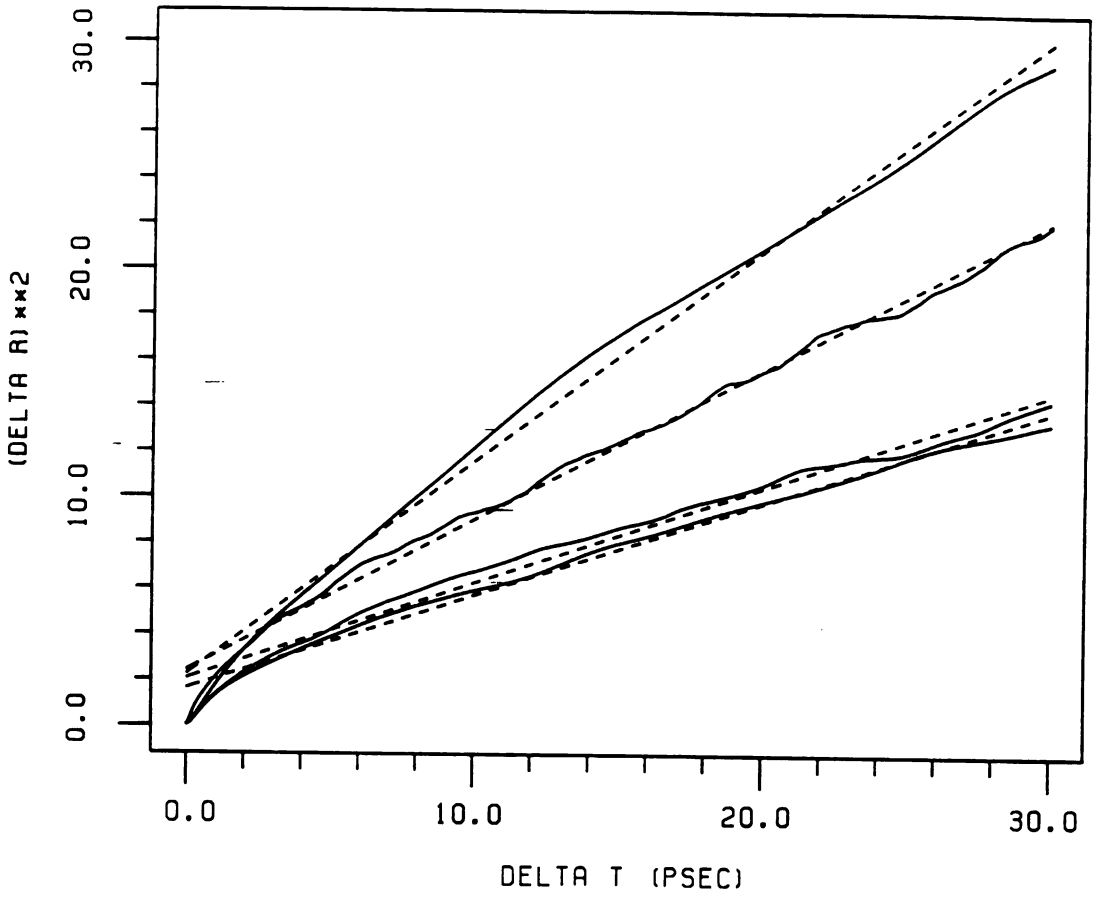


Figure 9. Plot of $\langle(\Delta r)^2\rangle$ versus Δt, where Δr is the difference in the position of the center of mass of a molecule at some time t and $t = 0$. Solid lines are actual data; broken lines are straight-line fits. Upper-most line is for pure 1-octanol at 75°C. Lower-most line is for pure 1-octanol at 40°C. Inner lines are for 40°C hydrated 1-octanol, where the higher line represents the diffusion of saturation waters, the lower line represents 1-octanols in the presence of saturating waters. Data gathered over 600 psec. Also see text.

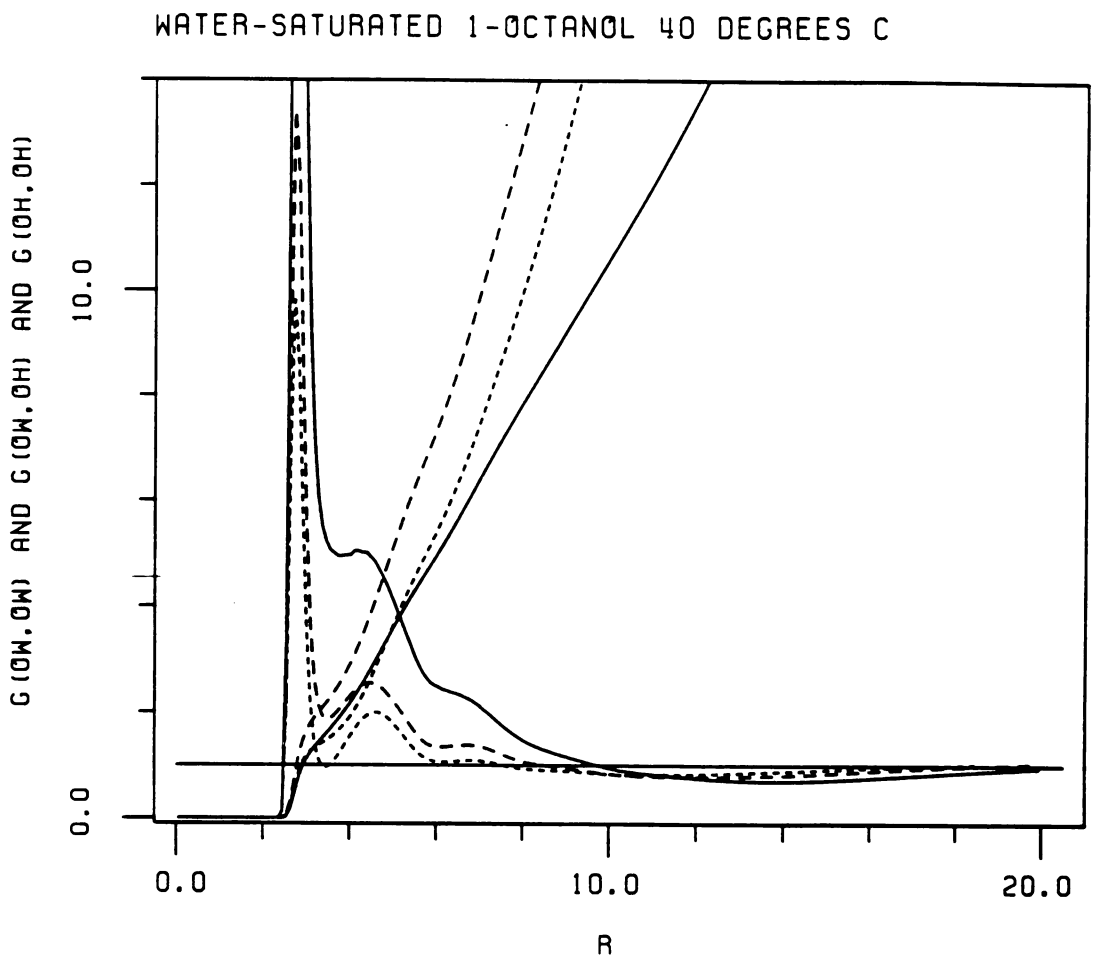


Figure 10. Radial pair distribution functions for water-saturated 1-octanol at 40°C. Plots are g(Ow,Ow) between water oxygens (solid line), g(Ow,Oh) between water oxygen and octanol oxygen (broken, long dashes), and g(Oh,Oh) between octanol oxygens (broken, short dashes). RDF g(Ow,Ow) rises to g=26.6 at R=2.7 Å. Coordination number plots follow same line designations. Data collected over 2.5 nsec (50000 coordinate sets).

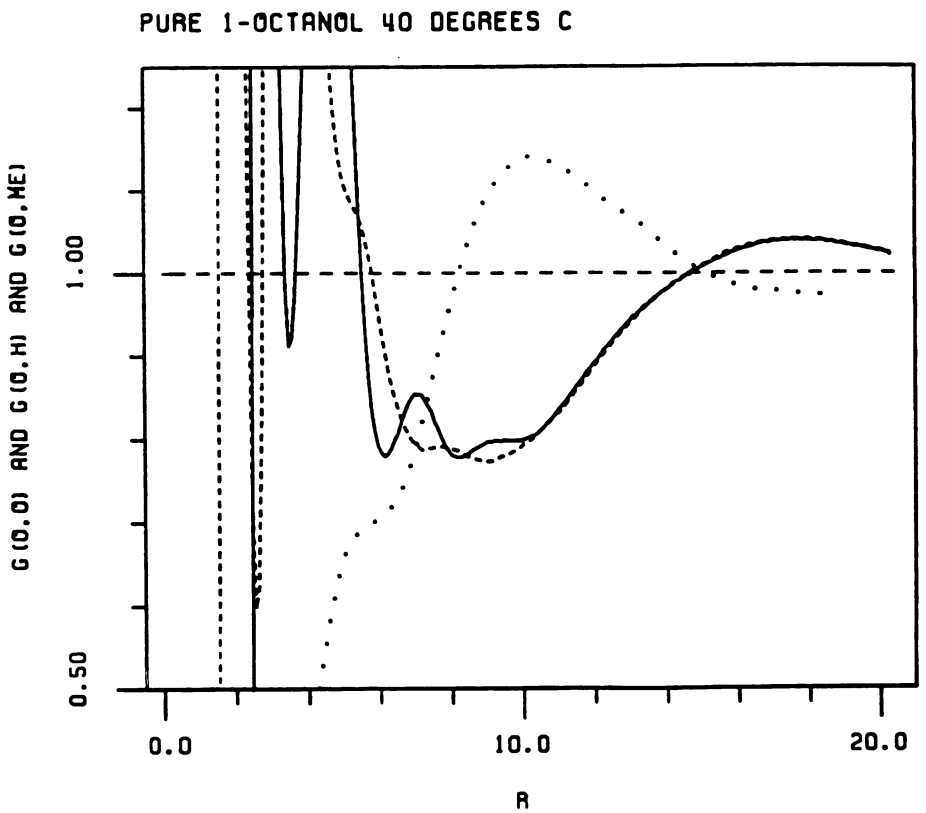


Figure 11. Closeup of radial pair distribution functions in pure 1-octanol. Plots are g(O,O) between hydroxyl oxygens (solid line), and g(O,H) between the oxygen and surrounding hydroxyl hydrogens (broken line). Dotted line is g(O,Me). Data from averaging 250 molecules of pure 1-octanol at 40°C, over more than 30000 coordinate sets, about 2 nsec.

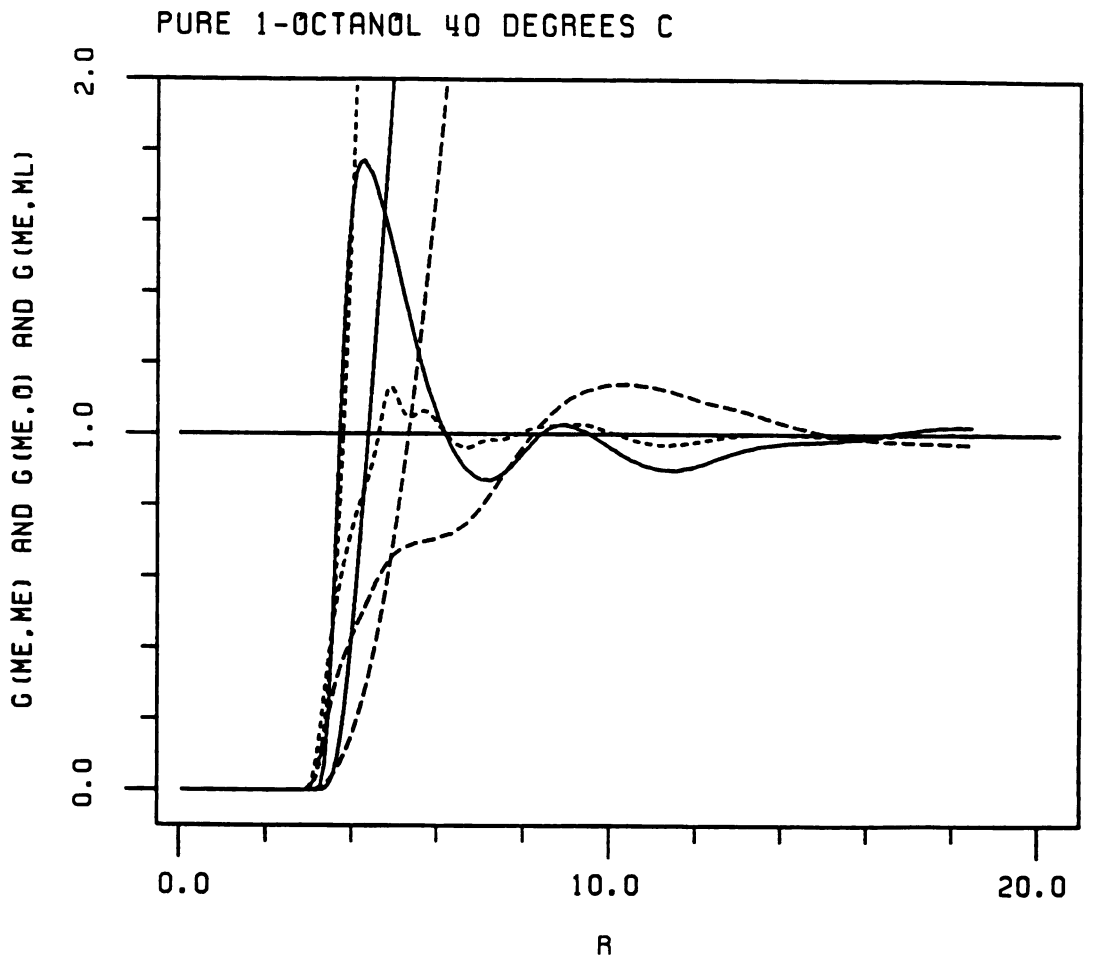


Figure 12. Radial pair distribution functions in pure 1-octanol. Plots are g(Me,Me) between end-methyls (solid line), g(Me,O) between end-methyls and hydroxyl oxygens (broken line, long-dashes), and g(Me.ML) between end-methyls and methylenes (short-dashes). Data collected as in Figure 11.

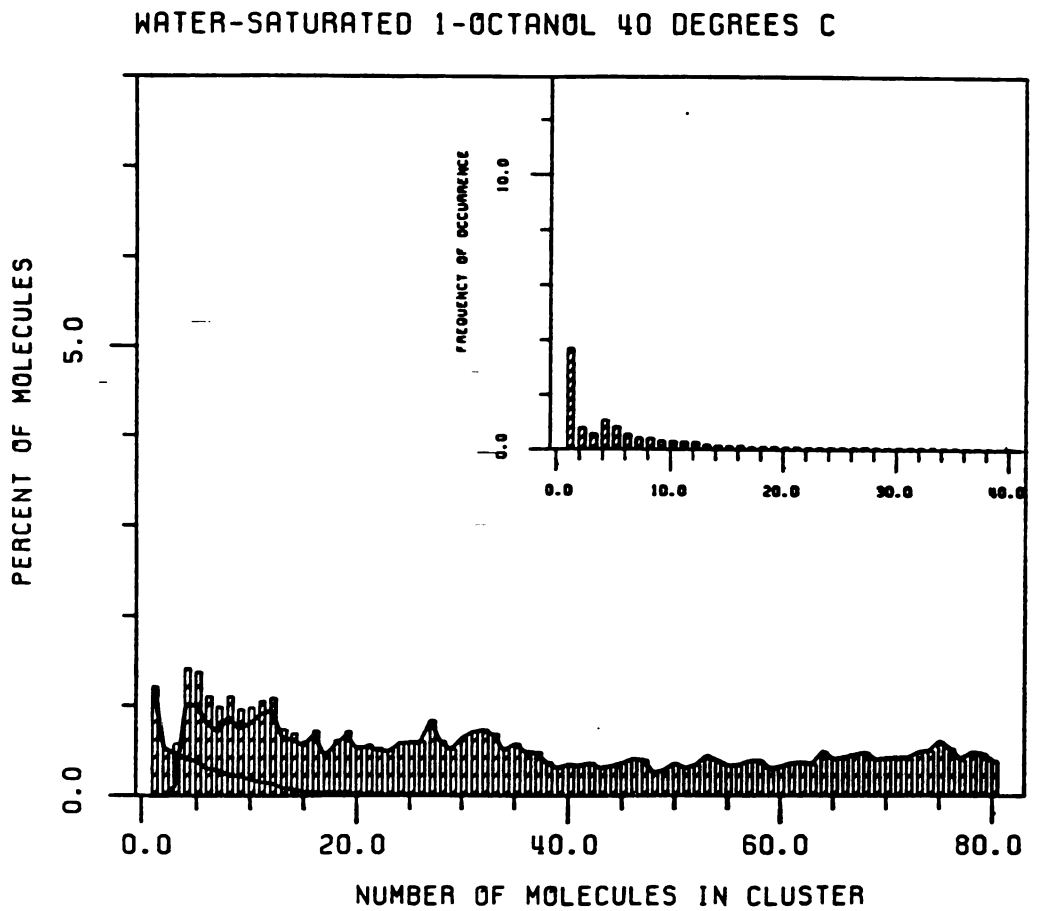


Figure 13. Frequency distribution of clusters of various sizes in hydrated 1-octanol at 40°C. The main display reports the percent of molecules which participate in a cluster of a given size. Data collected as in Table V. The inset reports the actual number of clusters of a given size found in a periodic box of 228 1-octanol molecules and 76 waters. Data collection as in Figure 10.

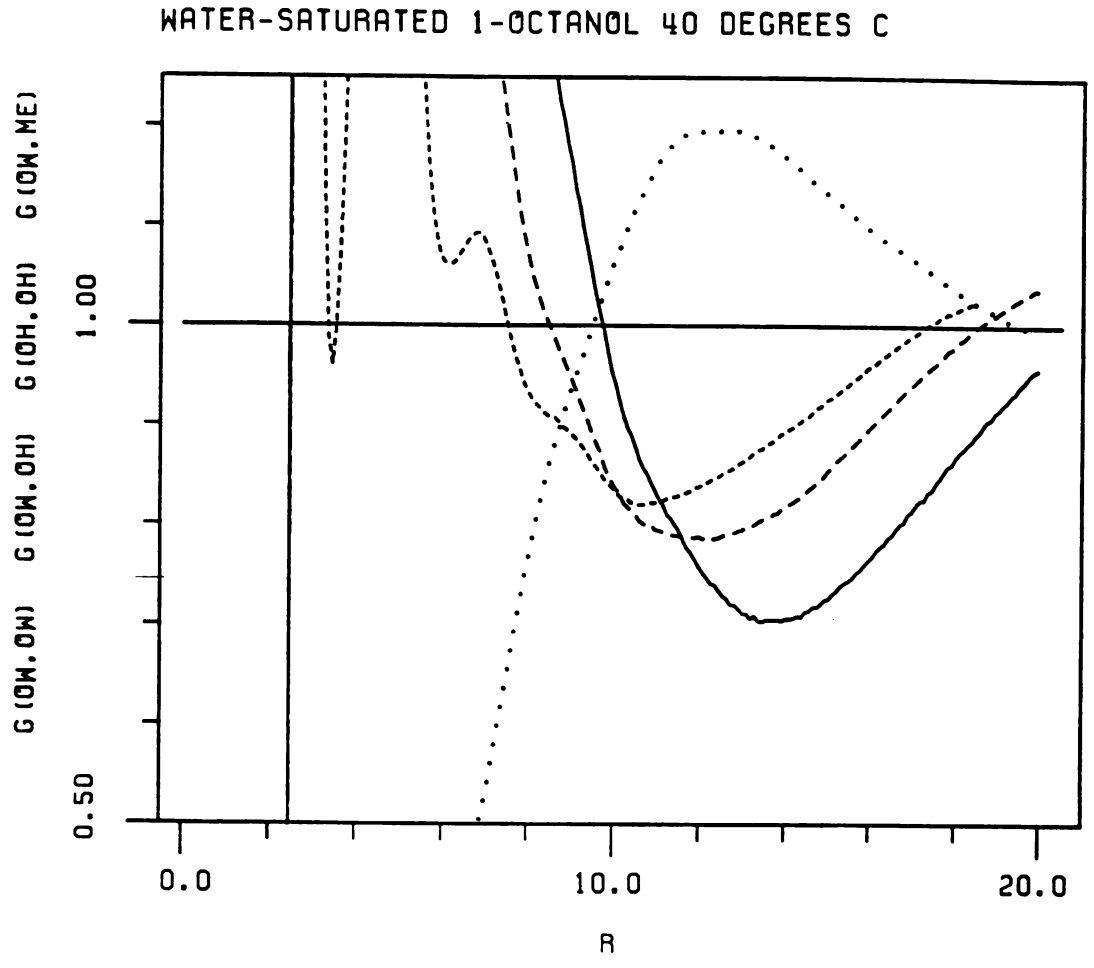


Figure 14. Closeup of radial pair distribution functions for water-saturated 1-octanol at 40°C. Plots are g(Ow,Ow) between water oxygens (solid line), g(Ow,Oh) between water oxygen and octanol oxygen (broken, long dashes), and g(Oh,Oh) between octanol oxygens (broken, short dashes). Dotted plot is g(Ow,Me). Data collected as in Figure 10.

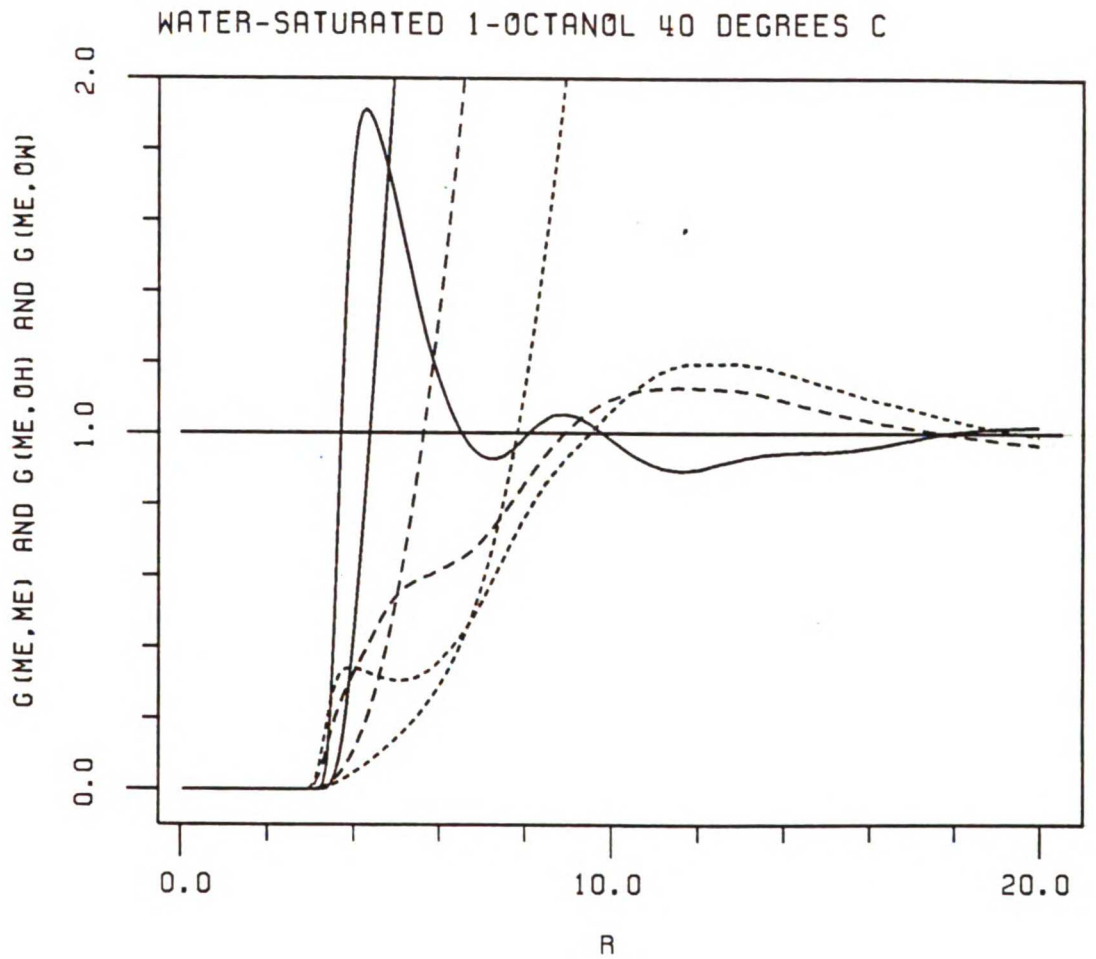
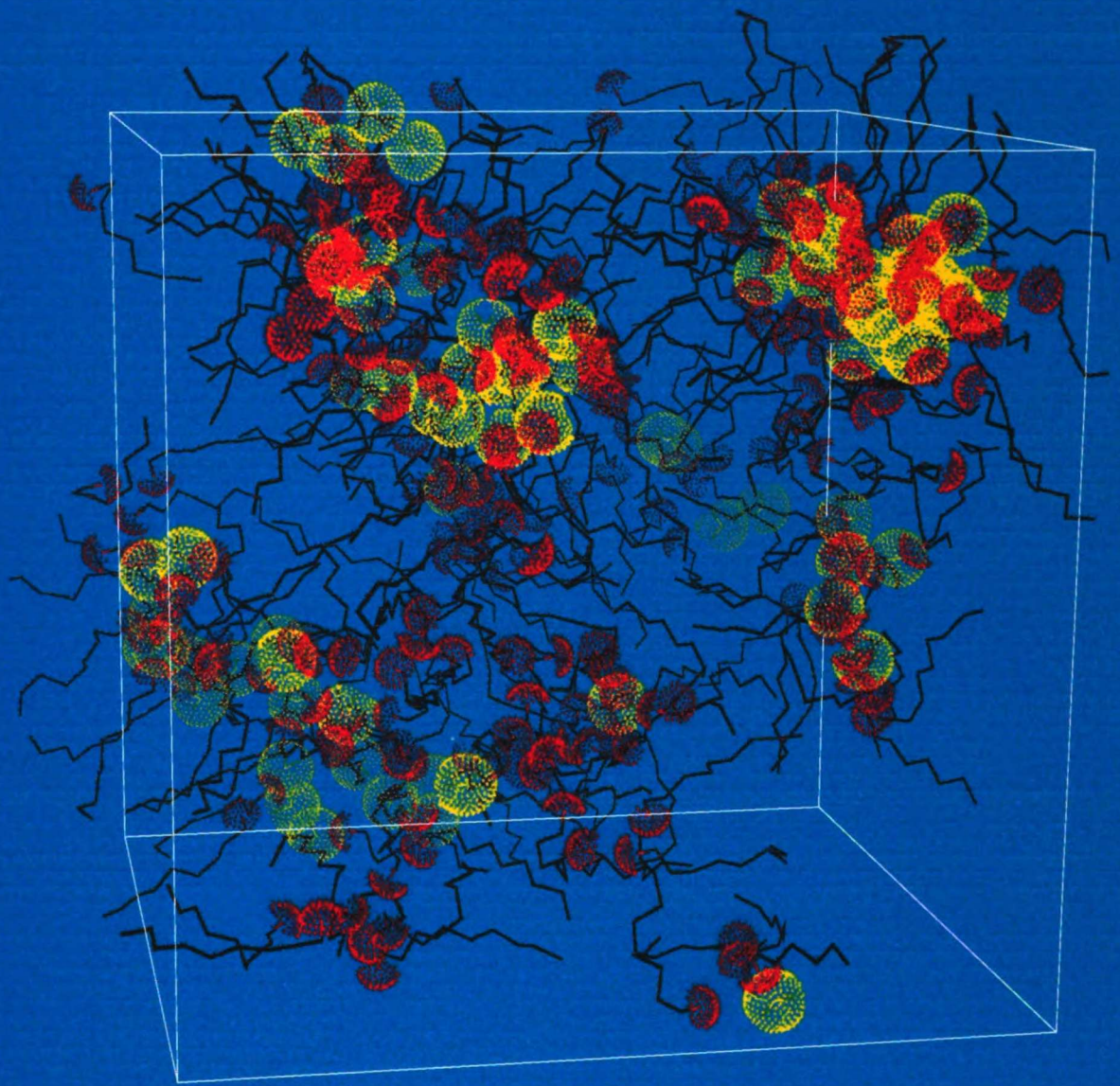
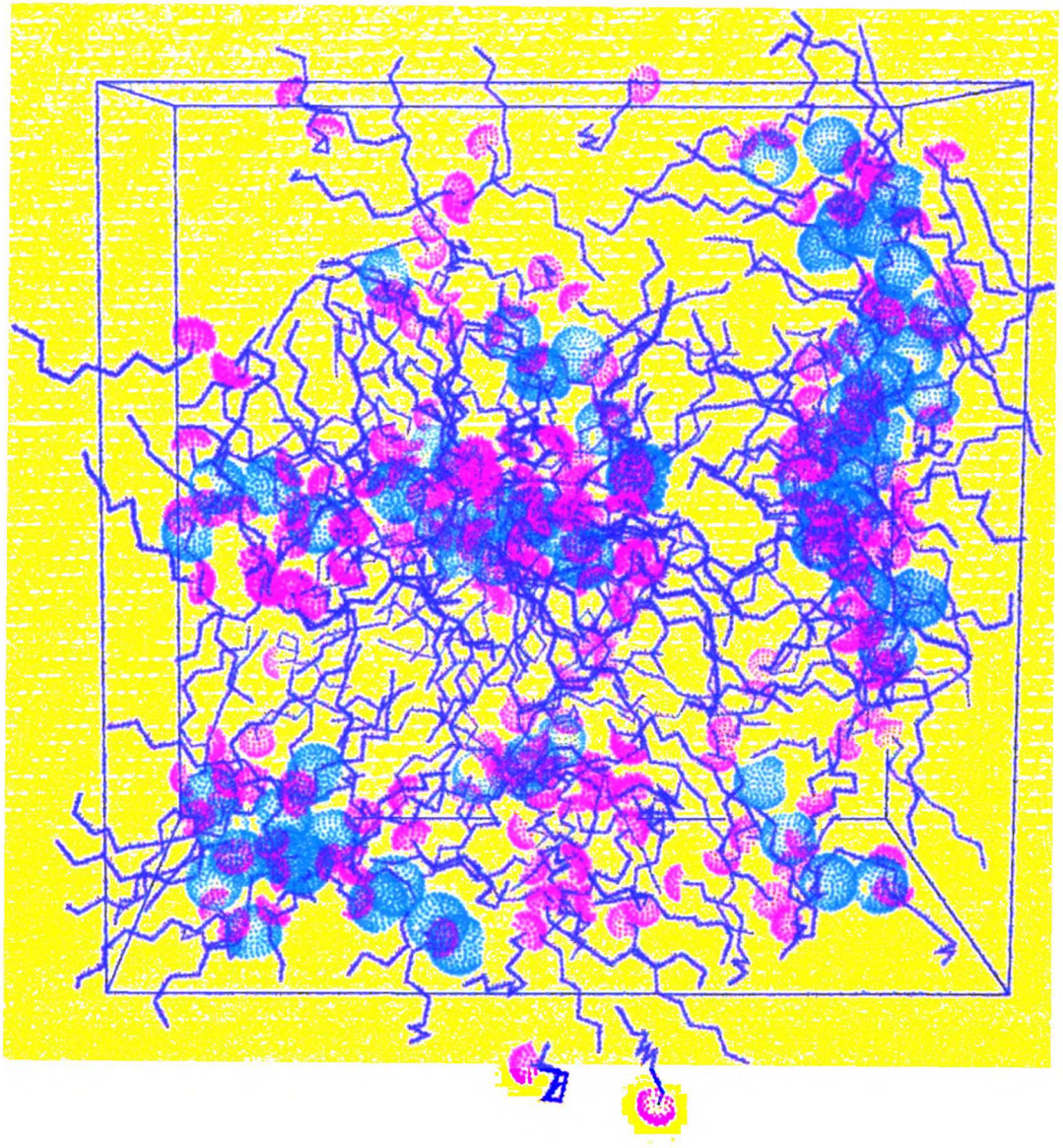


Figure 15. Radial pair distribution functions in hydrated 1-octanol. Plots are g(Me,Me) between end-methyls (solid line), g(Me,Oh) between end-methyls and hydroxyl oxygens (broken line, long-dashes), and g(Me.Ow) between end-methyls and water oxygens (short-dashes). Data collected as in Figure 10.



water-saturated 1-octanol UCSF MidasPlus



water columns in oct/wat UCSF MidasPlus

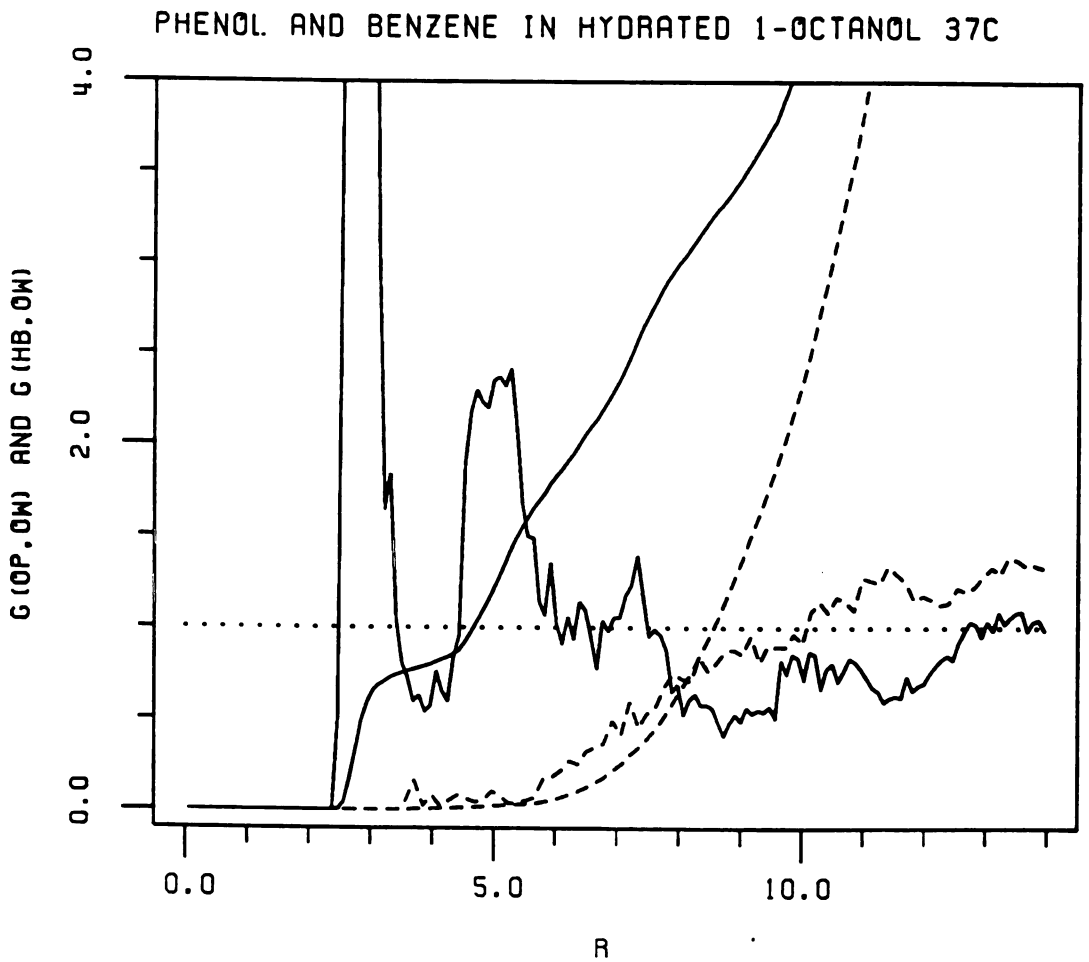


Figure 18. RDFs: Phenol/Benzene against water oxygen; g(Op,Ow) (solid line) and g(Hb,Ow) (broken line), and coordination numbers. Data collected over 2000 coordinate sets.

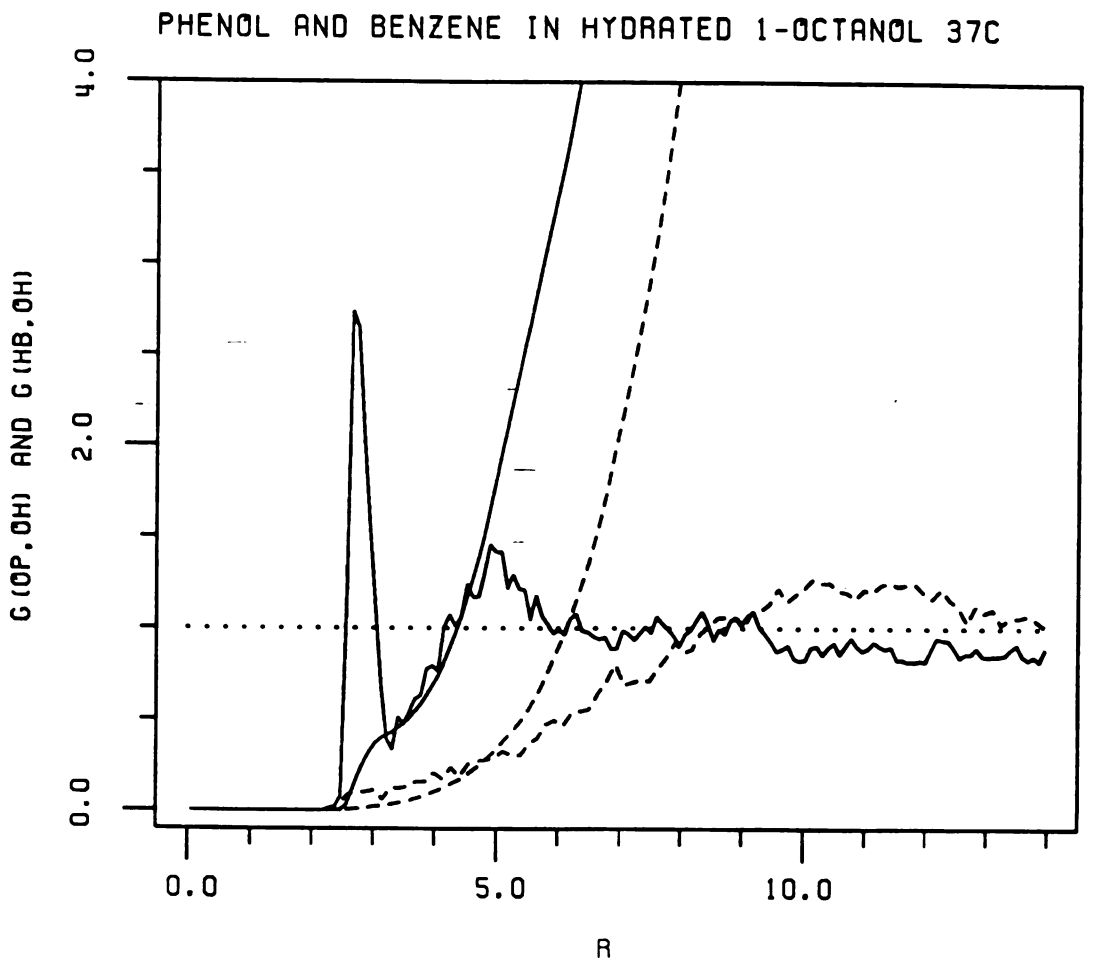


Figure 19. RDFs: Phenol/Benzene against octanol oxygen; g(Op,Oh) (solid line) and g(Hb,Oh) (broken line), and coordination numbers. Data as in Figure 17.

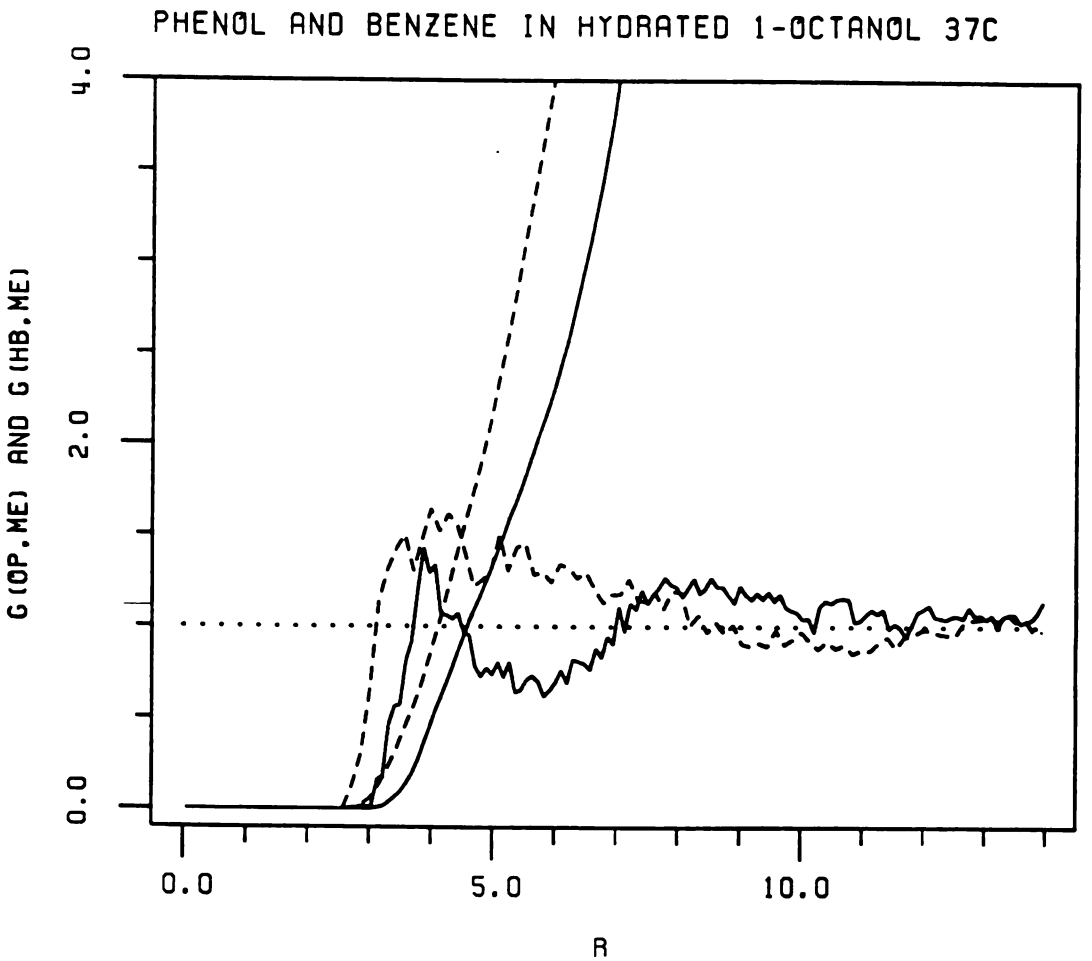
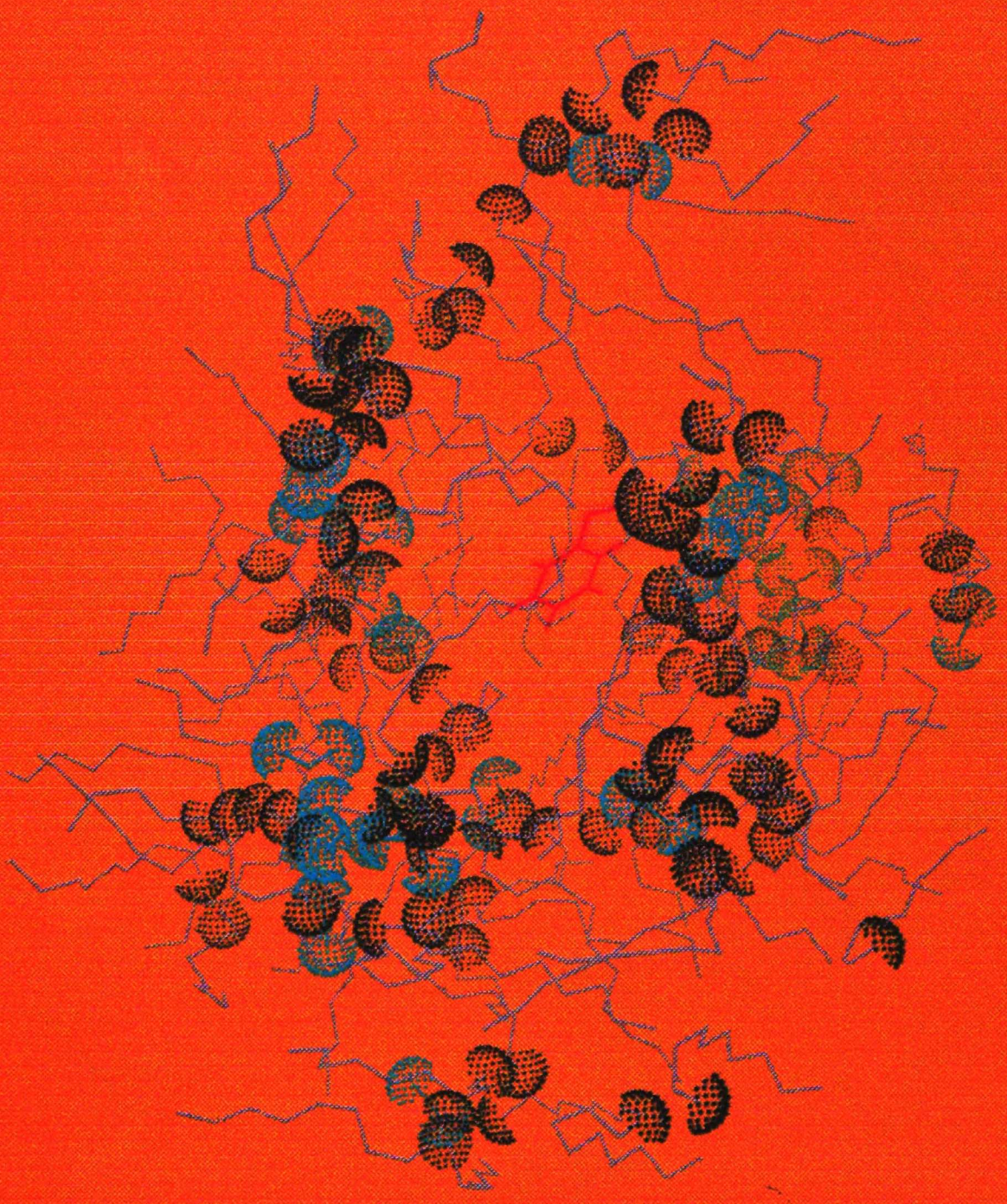
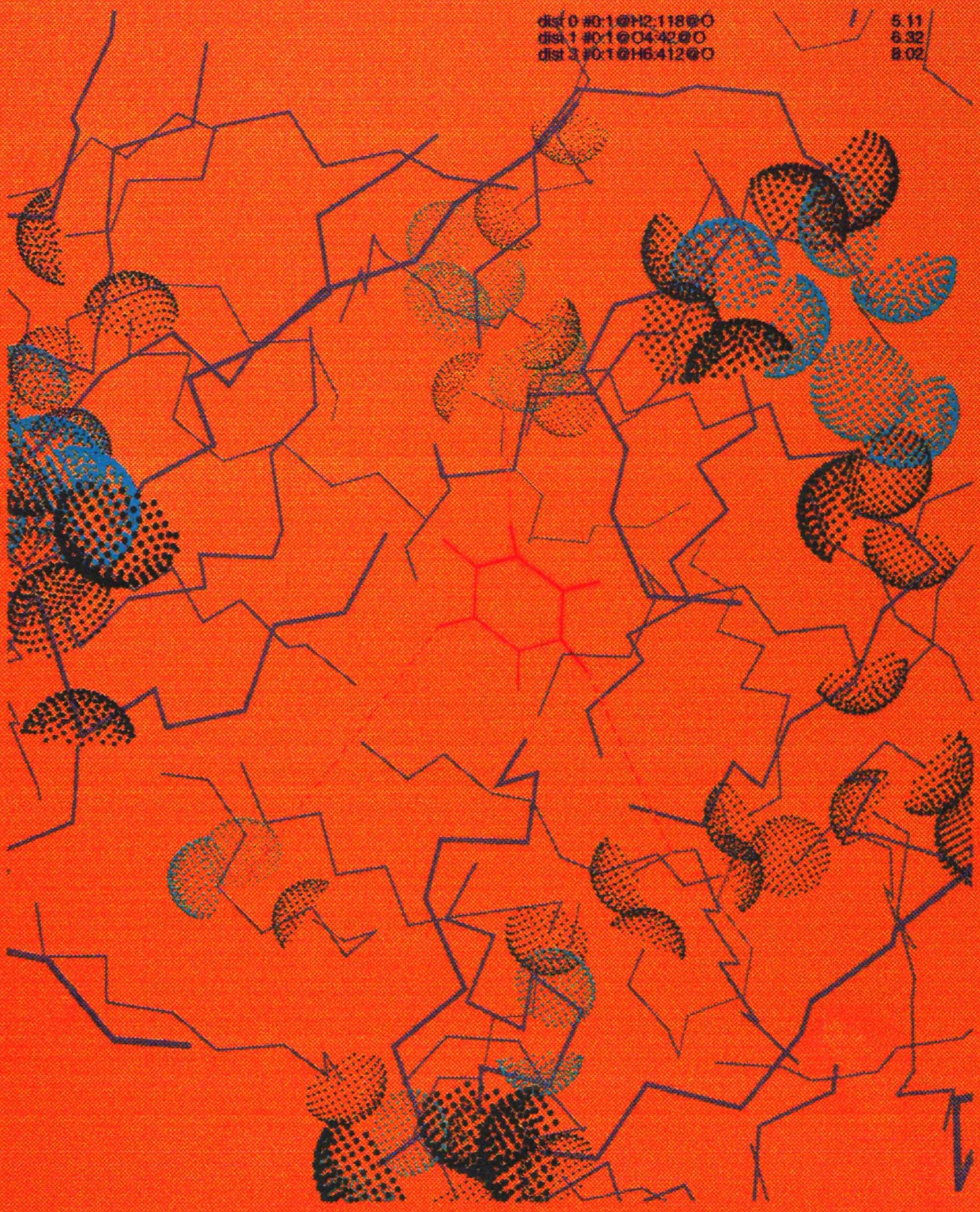


Figure 20. RDFs; Phenol/Benzene against octanol methyls; g(Op,Me) (solid line) and g(Hb,Me) (broken line), and coordination numbers. Data as in Figure 17.



phenol in hydrated 1-octanol UCSF MidasPlus



benzene in hydrated 1-octanol UCSF MidasPlus

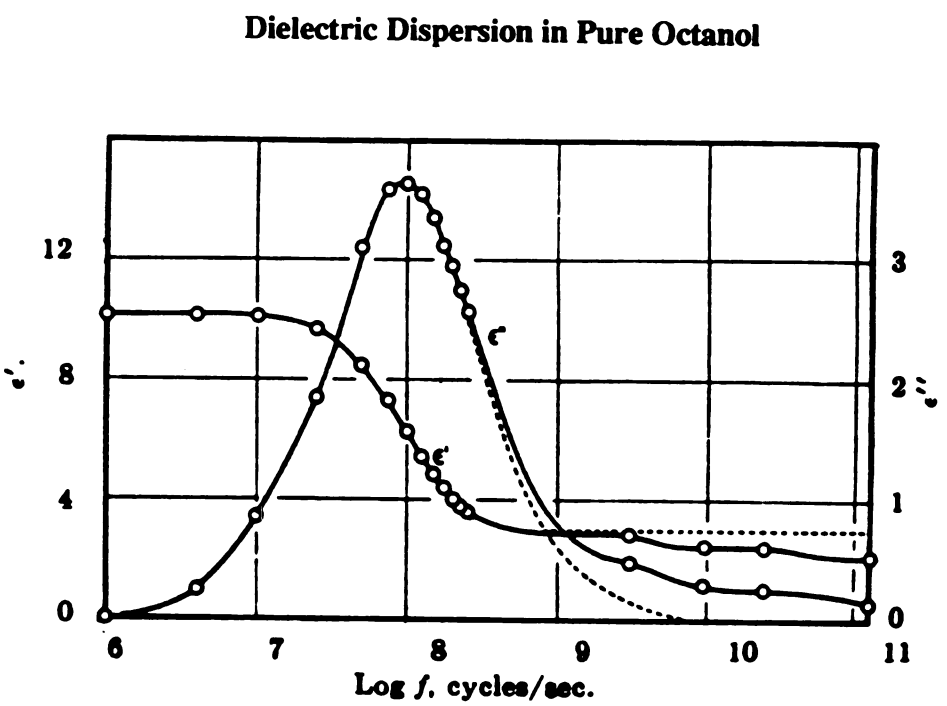


Figure 23. Dielectric constant and loss of *n*-octyl alcohol at 20° against log frequency (in cycles/sec.). Solid lines represent the observed data and dashed lines the calculated behavior for the first dispersion region.

Figure 23. Dielectric dispersion plot of ε′ versus log(f) and absorbance ε′′ versus log(f) where *f* is actually ω, the angular frequency in cycles per second. We thank authors S. K. Garg and C. P. Smyth and The American Chemical Society for allowing us to to reproduce this figure from reference 46.

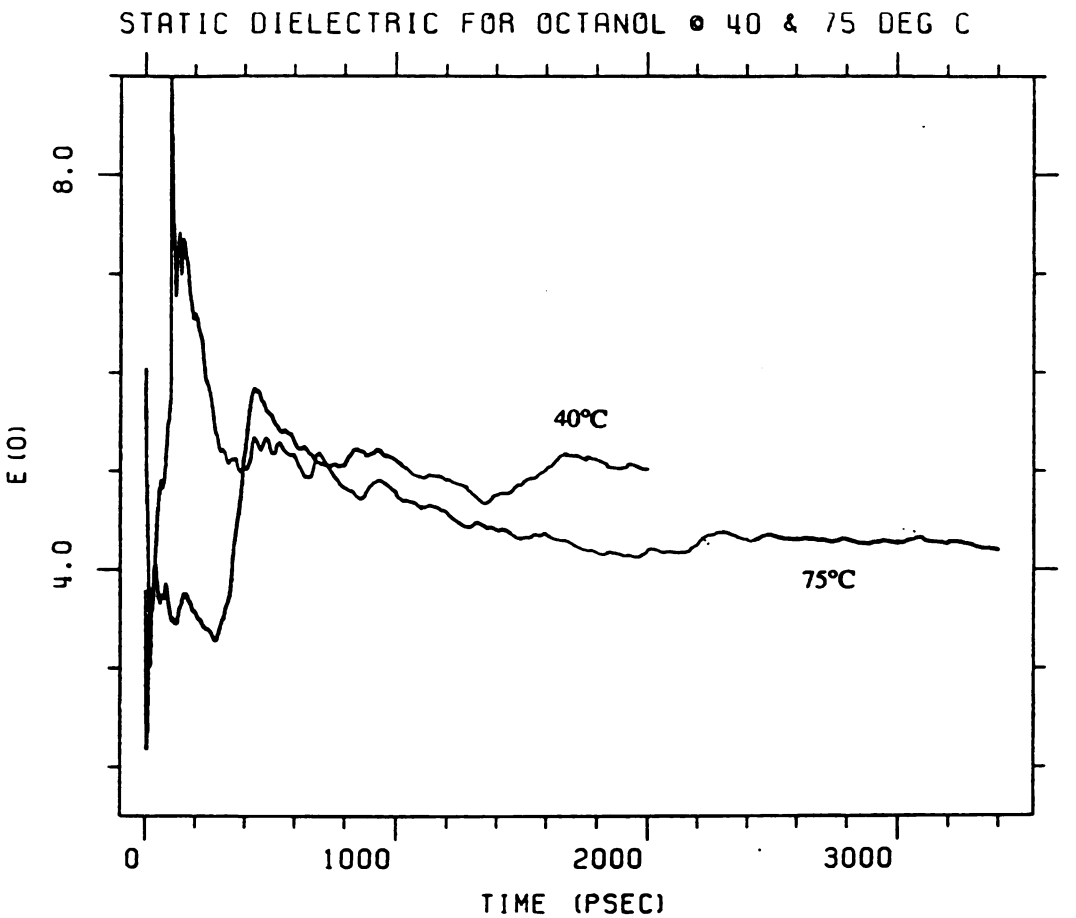


Figure 24. Plots of the cumulative static dielectric constant, ε_0 versus Time in picoseconds. The upper line is from 2.0 psec of pure 1-octanol MD at 40°C. The lower line is from 3.3 psec of pure 1-octanol MD at 75°C.

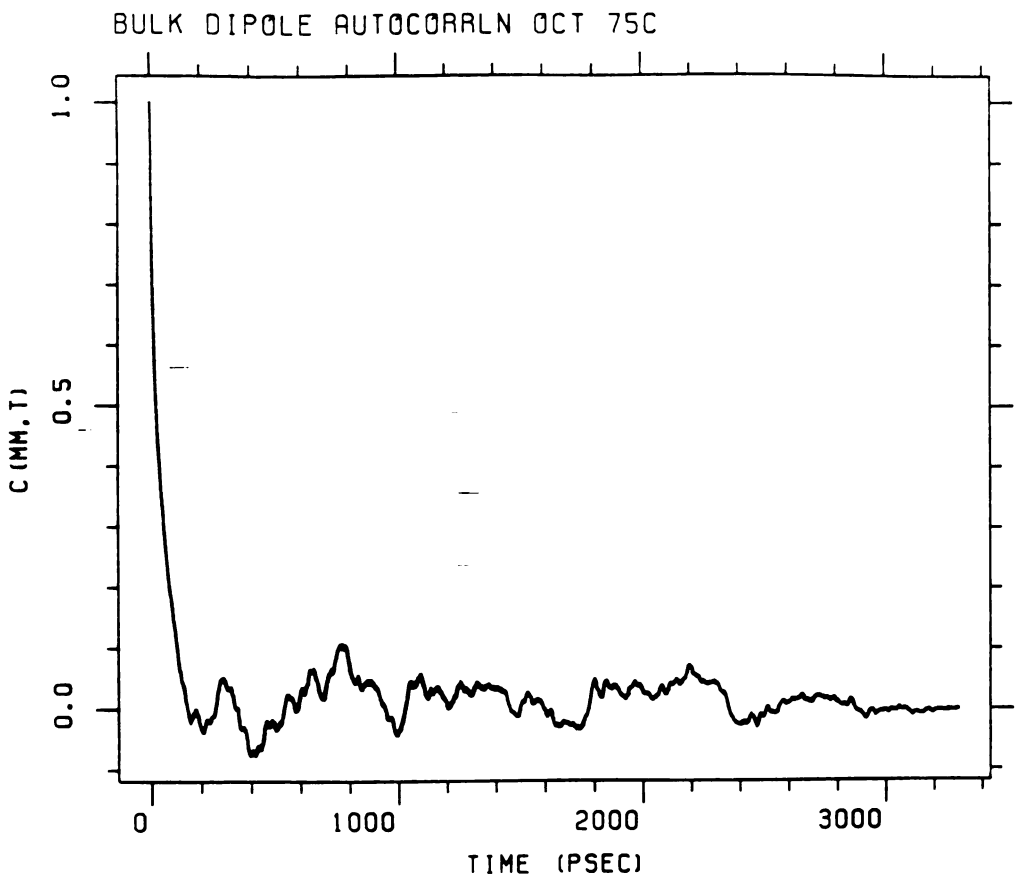


Figure 25. Plot of the autocorrelation of the bulk polarization moment, *C(***M**(t)*)*, ala equation 12, for pure 1-octanol at 75°C.

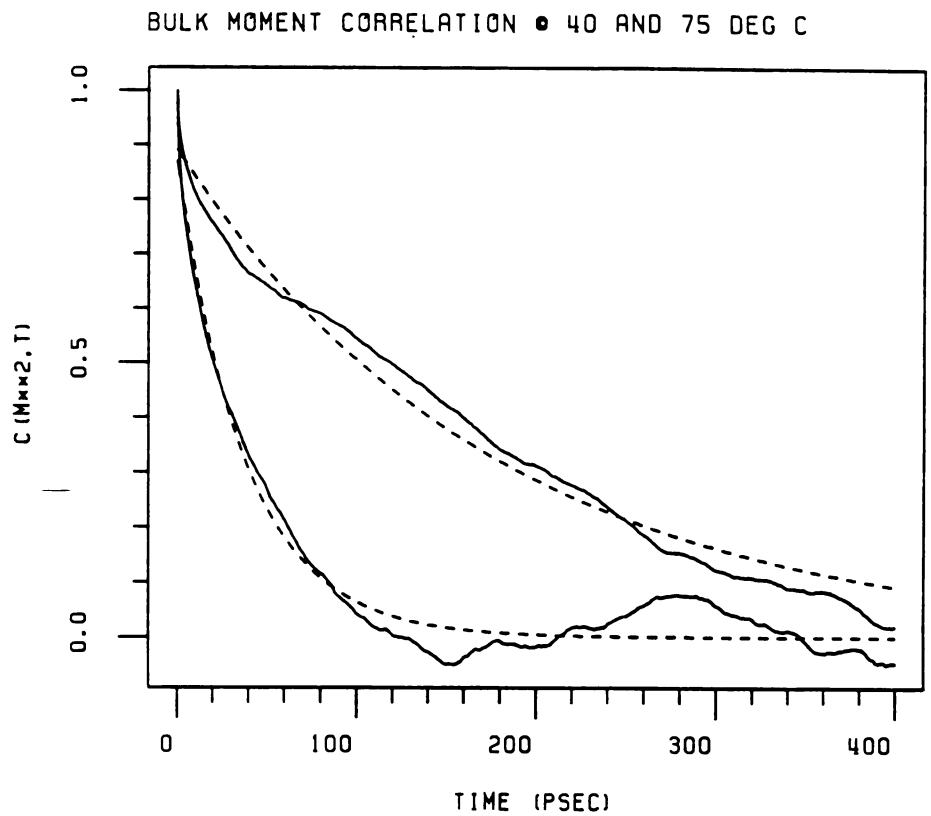


Figure 26. Plots of the autocorrelation of the bulk polarization moment, *C*(**M**(t)), ala equation 12, for pure 1-octanol at 40°C (upper solid line), and 75°C (lower solid), over 400 psec. The broken lines show the single-exponential fits to the raw data.

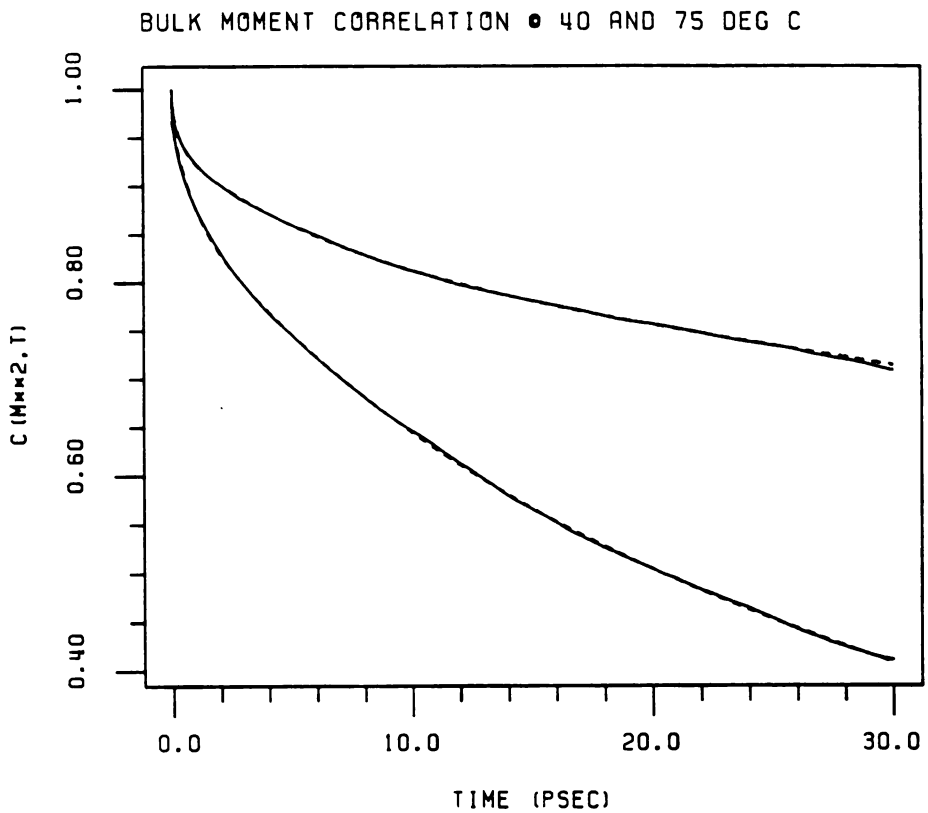


Figure 27. Plots of the autocorrelation of the bulk polarization moment, *C*(**M**(t)), ala equation 12, for pure 1-octanol at 40°C (upper solid line), and 75°C (lower solid), over 30 psec. The superimposed broken lines show the triple-exponential fits to the raw data.

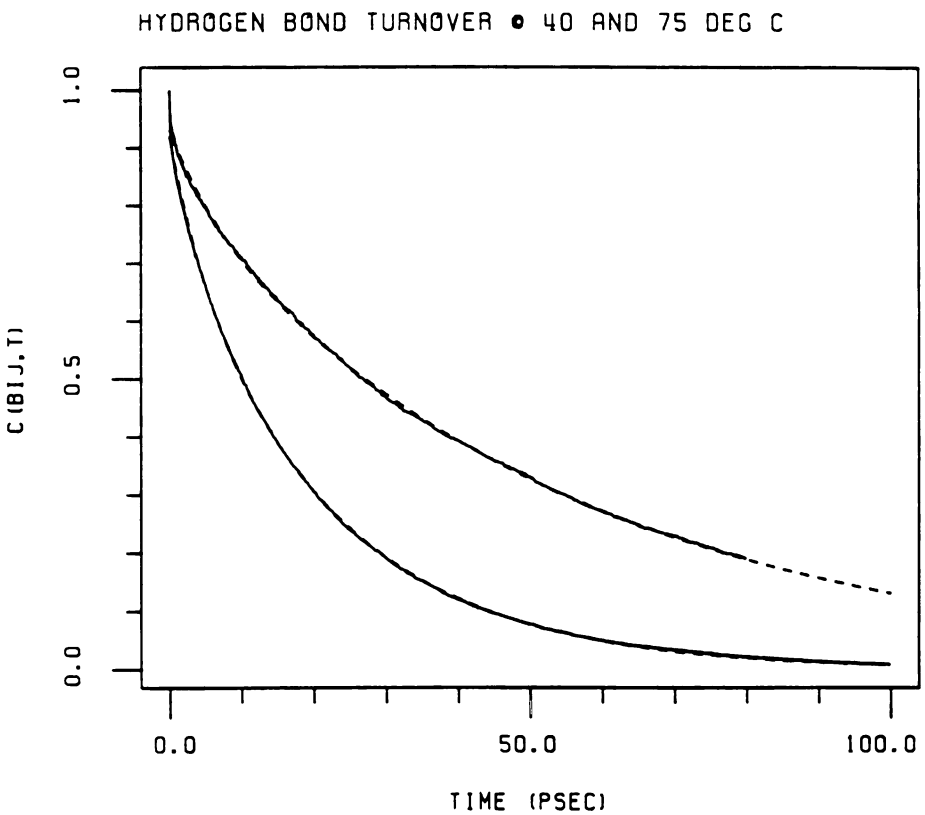


Figure 28. Plots of hydrogen-bond decay or "annihilation" in pure 1-octanol at 40°C (upper lines) and 75°C (lower lines). Double-exponential fits (broken lines) are superimposable over the raw data (solid lines).

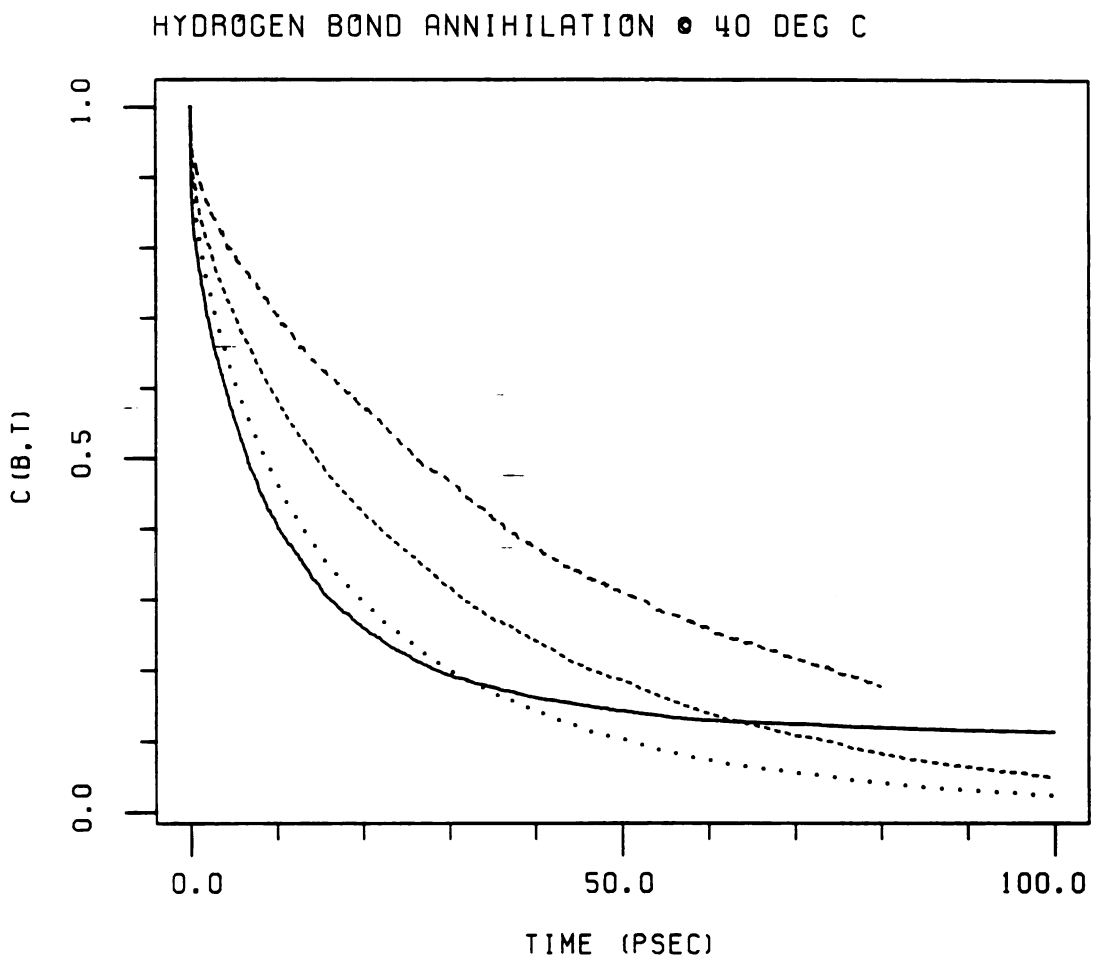


Figure 29. Plots of hydrogen-bond decay or "annihilation" in water-saturated 1-octanol at 40°C. The upper-most line (long dashes) for pure 1-octanol is shown for comparison. The second highest line (short dashes) if for octanol-octanol hydrogen bonds: the third highest line (dotted) is for octanol-water bonds; the solid line is for water-water hydrogen bonds.

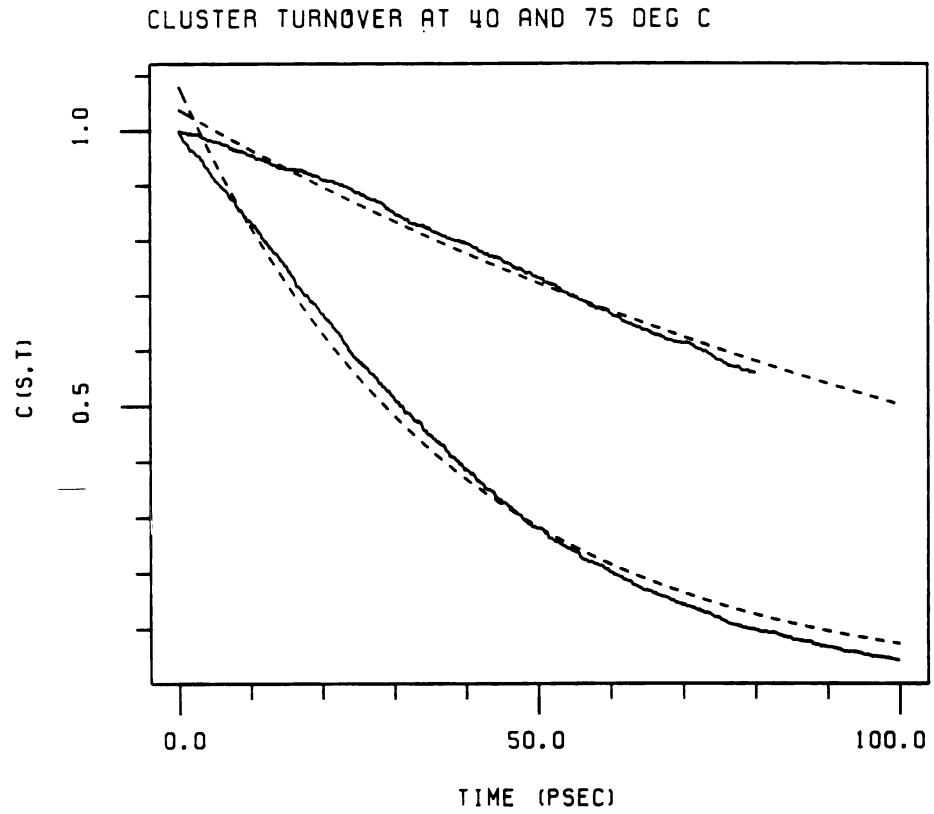


Figure 30. Plots of aggregate-cluster decay or "annihilation" in pure 1-octanol at 40°C (upper lines) and 75°C (lower lines). Single-exponential fits (broken lines) to the raw data (solid lines).

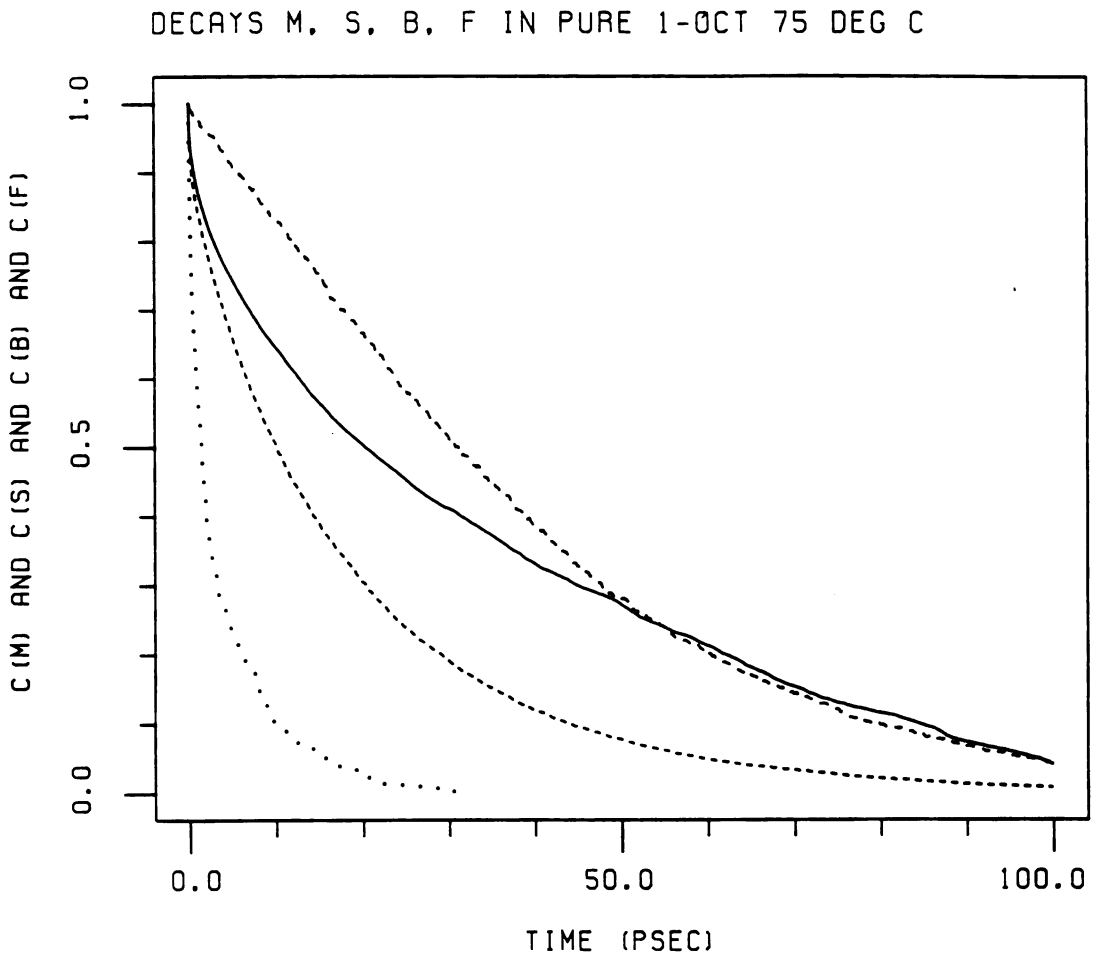


Figure 31. Plots of four different decays in pure 1-octanol at 75°C, (raw data). Time-autocorrelation of bulk moment fluctuations $\delta\mathbf{M}$ (solid line), aggregate-cluster decay $\mathbf{S}_{CL}$ (upper long dashes), hydrogen-bond decay $\mathbf{B}_{ij}$ (short dashes), and free monomer decay (dotted).

Figure 32. Squared fluctuations of the bulk moment in pure 1-octanol

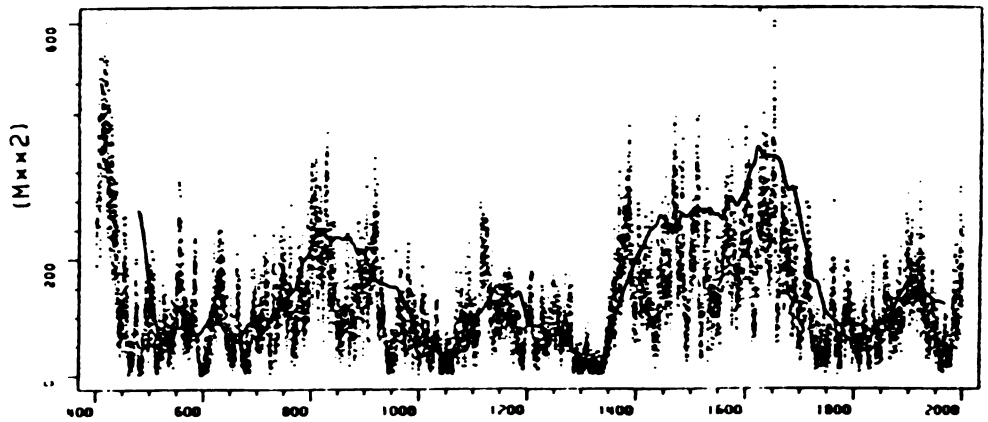


Figure 33. 6-membered ring frequencies versus M^2.

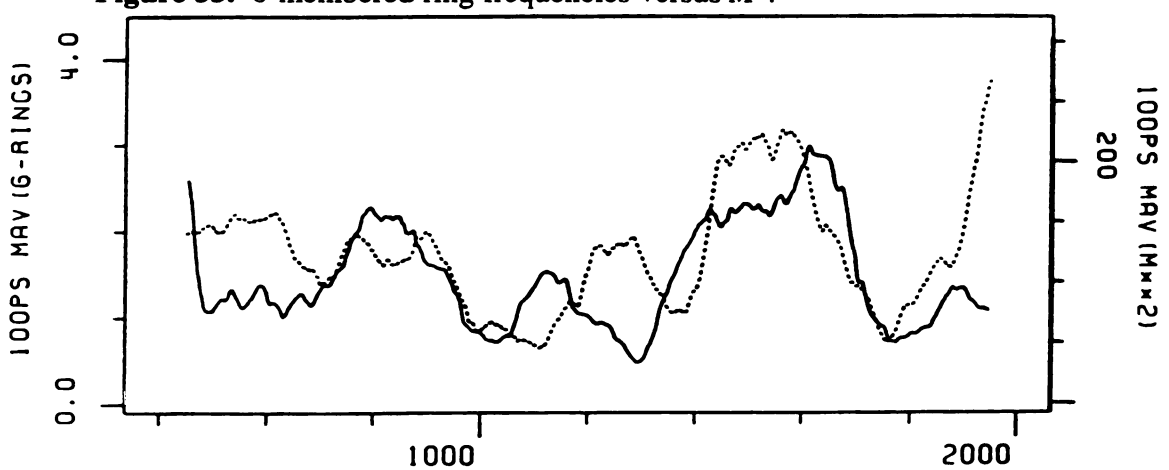
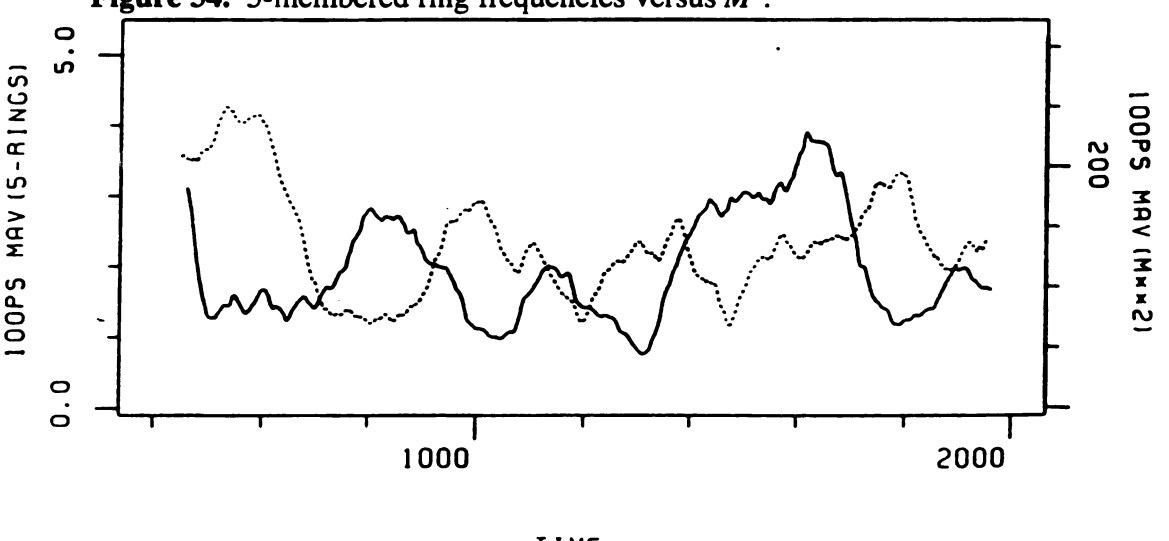


Figure 34. 5-membered ring frequencies versus M^2.



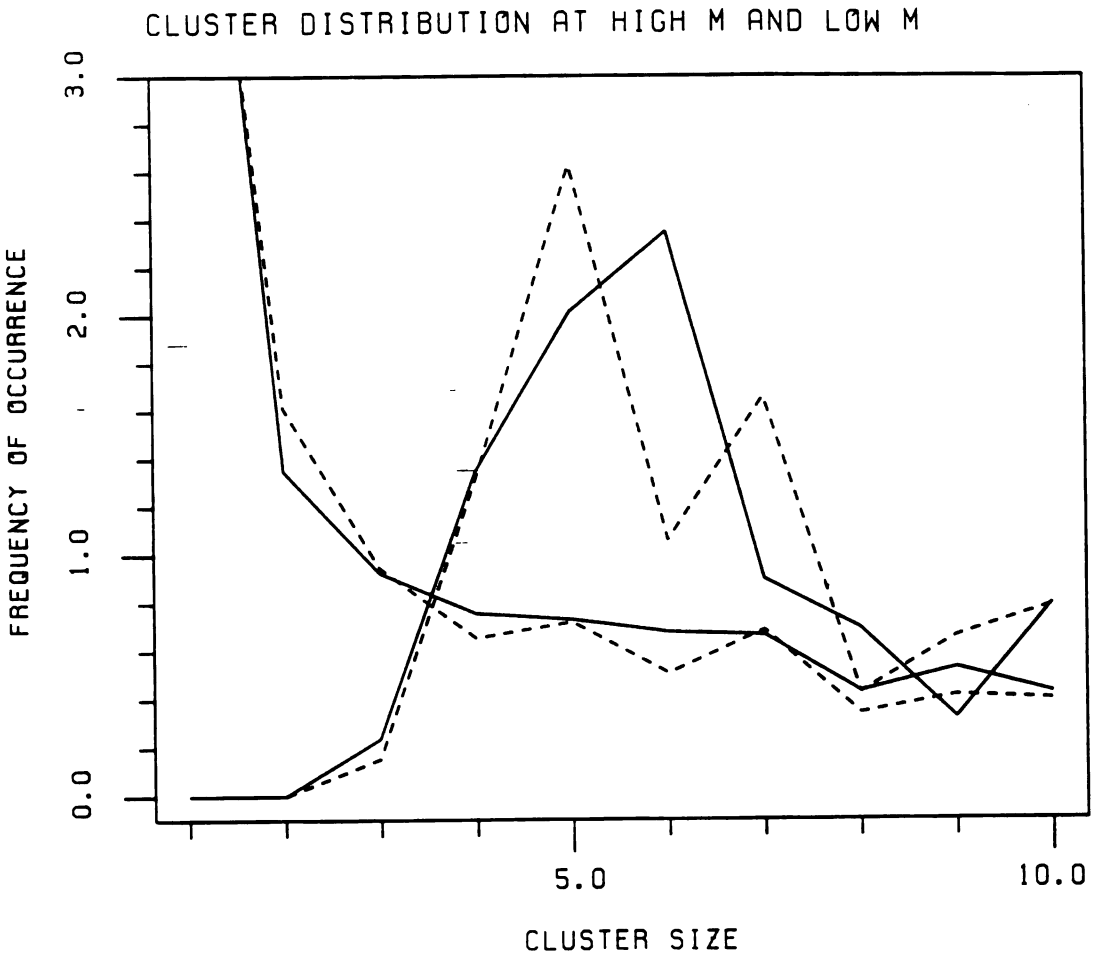


Figure 35. Cluster distribution at high values (solid lines) and low values (broken lines) of M^2. Y-axis is frequency of occurrence of clusters of the size indicated along the X-axis which were found in a box of 250 1-octanol molecules at 40°C.

CHAPTER III

AMBERCUBE MD, Parallelization Of AMBER's Molecular Dynamics Module For Distributed-Memory Hypercube Computers

The text of this dissertation chapter III is a reprint of a very similar article that appeared in the Journal of Computational Chemistry in 1993. The coauthor listed in the JCC article, Peter A. Kollman, directed and supervised the research which forms the basis of this chapter.

Permission to reprint has been granted from John Wiley & Sons, Inc.

Reprinted with permission from: S. E. DeBolt, P. A. Kollman, *Journal of Computational Chemistry*, **14**, 312-329 (1993). Copyright 1993 John Wiley & Sons, Inc.

AMBERCUBE MD, Parallelization Of AMBER's Molecular Dynamics Module For Distributed-Memory Hypercube Computers

ABSTRACT

A fully functional parallel version of the molecular dynamics (MD) module of AMBER3a [1] has been implemented. Procedures parallelized include the calculation of the long-range nonbonded Coulomb and Lennard-Jones interactions, generation of the pairlist, intramolecular bond, angle, dihedral, 1-4 nonbonded interaction terms, coordinate restraints, and the SHAKE bond constraint algorithm. So far as we can determine, this is the first published description where a distributed-memory MIMD parallel implementation of the SHAKE algorithm has been designed to treat not only *hydrogen-containing bonds* but also *all heavy-atom bonds*, and where "shaken" *crosslinks* are supported as well.[2]

We discuss the subtasking and partitioning of an MD time-step, load balancing the nonbonded evaluations, describe in *algorithmic detail* how parallelization of SHAKE was accomplished, and present *speedup, efficiency, and benchmarking results* which were achieved when this hypercube adaptation of the MD module AMBER was applied to several variant molecular systems. Results are presented for speedup and efficiency obtained on the nCUBE machine, using up to 128 processors, as well as benchmarks for performance comparisons with the CRAY YMP and FPS522 vector machines.

INTRODUCTION

Why parallelize MD? There is a need for faster computation to facilitate simulations of complex systems where sampling over **long continuous molecular dynamics trajectories** is desired. Long continuous trajectories are *essential* for simulations where long-time-scale motions are influential. Examples of studies where this is an important factor include the modeling of phenomena related to dielectric effects, diffusion, viscosity, heat capacity, interfacial behavior, enzyme active site dynamics, long-time structural reconfiguration as occurs in membranes, proteins, DNA, or other polymers, and whenever sufficient statistics must be generated to associate microscopic characteristics with time-dependent macroscopic traits of a system.

In MD, the time-step comprises the fundamental unit of divisible work; the program iterates over many time-steps, generating a discretized time trajectory during which the data of interest is accumulated. Because MD simulates particle motions as a function of time, it is by nature an inherently sequential algorithm. Because each MD time-step is dependent upon the result of the previous one, a time-step must be *"granularized"* or data-decomposed into time-sequential subtasks that can be distributed over multiple processors. Figure 1 is a schematic of the various subtasks into which an MD time-step can be usefully divided and indicates the portions executed in parallel in the implementation discussed here.

When converting a program from sequential to parallel, it is typical to strive for a scale-up in speed that is proportional to the number of processors applied to the job, i.e., linear scaleup. Obtaining greater speedup in a computation is then simply a matter of

applying more processors. The parallelized code decreases its processing-unit useage time to zero as the number of processors approaches infinity, leaving only the extant performance-limiting sequential code. Amdahl's Law, as applied to parallel computation[3a], expresses the reality that so long as there is *any* extant sequential code after parallelization it will limit the overall speedup that can be expected, regardless of the number of processors applied. If, after parallelization of some sequential code, the fraction of the original sequential code remaining is F, then the theoretical maximum achievable speedup factor is 1/F.

For instance, say the SHAKE routine is responsible for 3.6% of sequential execution time in some molecular system where bond length constraints are being used, and miscellaneous sequential code is responsible for another 3% of execution time. If the SHAKE routine were not parallelized, there would be a total of 6.6% extant sequential code after parallelization. The maximum possible overall speedup would be a factor of 1/0.066, or 15.1 times as fast as the sequential code, even if an infinite number of processors were applied. However, with the SHAKE routine parallelized, a speedup of 33.3 would be theoretically possible.

In this study it was found that, when considering MD simulation systems of typical size, the use of 64 or more processors sped up the parallelized routines to the point that internodal message-passing overhead and the remaining miscellaneous sequential code dominated the execution time. Although maximum theoretically achievable parallel speedup is rarely achieved in practice, one can envision a distributed-memory parallel MD that scales nearly linearly with the number of applied processors, even when

that number were greater than 64 or so. It would have to be either a very fine-grain implementation (as yet undescribed) with minimal message passing, or where interprocessor communications were lighteningly fast, and with next to no extant sequential code after parallelization. The cost of internodal communications and the specific consequences of Amdahl's limit for parallel molecular dynamics are demonstrated and discussed in the Performance section.

Several recent articles have addressed various issues regarding the parallelization of MD for a host of machine architectures.[2,4] Here, we present a discussion of MD data decomposition and a data-replication approach for the adaptation of sequential MD routines to parallel hypercubes and describe in algorithmic detail the load balancing of nonbonded interactions and parallelization of the SHAKE routine. One of the goals in this parallelization effort has been to retain the full general utility of the original sequential version. We demonstrate this with several molecular test systems that contain variant features: pure liquid TIP3P water, a binary liquid media octanol/water, fully solvated (crosslinked) crambin protein, and alphalytic protease protein with a covalently attached inhibitor ligand, coordinate restraints on a mobile α-helix, a "cap" of solvation waters over the active site, and where only a specified set (a *belly*) of atoms within a given radius of the active site are mobile.

The original AMBER3a MD fortran code[1] was retained wherever possible to limit reprogramming of data structures and evaluation routines, facilitate ongoing maintenance, and ensure backward compatibility. Consequently, this effort represents a conversion to medium-grain parallelism via the exploitation of implicit parallelism

rather than a complete redesign and reconceptualization of molecular dynamics down to the level of finest grain. Notably close agreement between the sequential and parallel output values is enjoyed as a convenient side benefit of this approach.

The procedures that were the prime targets of parallelization were the evaluation of pair-wise forces and energies for nonbonded interactions, pairlist generation, SHAKE (bond length constraints), dihedral angles, 1-4 nonbonded terms, angles, and bonds; any coordinate restraints are also handled in parallel. This includes all of the terms evaluated in AMBER's classical molecular mechanical potential energy function [5]:

$$V_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] +$$
$$\sum_{nonbonded} [\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}] + \sum_{Hbonds} [\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}] +$$
$$\sum_{1-4's} [\frac{1}{VDW_{scale}} (\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}) + \frac{1}{EEL_{scale}} (\frac{q_i q_j}{\varepsilon R_{ij}})]$$

The form and parameterization of this potential function has been discussed in detail elsewhere.[6] Support routines facilitating procedures related to data decomposition, load balancing, and the optimization of internodal communications during global regathering operations were added.

In brief, the various data structures are domain-decomposed over N_P processors, where each processor is responsible for a portion of the molecular system's force or coordinate array. The force array is globally regathered once per MD step via optimized interprocessor communication, and energies are globally summed once per MD

step after the calculation of all pairwise forces (refer to Figure 1). If the parallel version of the SHAKE routine (SHAKEP) is being used, then the array of "corrected" coordinates array also undergoes regathering once per each MD step via an interprocessor communication routine that was customized specifically for this purpose.

We present a section on domain decomposition in which the natures of the various MD domain types are described. There is discussion of how partitioning and assignment of domain fragments over N_P processors was accomplished. Following that, a description of the techniques used in load balancing the nonbonded interactions will be given along with an illustrative example.

A significant portion of this paper is devoted to a discussion of the important issue of parallelizing the SHAKE algorithm. Several methods are proposed, implemented, and their results discussed. The problem of shaking "long-distance" crosslinked bonds is solved. An internodal reciprocal exchange algorithm that facilitates heavy-atom to heavy-atom bond constraints is described.

DOMAIN DECOMPOSITION

To describe how the various data structures were decomposed to facilitate parallelization, some discussion of the several different domain classes characteristic of molecular dynamics systems is necessary. Molecular dynamics data structures belonging to the domain classes homogeneous-static (short-range intramolecular interactions), inhomogeneous-dynamic (nonbonded interactions), and inhomogeneous-static

(SHAKE), require distinctly different treatments.[3b] In this distributed-memory adaptation of the originally sequential AMBER MD code, each processor holds a full replica of the important data arrays, but is responsible for processing only a given segment within a data array.

Homogeneous-Static Domain

The short-range intramolecular interactions are all members of the most easily decomposable of domain types, the homogeneous-static domain. Members of a homogeneous-static domain are equivalent in their characteristics, require identical amounts of computation time, and are unchanging in number and in distribution of members (thus a static domain) throughout the execution of the program. These domains' members are structured as simple divisible vector arrays. Lists consisting of atoms appropriately associated to describe bonds, angles, dihedrals, 1-4 interactions, and coordinate restraints are decomposed simply by dividing their members equally among the available processors.

Using this breakdown scheme, the members of each fragment list remain in sequential order. This way, evaluation routines which do list processing using numerically ascending loop indices can be retained in their same form as found in the original sequential program. The only difference is that new beginning and ending indices are sent to the evaluation routines, and these routines now process only a reduced, "chopped out", portion of the total data vector. It should be clear that concurrent processing is "automatically" load balanced for these short-range interactions of the

homogeneous-static domain.

The most time-consuming *intra*molecular interactions to evaluate are the dihedral and 1-4 nonbonded interactions, usually contributing 2-10% of the execution time for sequential code. These can begin to significantly dominate parallel execution time, especially once pairlist generation, nonbonded interactions, and SHAKE have been parallelized. Accordingly, the dihedral and 1-4 interaction evaluations were domain distributed as described above for homogeneous-static domains. All of the other intraatomic interactions mentioned above were also parallelized even though they require minimal time to evaluate; these procedures parallelize easily, and would pose *added* computation during global regathering of the forces if they were left as sequential code.

Inhomogeneous-Dynamic Domain

In this implementation, both pairlist generation and nonbonded interaction evaluations are executed in parallel. The nonbonded interactions, represented by Coulomb and Lennard-Jones (12-6 or 12-10) potentials, exemplify an inhomogeneous-dynamic domain. The members of this domain are the interacting residue pairs with their corresponding components, the atom pairs. Before the nonbonded interaction for each atom pair can be evaluated, an atom pairlist must be generated. AMBER's molecular dynamics module generates a residue-based pairlist. This means that all residues are tested against all other residues (globally, all I=1,N residues are tested against all J=I,N residues) to determine whether there are any active atom pairs between two given residue pairs. If any one of a residue-pair's atom pairs test as active, then all atom pairs

encompassed by the residue pair are considered active, and are stored in the pairlist which serves to mark the *active atom pairs* until the next pairlist regeneration. In a molecular dynamics calculation, generation of the pairlist is usually the second most overall time-demanding procedure, after nonbonded term evaluation.

Ordinarily, only the atoms *j* within a specified cutoff radius from any atom *i* are considered as actively interacting with atom *i*, forming *i-j* atom pairs. As the atoms in a dynamic molecular system undergo changes in their Cartesian positions with time, the number and identity of atom pairs made by any one atom *i* will change as a simulation progresses. Upon pairlist regeneration this may cause an imbalance in the number of atom pairs per processor. Dynamic load rebalancing during program execution may be necessary. The object is to reassign the appropriate beginning and ending residue-pair indices to each processor so that, within this assigned range, an equivalent number of active atom pairs will be encompassed by every processor. The solution to the problem of load balancing the nonbonded interactions of this inhomogeneous-dynamic domain will be discussed more thoroughly and illustrated by example in a later section of this article.

Inhomogeneous-Static Domain

In the decomposition of the bondlist(s) for the SHAKEP routine (where bond lengths are constrained to their equilibrium values in parallel), the domain members are the bonds with their component atoms. Individual bonds may differ in several attributes: their equilibrium lengths, their component atom types (and thus atomic mass),

and the connective topology of their surrounding environment. But, because a molecular system's topological connectivity never changes during an MD simulation, this inhomogeneous-static domain requires only a single "smart decomposition" at program initiation. To parallelize the SHAKE procedure, the bondlist can be broken into approximately equivalently sized fragment lists, with each node possessing one piece of the list and operating as a loosely cojoint member of a "parallel pipeline".

LOAD BALANCING OF NONBONDED INTERACTIONS

Molecular dynamics simulations employ interatomic force calculations in order to integrate Newton's equations of motion.[8] Atom positions and velocities are repeatedly updated subject to this field of interatomic forces, leading to the evolution of time-dependent step-wise particle trajectories. With each new time-step, a set of component forces and energies between nonbonded interacting atoms are evaluated by referring to the (periodically updated) stored list of interacting atom pairs. This nonbonded evaluation procedure is by far the most computationally intensive portion of the code, ordinarily accounting for 75-95% of execution time. To effectively speedup the calculation, it is critical to achieve a well-balanced distribution of these nonbond evaluations over multiple processors.

The task of pairlist generation is also distributed among multiple processors in this parallel implementation. The initial pairlist must be generated by assuming domain homogeneity (the assumption that the number of atom pairs is directly proportional to the number of residue pairs, all residues are of exactly the same size and have the same

conformation, and the system is homogeneously configured), providing a first guess for the proper decomposition. If, after pairlist generation, the number of nonbonded atom pairs assigned to each processor for evaluation is not within a specified percentage (we chose 10%) of a given target value (the global number of pairs divided by the number of processors used), then load balancing ensues. The number of atom-pair interactions stored or "owned" by each processor must be examined; then, exchanges are made between those processors that have surplus pairs and those processors which are deficient in pairs.

Also, to take best advantage of AMBER's many routines and data structures it is necessary to keep the interacting atom pairs belonging to a processor in a sequentially ordered list. This preserves the ordered correspondences between the pairlist and other sequential arrays, e.g., the nonbonded pair-exclusion list, sequential loop indices, etc. Because it is convenient to keep the nonbonded atom pairs in a sequentially ordered list, the beginning and ending residue-pair indices used as brackets in subtasking the generation of the pairlist must also be arranged in a correspondent ordering. Thus, exchanges of active atom pairs can occur only "across" or "through" the node borders of processors that hold adjacent residue-pair ranges. Let us illustrate with a practical example: load balancing the nonbonded interactions in the cap-solvated alphalytic-protease/inhibitor system over a two-dimensional hypercube, four nodes. The schema presented in tables Ia, Ib, and Ic present numerical results corresponding to the following logic:

A) Assign the initial-guess residue-pair indices to each processor, assuming domain homogeneity, as explained above (Table Ia).

Before load balancing, use the initial-guess residue-pair indices to generate the *actual* number of interacting atom pairs associated with each node. (Generate the pairlist).

B) Determine which atom pairs need to be exchanged between processors in order to obtain a balanced distribution. The final number of atom pairs held in each processor should nearly equal the target value (Table Ib). Sequentially correlate the residue-pair indices to the atom pairlist. Reapportion and reassign the residue pairs by sending and receiving internodal exchanges of residue-pair indices.

C) Pairlist regeneration over these residue-pair indices will now produce the required (corresponding) number of atom pairs (Table Ic). Regenerate the atom pairlist and continue the MD computation.

Note that prior to load balancing (Table Ia), as great as 78% deviation from the target value occurs, but after load balancing (Table Ic), there is nearly a perfect distribution of interacting atom pairs among the four processors.

The parallel load balancing overhead is folded into the time required to generate the pairlist. This extra overhead cost is largely offset by the great reduction in the time needed to evaluate the equiproportioned nonbonded interactions in parallel. Also, remember that the load balancer is dynamic; rebalancing will *only* take place when a molecular system reconfigures to the extent that the number of non-bonded interactions evaluated by any one processor exceeds the specified percentage criterion, 10% of the target value.

In very homogenous systems such as pure liquid water, load balancing may never be required. Load balancing might occur fairly frequently, say every few pairlist generations, for a very inhomogeneous and configurationally evolving system such as a binary solvent media with ionophore and ions. In systems such as proteins, which usu-

ally do not fluctuate widely in their tertiary and secondary structures over the time of an MD simulation, one initial load balancing, with perhaps infrequent rebalancing, should be sufficient. Also, when a very large number of processors are used, small fluctuations in the (low) number of atom pairs assigned per processor may give large percentage differences (greater than 10%). Such a situation could lead to more frequent load balancing.

PARALLEL SHAKE

SHAKE is a procedure for constraining bonds to their equilibrium bond lengths. When used, the SHAKE procedure is called every MD timestep after evaluation of the pairwise forces/energies in order to readjust or "correct" bond lengths before the new particle velocities are determined. This enforcement of rigid bond lengths effectively filters out the high-frequency motions of bonds which vibrate with a period nearly equal to the typical MD step size. With these high-frequency motions removed, a larger MD step size can be taken, thus extending the time spanned by a simulation of a given number of steps. As originally implemented, SHAKE is an iterative and inherently sequential procedure.[7] Bond atom positions are incrementally adjusted along a vector toward their equilibrium value. Adjustments are processed in numerical order, from the first bond to the last, all hydrogen-containing bonds adjusted first, then (if desired) all exclusively heavy-atom bonds are adjusted. Each bond-atom's position adjustment is influenced by its (equilibrium) position at the previous MD step, its current force-determined position, the magnitude of the deviation of its bond from the equilibrium

value, and its atomic mass. Several sequential iterations through the bondlist are required to reach convergence over all "shaken" bonds.

Recently, accounts describing an analytical matrix inversion method for the SHAKE algorithm, where no iterative process is required, have appeared.[2a] However, this procedure is useful only in cases where there are minimal coupling effects between adjusted bonds, i.e., when adjustment of the atoms in one bond do not affect the length of another bond. This essentially limits the analytic version to the cases where only hydrogen-containing bonds are "shaken".

Arguments as to the effect of the SHAKE approximation on realistic particle trajectories or as to the necessity or wisdom of its useage in various circumstances are outside the scope of this paper. Often enough, a user specifies that *all* bonds in a system should be "shaken" during a simulation. It is clearly prudent to retain this option in the hypercube parallel version of AMBER MD, particularly in view of Amdahl's limit for concurrent processing. This requires the creation of a procedure that decomposes the bondlist for parallel processing, modifications to the SHAKE algorithm so it will handle internodal communications, and a post-SHAKE internodal communication routine which optimizes the global regathering of the distributed "shaken" coordinates.

Connective Topology Breakdown

There are two major obstacles to the parallelization of SHAKE. First, it is inherently sequential as originally implemented. Also, because adjacent bonds always share an atom, the adjustment of a bond's length by repositioning its component atoms

causes position adjustment dependencies, or constraint couplings, between this bond and any adjacent bonds in the list.[2a,7] As the bondlist is processed, the repositioning of each atom is dependent to some degree upon any prior repositioning of atoms in preceeding bonds (which has occurred during the current SHAKE iteration) or in succeeding bonds (which occurred during the previous SHAKE iteration). Dependencies therefore extend all throughout the bondlist. The final converged (corrected) coordinates are thus partly a function of the (sequential) order in which bonds are adjusted. Obviously, this "order dependence" condition does not facilitate parallelism.

Secondly, the decomposition of the bondlist is complicated by the existence of a large variety of possible connective topologies. There are four bonds possible to any one atom, initiation and termination sites, side chains, cyclic loops, multiple cyclic loops possibly terminating at a common atom, multiple molecules, solvent or not, long-distance crosslinks between residues in a molecule, crosslinks between different molecules as in a covalent enzyme-substrate complex, and so forth. Depending on the molecular system under study, the various topological possibilities may be encountered in any combination.

So, multiple topological criteria must be met simultaneously to designate a break in the bondlist and still ensure a robust SHAKEP routine. For instance, when "shaking" all bonds no bondlist breaks should occur within sidechains or within loops, or in crosslink chains, or at a main-chain atom connected to any of a side chain, loop, or crosslink, whose bonds have not yet been adjusted in the current iteration, or when any of the numerically precedent atoms connected to a particular main-chain atom have not

yet been adjusted, etc. Often, the "breaks" occur between main-chain atoms which link two residues. In the end, the bondlist must be broken into *N* sublists for *N* processors.

Initial Naive "Averaging" Approach

Once decomposition of the connective topology has been accomplished, how is the sequential SHAKE algorithm transformed for parallel operation? Figure 2a illustrates a hypothetical system where all bonds of a long polymer chain are being "shaken" and there are shared atoms at node borders. Figure 2b shows a blow-up of the node borders, where a nitrogen atom is shared between two processors. Why not simply distribute the bonding atom sublists to the individual processors, and after every SHAKE iteration have those processors sharing a node-border atom take an average of the shared atom's newly adjusted coordinates, then continue processing? Because when coordinate averages are taken, this "average middle ground" is maintained, and the shared atom quickly follows a vector path back toward the last MD step's (equilibrium) position, and then becomes virtually immobile. In order to meet all of the mutual bond length constraints, the rest of the molecular system must fall in line behind the node-border atoms, which are stuck at (nearly) their last MD step positions. The net result is that all atoms shared between nodes get "pinned" and the trajectory can no longer evolve. This eventually results in the SHAKE routine bailing out due to the inability to meet its convergence criteria. Fortunately, there is a way to circumvent this problem.

Parallel Pipeline

Our strategy will involve the construction of a "parallel pipeline", where the mechanism for internodal exchanges of atom positions at the node-border regions or "pipe joints" is kept as simple as possible. AMBER designates a main chain through the connective topology. It is important to seek decompositions where a processor shares one main-chain atom with another processor which contains a bond that is directly adjacent topologically and follows numerically along the mainchain. This way, communication between cojoint pipe-segment nodes facilitates the exchange of information relating to a shared cojoiner atom (at the node border) and complicated communication trees are avoided.

Table II exemplifies a typical topological decomposition for a system (crambin/water) in which all bonds will be shaken. The solvent waters (assigned to processors P4-P31 in this example) are noncovalently associated and small enough so that no molecular fragmentation is required; this part of the system is conveniently homogeneous. But, the bondlist for the inhomogeneous polymeric part (the protein in processors P0 through P4) will be fragmented into a number of segments by breaking the topology between certain main-chain bonds. Notice that processors P1 and P2 share atom 291; the "tail-atom" of P1's bondlist connects to the "head atom" of P2's bondlist, as illustrated in figure 2b. When "shaking" *all* bonds, this part of the system can be treated by using the parallel methods described next.

Two-Way Relay Shake-pipe

The following is a discussion of a parallel approach which exactly reproduces the results of the sequential SHAKE algorithm. The two-way or bidirectional relay pipe described here is an interesting method for parallelizing inherently sequential code. In the end, this method was the progenitor of an even better solution to the parallelization of SHAKE. It will serve as instructive to consider the advantages and limitations of this approach prior to discussing the SHAKEP algorithm that was finally derived from it (and uses the same data structures).

The parallel relay pipe operates much the same way as a pipeline mechanism is used to expedite data processing in vector machines. The differences lay in the necessity for communication of information between nodes, and the bidirectional data flow through the pipe. Each node works on its "own" fragment of the bondlist, ignorant of any coordinate adjustments made in other nodes. As illustrated in Figure 3, during the first SHAKEP iteration only the node (P0) containing the first segment of the bondlist operates, processing from the beginning (Figure 3, at 1A) to the end (1B) of its sublist. The adjusted coordinates of the cojoiner atom (at the terminal end of P0's sublist) are "handed off" via internodal communication to P0's cojoint pipe-segment partner (P1), the processor containing the next segment of the bondlist. Thus, a relay along the pipe-segment nodes has begun.

Upon receipt of pipe-segment P0's adjusted cojoiner atom coordinates, the succeeding cojoint pipe-segment node P1 then adjusts the appropriate adjacent bond(s) in its own sublist and thereafter returns the cojoiner atom's updated coordinates back to

the preceding cojoint partner node P0. Thus, a *two-way*, forward and reverse, relay is occurring. Node P1 then continues processing up to the end of its bond sublist and a two-way handoff also takes place at the next cojoint pipe-segment node (P2). This process continues along the nodes until the global end of the bondlist is reached. At that point, one full SHAKEP iteration has occurred, and the resulting first-iteration coordinates are exactly the same as they would be if the entire procedure had been done sequentially, as in the original SHAKE.

But, also notice that once pipe-segment node P0 has obtained the updated terminal cojoiner atom's coordinates from P1 (Figure 3, at 2A), it can complete the next SHAKEP iteration over its sublist. So, while node P1 is processing through the first SHAKEP iteration, node P0 is processing through the second iteration. When P2 is processing the first iteration, P1 will be processing the second, and P0 the third, iterations (Figure 3, at 3A), and so on. In this way, the bondlist segments are processed in parallel, very analogously to the way a pipeline is used in vector processing.

For our purposes, there are four drawbacks to this algorithm. First, there is overhead in loading and unloading the pipe. During the initial and final stages of the procedure, some processors are either waiting to start or waiting for the other processors to finish, i.e., there are times when not all nodes are processing in unison. Second, due to the complexities in decomposing inhomogeneous molecular connective topologies as discussed earlier, each node may not contain exactly the same number and types of bonds on which to operate. This may cause a delay whenever a processor must wait for its cojoint pipe-segment partner(s) to handoff updated cojoiner atom coordinates.

These "delay bubbles" can promulgate through the pipe with successive cojoiner coordinate handoffs. Third, it should be evident that this procedure, although operating in parallel, may not necessarily scale as the number of processors applied to the job. It will scale as *the order of the number of SHAKEP iterations* needed to reach global convergence *if* the number of iterations is less than or equal to the number of processors used. The number of iterations needed to reach convergence is system and protocol dependent but typically ranges between 4 to 30. So, practially speaking, the two-way relay shakepipe will not always scale with the number of processors used and may also incur significant "waiting time" overhead. Last, there is no good way to support the shaking of bonds which are crosslinked across long-distances, e.g., in proteins where internal disulfide bonds are present (bond *i—j* in figure 2a). In this case, pipe-segment node P0 has a crosslink to an atom in pipe-segment node P2. In the two-way relay scheme, P0 would have to wait until P2 sends back updated crosslink cojoiner-atom coordinates before it could begin the next iteration. There would be little advantage to a parallel pipeline in such a situation.

The bidirectional relay pipe, was implemented as we have described here, was found to significantly speed up the SHAKE procedure when a low number of concurrent processors was used and gave coordinate results which precisely matched those of the sequential version of the algorithm. But, as expected, it did not scale well with the number of processors when many concurrent nodes were used. This bidirectional relay pipe approach could be very useful to obtain parallelism in an algorithmically similar (inherently sequential) situation *where many iterations occur*, and/or when a result which exactly *reproduces a sequential version* is required.

Reciprical Exchange For Crosslink Support

We have devised a mechanism by which long-distance crosslinked bonds can be managed: the reciprical exchange (REX) algorithm. Referring to Figure 2a, notice that one cross-link-bond atom (*i*) is held in processor P0's bondlist, the other atom (*j*) in processor P2's bondlist. During any REX SHAKEP iteration, the crosslink atoms are treated as usual; processors P0 and P2 reposition them in an effort to adjust their parent bond's (bonds *h*—*i* and *j*—*k* in Figure 2a) lengths toward the appropriate equilibrium values. The list of crosslink bonds (only bond *i*—*j* in this case) is held in another processor, say, processor P3, which treats them just like any other bond; it is P3 that actually adjusts the crosslink bond's length. But, at the end of the SHAKEP iteration, a reciprical exchange occurs. Processors P0 and P3 each trade their independently adjusted crosslink atom *i* coordinates and processors P2 and P3 trade their independently adjusted atom *j* coordinates, after which processing continues.

While the involved processors do not contain the exact same coordinate values after a trade occurs, the REX method is facilitated by the fact that these values follow each other very closely during each iteration. Figure 4 illustrates REX crosslinks iterating to convergence. It is as if each processor involved in a crosslink atom exchange were challenged by its trading-partner processor to deal with the shared atom from the opposing processor's point of view (coordinates), and carry on from there. Crosslink atoms are thus free to migrate during REX SHAKEP; they do not get pinned as they would if a naive averaging method were used. After this interprocessor haggling has continued for several iterations, a bargain is reached (convergence) when the atom posi-

tion is agreed upon to within the user-specified SHAKE tolerance criterion.

By exchanging crosslink atoms' adjusted coordinates after each SHAKEP iteration, each cross-connected molecular segment contributes an influence over the final corrected positions. The parts of the molecular system on either side of a crosslink bond exert competitive leverage on the repositioning of the crosslink atoms during successive REX SHAKEP iterations. Appropriately, the effects of large-scale chain motions on either side of the crosslink are carried through the "conduit" crosslink bond, influencing the evolution of the polymer's overall MD trajectory.

This REX method might be used to solve the problem of unsupported crosslinked bonds in the bidirectional relay shake-pipe. But, instead, the REX idea was broadened; we took advantage of its potential to also eliminate other drawbacks which were encountered with the relay shake-pipe. This finally lead us to install REX for both crosslink and cojoiner atom exchanges in what turned out to be a far more efficient parallel SHAKE algorithm, the REX *nodechain*.

Reciprical Exchange Nodechain

After successfully applying the REX mechanism to deal with crosslinked bonds in parallel mode, it was further applied to enable the management of internodal communications at the cojoint pipe-segment node junctions. There are several "variations on a theme" for the REX strategy, which could be implemented to facilitate cojoiner atom coordinate interchanges at the pipe-segment node junctions. The best REX scheme is the one which generates the fewest iteration cycles in SHAKEP yet follows a "clean"

coordinate exchange and adjustment strategem.

We start by dispensing with the common practice of processing along the bondlist in strict numerical order as an artifact of sequential machine programming. Nearly any ordering of bond adjustments during a SHAKE iteration should be equally legitimate as long as the system-wide constraint couplings are preserved, the original SHAKE repositioning method is adhered to, and the criteria for convergence are met in the end. Because of this, A SHAKEP iteration can begin by letting all pipe-segment nodes initiate processing of their bond sublists in unison. Because all processors will always operate simultaneously, each processing nearly the same amount of work, one entire "global bondlist update" can be completely processed once every iteration, in a fraction of time 1/(number of processors) it takes the sequential algorithm. No pipe loading (nor unloading) overhead is incurred. The procedure should thus scale more nearly as the number of processors used instead of as the number of SHAKEP iterations to reach convergence.

Notice that this application of the REX approach is no longer analogous to pipeline processing. The bond adjustments are not processed sequentially "through" a multisegment pipeline as occurs with the bidirectional relay pipe. Nothing, aside from the mutual repositioning dependencies (constraint couplings of the atoms), "flows" through a pipe. Here, the structure is more analogous to a chain constructed of relatively independent linked units, where the adjacent nodes sharing border atoms form a nodechain.

These "loosely linked" REX nodes are involved in handoffs only with their directly adjacent interlinked nodechain partners (and any crosslinked nodes). And, because REX handoffs of colinker (cojoiner) atom coordinates do not progress along the linked nodes, handoff-waiting delay bubbles should not readily promulgate through the REX nodechain. In the final analysis, though, some minimal waiting delays (due to domain inhomogeneity) will probably occur regardless of the specific REX protocol used, although improvement over the bidirectional relay shake-pipe is certainly expected.

The implementation of a REX nodechain is somewhat complicated due to several factors: the presence of the two bondlists, one for the hydrogen-containing bonds and one for heavy-atom only bonds, the diversity of possible connective environments for the chainlinker atoms, the support of crosslinked bonds, and the various iteration-stop and exit conditions. But, the approach is conceptually simple. The decomposition of the bondlist is exactly the same as for the bidirectional relay shake-pipe, where the bondlist was divided into N_P sublists, each handled by one pipe-segment node, with cojoiner atoms shared between cojoint pipe-segment node partners.

Figure 5 illustrates the internodal communication protocol for an implementation of a 4-processor REX nodechain with crosslink support. All processors begin processing in unison. After chainlink node P_i adjusts all bonds that affect the position of its "head of the sublist" chainlinker atom the newly adjusted coordinates are sent over to the interlinked node-partner P_{i-1} (actually, this data is temporarily held in P_{i-1}'s message-read buffer). Both chainlink nodes P_{i-1} and P_i continue processing. When

node P_{i-1} nears the end of its sublist, just before adjusting its terminal chainlinker atom it will read the P_i-adjusted chainlinker atom coordinates sent over from P_i earlier. Processor P_{i-1} adjusts this terminal chainlinker atom, then "reciprocates" by returning these adjusted coordinates forward over to P_i's message-read buffer. Any crosslink atoms are handled at this point, as described previously in the REX crosslink section. At the initiation of a new SHAKEP iteration, processor P_i reads and uses the coordinates which were sent forward from P_{i-1}. SHAKEP then continues to iterate, repeating this protocol until convergence is reached over all bond lengths. In Figure 5, crosslink communications between P_3—P_0 and P_3—P_2 are illustrated (with bold broken lines) at the termination of iteration 2B and initiation of iteration 3A.

The REX nodechain differs from the pure reciprical exchange mechanism as implemented for crosslinks. A simultaneous mutual "trade" of coordinates does not occur when the nodechain is processed. During each SHAKEP iteration, the REX *crosslink atoms* are independently adjusted by their parent pipe-segment nodes and immediately traded (see Figure 4), whereas the REX *chainlinker atoms'* adjustments and handoff/retrievals are time staggered as shown in Figures 5 and 6. The adjustment and handoff of P_i's "head of the list" chainlinker atom is completed long before P_{i-1} 's terminal chainlinker atom is processed. By arranging communication before computation, the preceding chainlink node's "need" is anticipated and no delay waiting should ever occur for this exchange. Although some minimal delay *might* occur when P_i reads coordinates sent forward by P_{i-1}, this should not promulgate through the interlinked chain due to each processor's relative independence.

The net effect of this protocol is simply that the chainlinker atoms are adjusted out of numerical sequence, or one could say that a new order of adjustment is established. Although this new order is no longer numerically sequential across the entire bondlist, all atoms are subject to the same constraint couplings, system-wide, and all SHAKE criteria for convergence are met in an overall fully consistent manner. When *all* bonds are "shaken", the resulting trajectory will gradually diverge from one generated using the sequential SHAKE. So, the REX trajectory is different, but every bit as legitimate.

Tables III and IV present efficiency, energy deviation, and percentage of CPU useage for the SHAKE and SHAKEP routines. Note that over a range of hypercube sizes ranging from 1-128 processors the calculated average total energies (Table III) are in close agreement and well within the rms deviations over the divergent trajectories. Divergence of the total energy begins after about 500 MD steps. There is apparently more variance between the trajectories generated by different machines (due to variant numerics such as floating-point representations, precision, order of operations, etc.) than between the trajectories of various sized hypercubes where differences are attributable to the REX SHAKEP mechanism. These data affirm our assertion that trajectories influenced by REX SHAKEP are equally as legitimate as one that results from applying the sequential SHAKE method. The REX nodechain with crosslink support is a robust way to bring this originally sequential procedure into the parallel mode.

Communication Optimization for Coordinate Regathering

The SHAKEP algorithm as discussed above should apparently scale nearly as the number of processors allocated to the job. But, there is some overhead due to the extra test statements and interprocessor communications which facilitate the REX mechanism. Because the ratio of overhead/computation (of position adjustments) in SHAKEP will grow as an individual processor's fraction of the global bond list is reduced, this overhead increases with an increase in the number of applied processors. But, another contributor of extra overhead must be considered if a direct comparison with sequential SHAKE performance is to be made; there is the communication required to globally regather the SHAKEP-adjusted coordinate array after convergence is reached. The procedure SHAKEPCOM uses a version of Gray Code [3c,9] which we have customized to suit this purpose. In brief, this is an iterative procedure for globally distributing (via internodal message passing) all updated coordinates to all processors. The number of Gray Code iterations equals the *dimension* N of the 2^N processors of the hypercube.

Unlike the situation during global regathering of the atomic forces, where every processor may contain nonzero force array values for nearly every atom in the system (the force array is often "fully loaded"), here each processor updates only N_A/N_P data elements, where N_A is the number of atoms and N_P is the number of processors. As the global accumulation progresses with each successive Gray Code iteration, the number of data elements each processor must communicate increases by a factor of two. But, no processor's array is ever fully loaded until after the last iteration is *over*.

Thus, because much smaller data messages can be transferred between processors, the communication time required to complete the global regathering of SHAKEP-adjusted coordinates is insignificant compared to that needed to regather the forces after evaluation of the nonbonded terms. But, it does decrease the efficiency of the otherwise nearly linear scaling of the parallel SHAKEP routine. As the number of processors (dimension of the hypercube) increases, the number of iterations of Gray Code message passing must also increase, and true scaling with number of processors is thus progressively tempered due to internodal communication overhead. The SHAKEP/SHAKEPCOM efficiencies listed in Tables III and IV provide a quantitative measure of this trend.

Nevertheless, as illustrated in Figure 7, parallel SHAKEP enables marked overall program speedup. Amdahl's limit for concurrent MD is raised by "shaking" bonds in parallel. Figure 7 demonstrates that at the 64-processor level, and after all other relevant routines have been parallelized, the *overall program* is sped up by nearly a factor of about 20 instead of 10, due exclusively to having parallelized the SHAKE routine. Tables III and IV show that regardless of the number of processors used the time spent in SHAKEP/SHAKEPCOM is not a significant portion (about 5%) of the overall run time despite the routine's nonlinear scaling. Although it is clear that SHAKEP/SHAKEPCOM performance leaves the linear realm after about eight processors (Tables III and IV show about 42-76 percent efficiency), further speedup is still obtained with each doubling of nodes. This confirms that the REX strategy scales positively with added processors, surmounting the problem of scaling only as the number of SHAKEP iterations as was found for the bidirectional relay pipeline method.

PARALLEL PERFORMANCE

Table V reports the parallel MD performance for the octanol/water system when 128 processors are compared to a single processor. Notice the change in the percentage of the total time contributed by any particular subroutine when 128 processors are used. For the single processor, the evaluation of the nonbonded "Nonbond" forces/energies is the primary contributor, accounting for 86.5% of the overall program's total run time. All other individual subroutines account for less than 5% each, with "Other" contributing only 1%.

With 128 processors, the nonbonded evaluation (which scales linearly, as shown in Figure 8) no longer dominates the calculation. Rather, it is the internodal communication, "NB_Comm" at 35%, required to globally regather the updated forces that dominates the calculation, along with "Other", now at 40%. The "Other" category consists of miscellaneous interprocessor communication and parallel overhead folded in with some unparallelized sequential code. "Shake" and "Pairlist" still contribute less than 5%. The subroutines that evaluate angle and dihedral forces/energies have become insignificant contributors. Because internodal communication and miscellaneous sequential code are now dominating, there is not much parallel speedup left to be had. The addition of more than 128 processors would only increase the magnitude of NB_Comm and Other by an amount more than the speedup to be gained by decreasing the Nonbond time. The point of no (parallel) return for added processors has been reached, essentially as a result of a communications bottleneck. If future distributed-memory machines are designed so that the internodal communications speed is

increased relative to processor speed, this limitation could be alleviated. It would also become worthwhile to parallelize most of the sequential code in the Other category. This will result in more nearly linear *overall speedup*, even when large numbers of processors are used.

In Figure 8, average scaling performance behaviors of all the relevant individual procedures is plotted. Here again, the speed of these routines is limited by the internodal communications rate. Evaluation of forces/energies for the nonbonded, bond, angle, and dihedral interactions all scale linearly with the number of concurrent processors used. But, when the internodal communications for the gathering of the force array, NB_Comm, is folded in with the nonbonded evaluation time, Nonbond, the result is nonlinear scaling. Alone, the pairlist generation code scales linearly, but the load-balancing work and the internodal communication associated with load balancing should be included as a necessary part of generating the pairlist. This accounts for the nonlinear behavior of the pairlist as shown in Figure 8.

The performance of SHAKEP, including SHAKEPCOM, is also reported in Figure 8. Although the scaling is nonlinear with number of processors, it scales positively with added processors. It should be kept in mind that while the performance of any individual subroutine is often of some interest, linear scaling with number of processors may not always be required in order to get good speedup of the *overall program*, the most important issue afterall. Recall that regardless of the number of processors used, the time required to execute SHAKEP/SHAKEPCOM reliably remains near about 5% of the overall total and is thus nondominating (*when* executed in parallel). The critical

factor in obtaining overall program speedup is that all parts of the code must be operating in parallel.

Tables III, IV, and VI tabulate various performance data for the four test systems studied. It is gratifying to note that at the level of 64 processors the code is executing at about one third the speed of a Cray YMP, or three times that of an FPS522. The machine specifics for the benchmark comparisons are: nCUBE2 hypercube running v3.0_beta with the distributed FORTRAN77, Cray Y-MP8/864 running Unicos v6.1.5a on a single head using compiler CF77_5.0, and FPS FPX522 machine (with vector units) running v.4.3.3 with the distributed version of FORTRAN77. Benchmarks on all machines were run using 64-bit precision. For both sequential machines, the code was optimized to run as fast as possible by using the available compiler options. Also remember that the AMBER3a code is completely vectorized, providing a rigourous (sequential) standard to measure against.

For each molecular system studied, the maximum overall program speedup obtained (over that of a single node) is in the neighborhood of a factor of 20-25 or so, at which point a monotonic limit is reached due to an interprocessor communications bottleneck. Speedup is measured as the total time to run 1000 MD steps on a single processor divided by the time it takes on N processors. When comparing multiprocessor timings with single-processor timings, we run an optimized *sequential* version of the code for the single processor timing. That way, absolutely no parallel overhead (such as domain decomposition, internodal communication, miscellaneous testing, etc.) is included in the single-processor timing. When timing multiprocessor jobs, every pro-

cessor is individually timed and the average of these is reported.

Tables IV and VI note the progressive decrease in efficiency of the overall program as more processors are applied. Efficiency is simply the overall program speedup achieved divided by the number of processors used. It measures the percentage of processor capability directly applied to speeding up the code. As the number of processors increases, the overhead grows, decreasing efficiency, mostly due to the more extensive interprocessor communications.

Figure 9 shows the increase in the cost of globally gathering the force array as processors are added. Each processor's set of freshly calculated forces must be summed into every other processor's force array. For this procedure, the curve is neither linear nor concave increasing, as one might expect. As the number of processors increases, the pairlist is broken into progressively smaller fragments, so each processor is responsible for calculating forces between fewer atom pairs. As the size of a certain processor's atom-pair sublist decreases, an increasing number of the $1,N_A$ atoms under consideration will form *no* pairs (within the atom-pair subdomain of this certain processor) and thus contribute no forces. So, in order to minimize interprocessor communication, only the nonzero forces are globally gathered. This way, with each doubling of processors the communication cost increase is slightly less than it was for the previous doubling of processors. The curve is flattening out. This indicates that for a simulation of massive size and many processors the time required to intercommunicate the forces would cease to be dominant. Unfortunately, we did not have access to a machine of greater than 64 nodes, each with an appropriate memory size, with which to demon-

strate such a system.

Another very important consideration for parallel versus sequential performance comparisons is the length of the nonbonded interaction cutoff radius. This affects the ratio of communication/computation for the overall program. The longer the cutoff radius, the greater the speedup and higher the efficiency that will result for multiprocessors versus a single processor. In this article, cutoffs that are typical for a production MD simulation have been used (see text of tables III, IV, VI). Had we used an "infinite" cutoff length, where all system-wide pairwise interactions are calculated, the speedups would have been much greater. Interprocessor communication would not dominate in that case, nor would extant sequential code until very high numbers of processors were applied, where Amdahl's limit would be reached. Overall program speedup would remain more nearly linear even at high N_P.

We can demonstrate (by extending the interaction cutoff) what the effect of increased simulation system size or reduced internodal communications time would be on parallel performance. Extending the cutoff gives a greater number of pairwise atomic interactions without actually expanding system size or increasing the communications requirements. Figure 10 illustrates the average trends in overall program speedup and efficiency as a function of increasing numbers of concurrent processors when a typical cutoff is used (broken line), versus when a very long cutoff is used (solid line).

With the typical "short" cutoff, a monotonic limit is approached at 64 processors. Past that number, more parallel overhead is added than the speedup obtained; thus a

downward turn results in the broken plot at 128 processors (not shown). It is very noteworthy, however, that even with a typical cutoff, the overall speedup stays near the linear realm, for up to 8 or 16 processors, where the processor efficiency ranges between 73-86%. Although there is less efficient use of added processors when 32 or 64 processors are applied, significant speedup is still gained. But notice (Table VII and Figure 10, solid line) that, as expected, parallel performance is greatly enhanced with the cutoff extension. Sixty-four processors give a full one-half Cray YMP speed, a factor of 50 speedup over a single nCUBE node. That translates into a very satisfactory 75-80% efficiency, or 90% at 32 processors, and essentially complete linearity when fewer processors are applied. AMBERCUBE MD would run much more efficiently on a distributed-memory MIMD hypercube for which the ratio of (internodal communications speed) / (processing speed) is lower than that of the nCUBE2.

SUMMARY AND CONCLUSIONS

We considered several molecular simulation systems of typical size and complexity in order to exemplify the general utility and performance properties of this fully parallel implementation of molecular dynamics, AMBERCUBE MD. The decomposition of the various MD data domains which partition over multiple processors was discussed. An example was given describing the task of load balancing the concurrent evaluation of the computationally intensive nonbonded interactions. A successful method for robustly parallelizing the SHAKE routine, the Reciprical Exchange Nodechain (REX SHAKEP) Methodology was introduced and explained in algorithmic

detail. All bond types — hydrogen-containing, heavy-atom-only, and crosslinks — can be "shaken" using this method.

Data demonstrating the degree of scaleup and efficiency with increasing numbers of processors has been presented for the overall program and for various parallelized subroutines. For four variant molecular systems, benchmarks were presented that compare the results obtained for an nCUBE machine, using from 1-128 processors concurrently, to the performance of two sequential vector machines. Procedures which did not rely on message passing (evaluation of bonds, angles, dihedrals, 1-4 nonbonds, and long-range nonbonds) scaled linearly with the number of processors applied. As expected, procedures which require significant amounts of internodal communication did not scale linearly. However, nonlinear speedup of a routine was acceptable in cases where a routine's percentage of the total run time remained nonlimiting as processors were added, as for the SHAKEP/SHAKEPCOM routine (5%). In the end, the interprocessor communication bottleneck was an impediment to obtaining maximal efficiency and overall program speedup.

When a typical nonbonded interaction cutoff radius was used, the overall program scaled up nearly linearly for as many as 16 processors, with overall efficiency declining but still scaling positively for as many as 128 processors. AMBERCUBE MD performed three times faster than an FPS522, or one third the speed of a single-processor Cray YMP when a typical spherical interaction cutoff radius length was used. When an nonbonded interaction cutoff was used, internodal message-passing communications did not dominate execution time. This simulates the effect that increased molecular

system size or faster internodal communication channels would have. In this case, the speedup and efficiency obtained (for many processors relative to one) were much greater, giving nearly linear efficiency with 64 nCUBE2 processors, achieving one half the speed of a Cray YMP.

While it is gratifying to demonstrate the successful adaptation of AMBER MD for parallel computing, an even greater speedup of MD will be needed in order to simulate processes which occur over timescales longer than tens of nanoseconds. Because the molecular dynamics methodology is used to simulate time-dependent events, it is an inherently sequential algorithm which does not lend itself to a coarse-grain linearly scaling form of parallelism. Yet, MD also resists parallel fine-graining because of the necessity that each processor must have "global" knowledge of the system-wide forces, a large array which requires updating and global regathering at every MD step. The resulting dilemma is a seemingly unavoidable communication bottleneck in the case of distributed memory message-passing architectures, or the delay due to memory contentions in shared-memory machines, and the problem invariably grows as the number of processors increases.

But we have demonstrated that the next generation of distributed-memory machines, which will consist of higher-performance individual processors and much faster interprocessor message-passing rates, should give excellent performance for this medium-grain parallel implementation of molecular dynamics. We are optimistic that, with advances in communication hardware and processing units on the horizon, an even greater extension of molecular dynamics calculations will certainly be realized in the

future. The nCUBE2 hypercubes used for this study send internodal data at a rate of 2.2 megabytes per second per channel, with 13 channels interconnecting nodes and one I/O connection. The nCUBE3 (due in 1994) will send internodal data at a rate of 50 megabytes per second per channel, with 16 channels, and two I/O connections. The processing unit speed will increase by approximately a factor of 30, giving an overall factor of 25 to 30 times faster machine, for nCUBE3 over nCUBE2[10]. While we are disappointed that the interprocessor communication / computation rate will remain essentially unchanged, we are happy to note that a simple extrapolation based on our results in this article indicates that the nCUBE3 will run AMBERCUBE MD (using a typical nonbonded cutoff) at six to eight times the speed of a Cray YMP.

REFERENCES FOR CHAPTER III

(1) (a) G. Seibel, **AMBER** 3.0 Revision A, Department of Pharmaceutical Chemistry, University of California, San Francisco, (1990). (b) U. C. Singh, P. K. Weiner, J. W. Caldwell, P. A. Kollman, **AMBER (UCSF)**, version 3.0, Department of Pharmaceutical Chemistry, University of California, San Francisco, (1986). See also: P. K. Weiner, P. A. Kollman *J. Comp. Chem.* **2**, 287, (1981).

(2) (a) J. E. Mertz, D. J. Tobias, C. L. Brooks III, U. C. Singh, *J. Comput. Chem.*, **12**, 1270-1277 (1991). This paper addressed parallelization of SHAKE on a shared memory MIMD machine, and implemented it for hydrogen-containing bonds, but not heavy-to-heavy atom bonds or crosslinked bonds. A clear mathematical description of constraint coupling in SHAKE is presented. (b) The next two papers addressed issues of constraint de-coupling and vectorization of SHAKE. Neither treated the problem of crosslink bonds. The method discussed in the second reference was more approximate and sometimes subject to nonconvergence. F. Muller-Plathe, D. Brown, *Comput. Phys. Comm.*, **64**, 7-14, (1991). G. Ganti, J. A. McCammon, *J. Comput. Chem.*, **7**, 457, (1986).

(3) G. C. Fox, M. A. Johnson, G. A. Lyzenga, S. W. Otto, J. K. Salmon, D. W. Walker, in *Solving Problems On Concurrent Processors. General Techniques And Regular Problems*, Vol **1**, Prentice-Hall Inc., Englewood Cliffs, New Jersey, ISBN 0-13-823022-6, (1988). (a) pp. 57 and 475. (b) chapter 3. (c) pp. 260-262, 482.

(4) H. L. Nguyen, H. Khanmohammadbaigi, E. Clementi, *J. Comput. Chem.*, **6**, 634 (1985). D. Fincham, *Molecular Simulation*, **1**, 1 (1987). T. W. Clark, J. A. McCammon, *Computers Chem.*, **14**, 219 (1990). R. D. Skeel, *J. Comput. Chem.*, **12**, 175 (1991).

(5) In the AMBER potential function V_{totl} is the potential energy of the system; K_r and r_{eq} are the bond stretching constant and equilibrium bond distance; K_Ω and Ω_{eq} are the bond angle stretching constant and equilibrium bond angle; V_n, n, and γ are the torsional force constant, periodicity of the torsional term, and the phase angle; A_{ij} and B_{ij} are the (Lennard-Jones) non-bond interaction parameters for particles i and j; R_{ij} is the interatomic distance between i and j; q_i and q_j are the atomic partial charges on i and j; ε is the effective dielectric constant; C_{ij} and D_{ij} are the hydrogen-bond interaction parameters for particles i and j; VDW_{scale} and EEL_{scale} are the van der Waals and coulomb scaling parameters for the intramolecular 1-4 nonbonded terms.

(6) S. J. Weiner, P. A. Kollman, and D. T. Nguyen, D. A. Case, *J. Comput. Chem.*, **7:2**, 230-252, (1986).

(7) W. F. van Gunsteren, H. J. C. Berendsen, *Mol. Phys.*, **34**, 1311-1327, (1977). J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **34**, 327-341, (1977). G. Ciccotti, M. Ferrario, J. P. Ryckaert, *Mol. Phys.*, **47**, 1253-1264, (1982).

(8) Wilfred F. van Gunsteren; Herman J. C. Berendsen *Angew. Chem. Int. Ed. Engl.* in *Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry*, **29**, 992-1023, (1990) M. P. Allen, D. J. Tildesley, in *Computer Simulation of Liquids*, Clarendon Press, Oxford University Press, ISBN 0-19-855375-7, chapt. 3, (1987).

(9) E. N. Gilbert, *Bell System Technical Journal*, "Gray Codes and Paths on the N-cube", **37**, p. 815, (May 1958).

(10) Personal communication from Rich Wyckoff, Director of Corporate Communications, nCUBE Corporation.

Table Ia. Assign initial (homogeneous) guess for residue-pair indices.

	I_{resI} J_{resI}	I_{resN} J_{resN}	Atom pairs	Deviation (%)
P0	(1, 1)	(76, 120)	196,207	+48
P1	(76, 121)	(165, 230)	237,264	+78
P2	(165, 231)	(281, 350)	55,705	−58
P3	(281, 351)	(560, 560)	42,858	−68

Initial guess "per processor" residue-pair range indices before load balancing assumes the number of atom pairs found will be directly proportional to the number of residue pairs tested. Processor P0 tests all (I, J) residue pairs in the range (1, 1) to (76, 120) for "active atom pairs." The global number of active atom pairs after pairlist generation on this MD step is 532,034. The number of residue pairs tested per processor is $(N * (N + 1))/2 = 157{,}080$, where $N = 560$, the total number of residues in the system. Also presented: the initial number of corresponding atom pairs found per processor and % deviation from the target value (532,034/4 = 133,008.5).

Table Ib. Determining which node-local atom pairs should be exchanged between processors.

	P0	P1	P2	P3
Atom-pair ranges initially associated with each processor before load balancing:				
P0	1	0	0	0
	196,207	0	0	0
P1	0	196,208	0	0
	0	433,471	0	0
P2	0	0	433,472	0
	0	0	489,176	0
P3	0	0	0	489,177
—	0	0	0	532,034
Atom-pair range exchanges between processors during load-balancing:				
P0	1	133,009	0	0
	133,008	196,207	0	0
P1	−133,009	196,208	266,017	399,025
	−196,207	266,016	399,024	433,471
P2	0	−266,017	0	433,472
	0	−399,024	0	489,176
P3	0	−399,025	−432,472	489,177
	0	−433,471	−489,176	532,034

A positive number on the diagonal of this matrix indicates the atom-pair range (APR) for pairs held and *kept* by a processor. A positive number on an off-diagonal indicates an APR that the processor of that row *sends* to the processor of that column. A negative sign indicates an APR that the processor of that row *receives* from the processor of that column. For example, processor P0 keeps its node-local atom pairs 1–133,008 but sends atom pairs 133,009–196,207 forward to processor P1.

Table Ic. Results of load balancing the atom pairs.

	I_{res1} J_{res1}	I_{resN} J_{resN}	Atom pairs	Deviation (%)
P0	(1, 1)	(36, 413)	133,054	.0003
P1	(36, 414)	(105, 542)	132,965	−.0003
P2	(105, 543)	(153, 329)	133,035	.0002
P3	(153, 330)	(560, 560)	132,980	−.0002

New "per-processor" residue-pair range indices, reassigned after load balancing. Also, the final number of node-local atom pairs held per processor after pairlist regeneration and percent deviation from target value (133,008.5).

Table II. Topological decomposition of the crambin/water system over 32 processors for the "shaking" of all bonds.

PROC	NBOND	FIRST-HB	LAST-HB	FIRST-AB	LAST-AB	HEADATM	TAILATM	HEADCOM	TAILCOM
P0	149	1	74	4093	4167	1	149	F	T
P1	144	75	144	4168	4241	149	291	T	T
P2	140	145	208	4242	4317	291	429	T	T
P3	138	209	278	4318	4385	429	566	T	T
P4	138	279	360	4386	4441	566	699	T	F
P5	138	361	498	0	0	700	837	F	F
P6	138	499	636	0	0	838	975	F	F
P7	138	637	774	0	0	976	1113	F	F
P8	138	775	912	0	0	1114	1251	F	F
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
P31	138	3955	4092	0	0	4294	4431	F	F

PROC is the processor identification number. NBOND is the total number of bonds assigned to PROC. FIRST-HB and LAST-HB are the first and final hydrogen-containing bonds contained within the section of the bondlist assigned to PROC. FIRST-AB and LAST-AB are the first and final heavy-atom-only bonds contained within the section of the bondlist assigned to PROC. This part of the bondlist undergoes fragmentation only when *all* bonds are shaken. A zero in either of these columns indicates no bonds of this type are assigned to PROC. Note that the AB bondlist follows numerically after the HB bondlist. Bonds 1–4092 are HB type; bonds 4093–4441 are AB type. HEADATM and TAILATM are the atoms at the extreme endpoints of the bondlist segment assigned to PROC. HEADCOM and TAILCOM indicate whether the HEADATM and/or TAILATM are shared between two processors and communication between the flanking processors will be required. F indicates a nonshared border atom and T indicates a shared border atom. Segments of the bondlist assigned to P0–P4 are interconnected, forming a pipeline.

Table III. Parallel performance data for octanol/water binary media system.

PROC	TIMETOT	f(FPS)	f(YMP)	TIMESHK	EFFS %	CPU %	ENERGY	EDEV
1	27,416	1/7	1/71	981	100	3.6	−841.6	17.8
2	14,278	1/4	1/37	637	77	4.4	−840.9	16.5
4	7526	1/2	1/19	392	63	5.2	−840.4	16.5
8	4356	1	1/11	290	42	6.7	−837.1	14.7
FPS	4080	1	1/10	268	—	6.6	−837.1	13.7
16	2507	1.6	1/6	174	35	6.9	−838.3	15.4
32	1697	2.4	1/4	109	28	6.4	−838.4	15.9
64	1380	3.0	1/3	74	21	5.4	−838.2	14.4
128	1239	3.3	1/3	60	13	4.8	−836.8	15.6
YMP	388	10.5	1	63	—	16.3	−854.1	22.8

All data is for double-precision calculations, using the same starting coordinates and protocol for all runs, 1000 MD steps: 2508 atoms, 228 *n*-octanols, 76 waters, constant temperature (348 K) and pressure (1 atm), periodic boundary conditions, 2-fs step length, a nonbonded cutoff of 12 Å, pairlist regeneration every 15 steps, with SHAKE tolerance (all bonds) set to 0.0001 Å. PROC indicates the number or type of (nCUBE) processors; FPS and YMP are described in the text. TIMETOT is the overall execution time for 1000 MD steps, in seconds. f(FPS) and f(YMP) are "the fraction of an FPS or YMP," i.e., one node of an nCUBE will execute the job at 1/7th the speed of an FPS, whereas 32 nCUBE nodes will execute the job at 3 times the speed of an FPS. TIMESHK is the time spent executing SHAKE (1 processor) or SHAKEP (multiple processors), in seconds. EFFS % is a measure of the processor efficiency with which SHAKEP/SHAKEPCOM is being executed. CPU % is the percent of the overall execution time spent on SHAKE(P), including the time spent globally regathering the converged coordinates in routine SHAKEPCOM. ENERGY refers to the final 1000-step average of the total system energy (potential plus kinetic) in kcal/mol. EDEV is the final average rms deviation (±) in the total energy in kcal/mol.

Table IV. Parallel performance data for crambin protein/water-solvated system.

PROC	TIMETOT	f(FPS)	f(YMP)	SPEEDUP	EFFT %	TIMESHK	EFFS %	CPU %
1	80,995	1/9	1/76	1.0	100	2633	100	3.3
2	41,947	1/5	1/39	1.9	97	1475	89	3.5
4	21,774	1/3	1/20	3.7	93	784	84	3.6
8	11,871	7/10	1/11	6.8	85	434	76	3.7
FPS	8520	1	1/8	—	—	1010	—	11.9
16	6847	1.2	1/6	11.8	74	257	64	3.8
32	4405	2.0	1/4	18.4	58	180	46	4.0
64	3270	2.6	1/3	24.8	39	175	24	5.5
YMP	1066	8.0	1	—	—	165	—	15.5
1	85,989	—	1/74	1.0	100	3285	100	6.7
32	4691	—	1/4	18.3	57	314	33	3.8
YMP	1162	—	1	—	—	208	—	17.9

Data in the top section refer to shaking hydrogen-containing bonds only. Data in the bottom section are shaking all bonds, including protein crosslinks. All data is for double-precision calculations, using the same starting coordinates and protocol for all runs, 1000 MD steps: 4431 atoms, 46 amino acids, 1259 waters, constant temperature (300 K) and pressure (1 atm), periodic boundary conditions, step lengths of 1 fs for SHAKE hydrogens only and 1.5 fs for SHAKE all bonds, a nonbonded cutoff of 10 Å, pairlist regeneration every 15 steps, with SHAKE tolerance set to 0.0001 Å. Column headings are as described in Table III. SPEEDUP is the relative speedup achieved over 1 nCUBE node when multiple nodes are applied in parallel. EFFT % indicates the total overall efficiency with which the multiprocessors are being used. For the SHAKE hydrogens-only system, the values for average ENERGY ($-10{,}856.2$) and the energy deviation ($\pm$ 684.8) agreed identically to seven-place accuracy for all nCUBE runs and to five-place accuracy when the FPS, YMP, and nCUBE are all considered; agreement was to four-place accuracy for the SHAKE all-bonds system and well within average deviations.

Table V. Performance of octanol/water system on 1 processor vs. 128 processors.

Routine	Time(s)	%
1 Processor		
Pairlist	944.23	3.43
Nonbond	23,721.95	86.53
NB_Comm	0.00	0.00
Bond	0.00	0.00
Angle	265.76	0.97
Dihedral	1213.16	4.43
Shake	980.54	3.58
Other	289.88	1.06
Overall	27,415.52	100.00
128 Processors		
Pairlist	30.28	2.44
Nonbond	203.66	16.44
NB_Comm	436.00	35.20
Bond	0.00	0.00
Angle	2.19	0.18
Dihedral	9.75	0.79
Shake	59.91	4.84
Other	496.76	40.11
Overall	1238.54	100.00

This data is for calculations referred to in Table III. Time is the execution time for each routine in seconds. Overall is the total execution time for 1000 MD steps. NB_Comm is the time cost for global gathering of the Nonbonded forces via internodal communication. Bond forces were not calculated because SHAKE was used. For 128 processors, Pairlist includes load balancing, Shake includes global regathering of adjusted coordinates, and Other includes miscellaneous internodal communications, parallel overhead, and extant sequential code.

Table VI. Performance data for alphalytic protease/inhibitor/cap water and TIP3P pure-water systems.

	Alphalytic protease			Liquid water		
PROC	TIMETOT	SPEEDUP	EFFT %	TIMETOT	SPEEDUP	EFFT %
1	39,809	1.00	100	9965	1.00	100
2	20,354	1.96	98	5106	1.95	98
4	10,629	3.75	94	2264	3.74	94
8	5778	6.89	86	1448	6.88	86
16	3284	12.12	76	860	11.59	73
32	2078	19.16	60	578	17.24	54
64	—	—	—	491	20.30	32
FPS	4299			1606		
YMP	444			161		

All data is for double-precision calculations, using the same starting coordinates and protocol for all runs, 1000 MD steps. For the protease system: 3152 atoms, 203 amino acids, and 357 cap waters, constant temperature (298 K), 12-Å mobile "belly," 1-fs step length, a 12-Å nonbonded cutoff, pairlist regeneration every 20 steps, no SHAKE. For TIP3P water: 648 atoms, 216 waters, constant temperature (298 K) and pressure (1 atm), 2-fs step length, a 10-Å nonbonded cutoff, pairlist regeneration every 10 steps, with SHAKE tolerance set to 0.0001 Å. Column headings are as described in Tables III and IV.

Table VII. Parallel performance of octanol/water and crambin/water with long cutoffs.

	PROC	TIMETOT	f(YMP)	SPEEDUP	EFFT %
Octanol/water: 1000 steps	1	221,107	1/90	1.0	100
	32	7796	1/3	28.4	89
	64	4461	1/2	49.7	78
	YMP	2450	1/1	—	—
Crambin/water: 5000 steps	1	1.95E6	1/92	1.0	100
	32	69646	1/3	28.0	88
	64	41830	1/2	46.7	73
	YMP	21212	1/1	—	—

A spherical cutoff of 30.0 Å was used to give about 3.1 million atom-pair interactions in the octanol/water binary media system. A cutoff of 18.0 Å gave about 5.5 million pairs for crambin/water. All data is for double-precision calculations using identical starting coordinates and protocol for all runs. Column headings are as described in Table III.

PARALLELIZATION OF MD MODULE AMBER

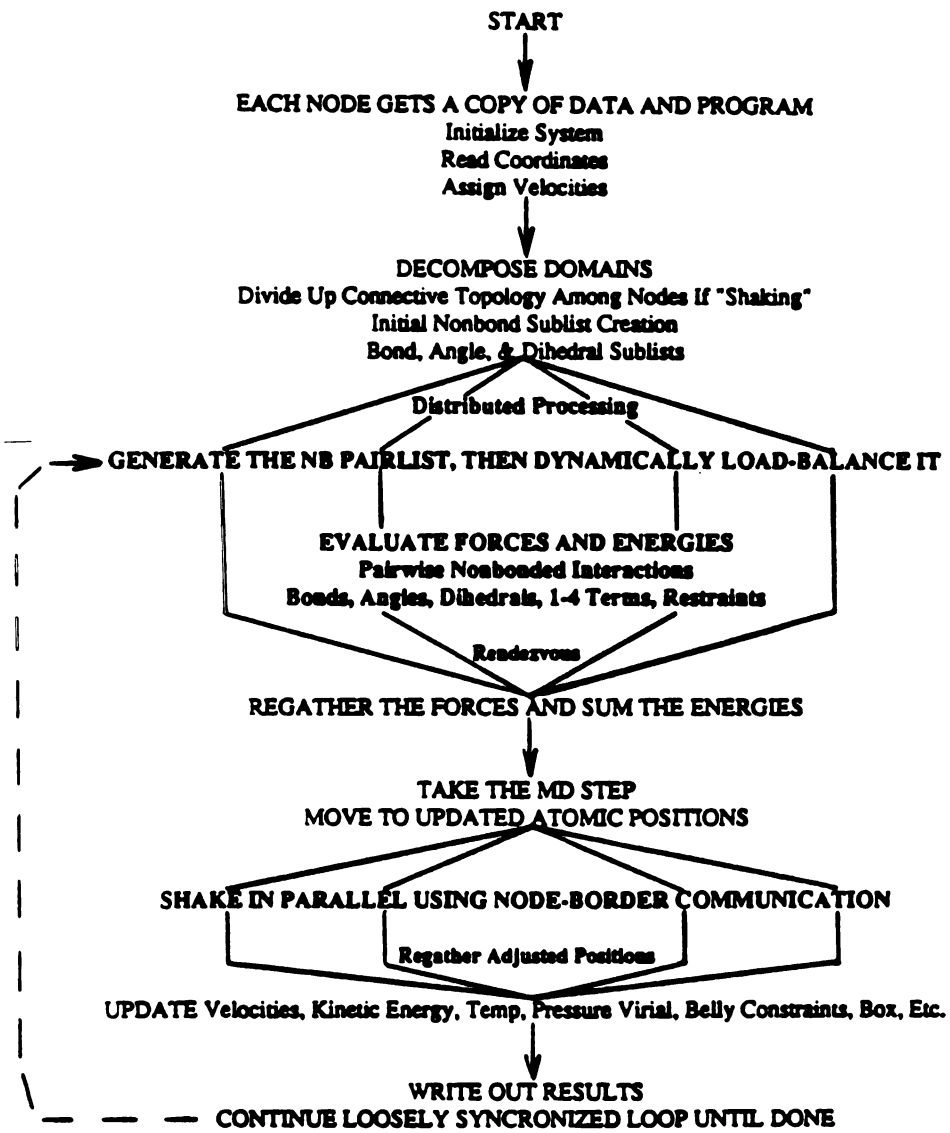


Figure 1. Decomposing an MD time-step for subtask distribution over multiple processors. Procedures carried out in parallel are indicated by multiple line expansions.

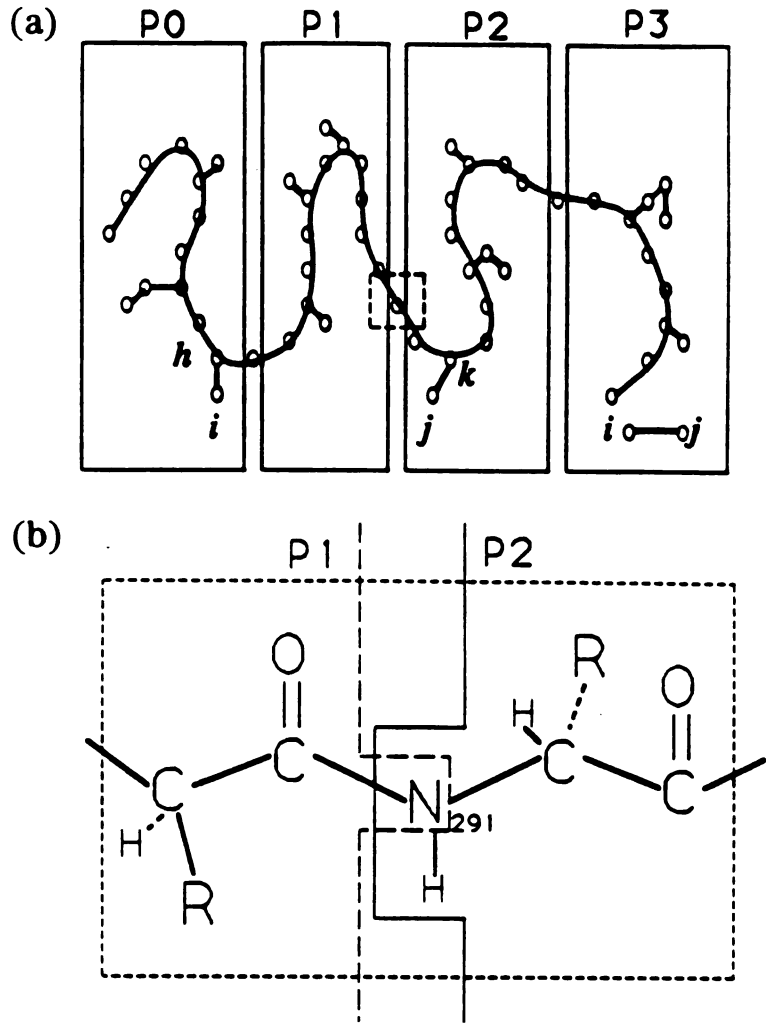


Figure 2. (A). Section of a long polymer chain with various side groups at low resolution, whose bondlist is divided into segments, each assigned to a different processing node. Atoms at node borders are shared. The section enclosed by a broken box between processors P1 and P2 is enlarged in Figure 2b. Atoms i and j are crosslinked in the molecular topology. Bond i—j is assigned to processor P3. **(B).** Blow-up of the indicated section of Figure 2a, showing the division of the connective topology at the mainchain peptide nitrogen, atom 291. There is a C—N(291) bond assigned to processor P1 and a N(291)—C bond assigned to P2. The N(291)—H bond also belongs to P2.

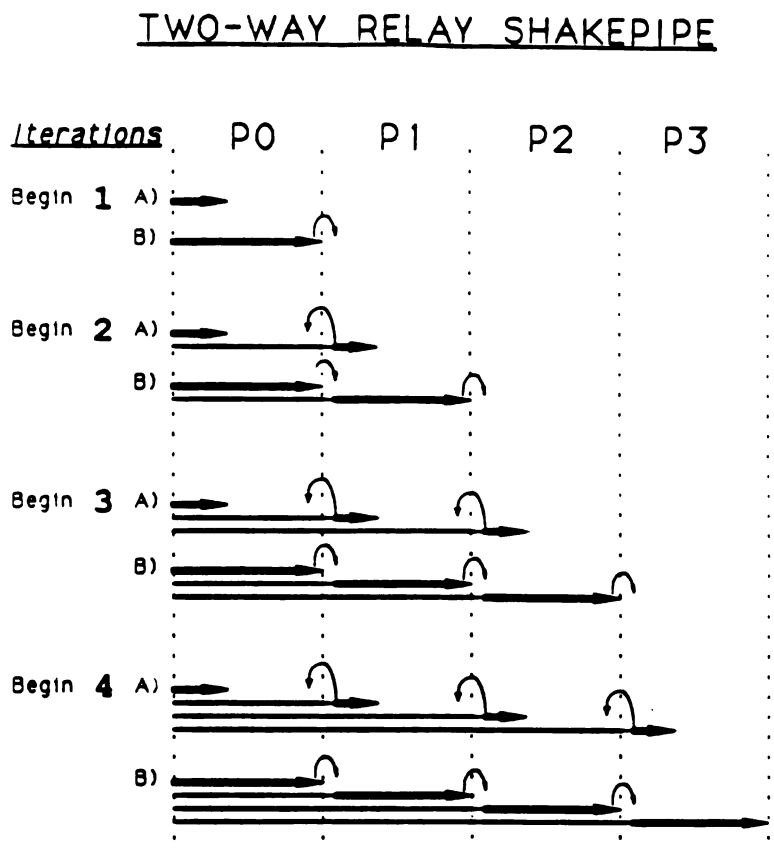


Figure 3. Bold arrow-headed lines indicate active processing along bondlist. Thinner lines denote previously processed territory. Thin arrow-headed curved lines indicate the "send" direction and time of interprocessor communication. **(A).** Processing of the bond list begins and the adjusted cojoiner atom's coordinates at the head of a processor's sublist is "handed back" to the processor containing the immediately adjacent segment of the bondlist (except in iteration 1). **(B).** sublist processing is completed and the adjusted coordinates of the terminal cojoiner atom are "handed forward" to the processor containing the immediately succeeding segment on the bondlist. The bondlist has been processed, once completely through, after iteration 4B. Note that during these initial steps in pipeline processing several nodes are in a nonactive wait state, illustrating pipe-loading overhead. By iteration 4, all nodes are processing concurrently.

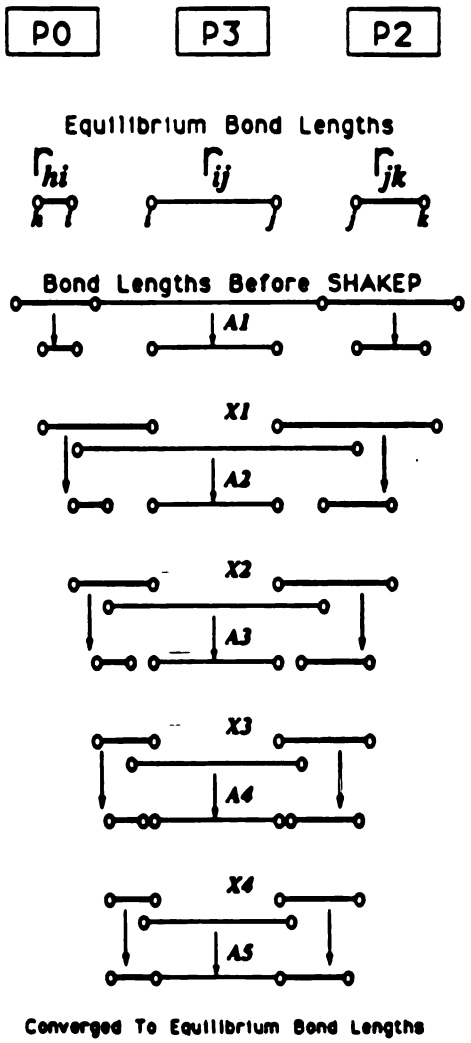


Figure 4. REX mechanism for atoms in crosslink bonds. Bond h—i is held by processor P0, bond j—k by processor P2, and bond i—j is a crosslink as indicated in Figure 2a. The equilibrium bond lengths are shown at the top. The situation after movement of the atoms due to impinging forces is referred to as *Bond Lengths Before SHAKEP*. At *A1*, the first SHAKEP bond length adjustment step is taken, with all three bonds tending toward their individual equilibrium values. In this schematic, bond adjustments are exaggerated. In the program, each adjustment is typically less than hundredths of an Å, and more like 10,000ths of an Å as convergence is approached. At *X1*, the first reciprocal exchange occurs. P2 and P3 trade atom j positions; P0 and P3 trade atom i positions. Iterations continue similarly until convergence is reached at *A5*.

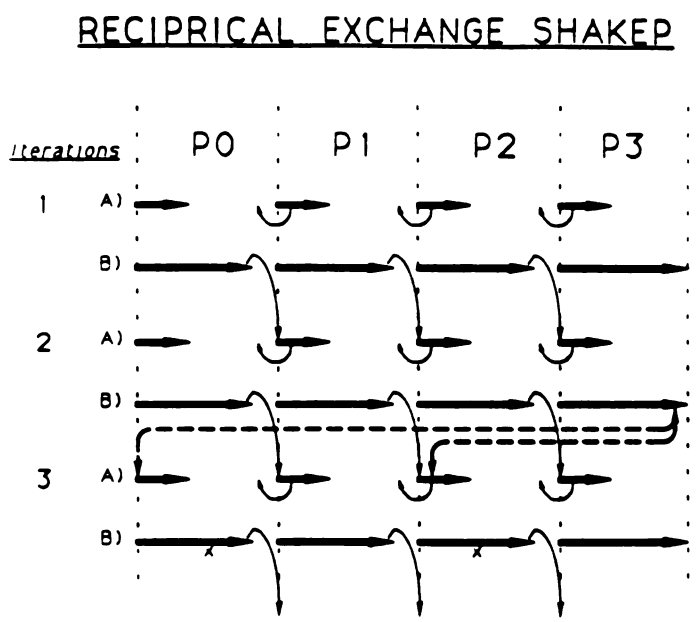


Figure 5. Parallel processing of the bondlist using the REX mechanism. Line types have the same meaning as in Figure 3. Crosslink communication (as described in Figure 4 and in the text) is indicated by bold broken lines shown only between iterations 2B and 3A for purposes of clarity. The Xs on processor P0's and P2's bondlist at iteration 3B denote the list position of the crosslink atoms i and j.

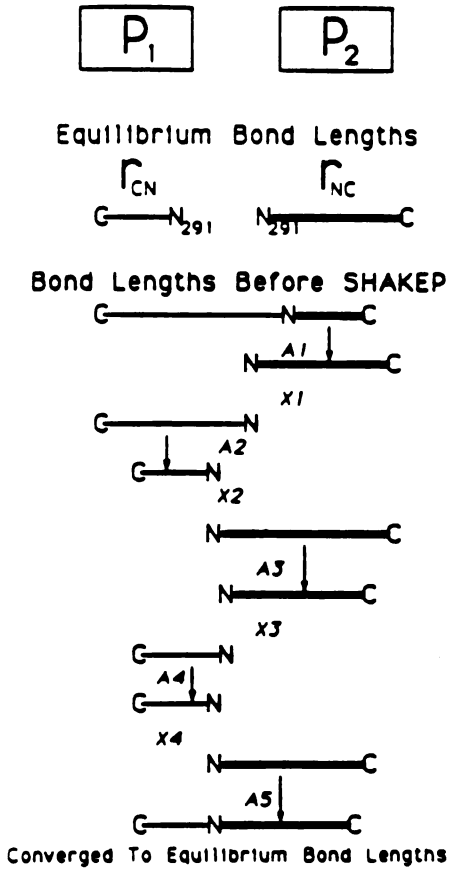


Figure 6. REX mechanism for node-linker atoms. Bond C—N is held by processor P1 and bond N—C by processor P2. The equilibrium bond lengths are shown at the top. The situation after movement of the atoms due to impinging forces is referred to as *Bond Lengths Before SHAKEP*. At *A1*, the first SHAKEP bond length adjustment step is taken, with both bonds tending toward their individual equilibrium values. At *X1*, the first "half" of a reciprical exchange occurs; P1 takes the atom N291 position that resulted from the adjustment of bond N—C in processor P2. At X2, the second half of the reciprocal exchange is made; P2 takes the atom N291 position that resulted from adjustment of bond C—N. Iterations continue similarly until convergence is reached at *A5*.

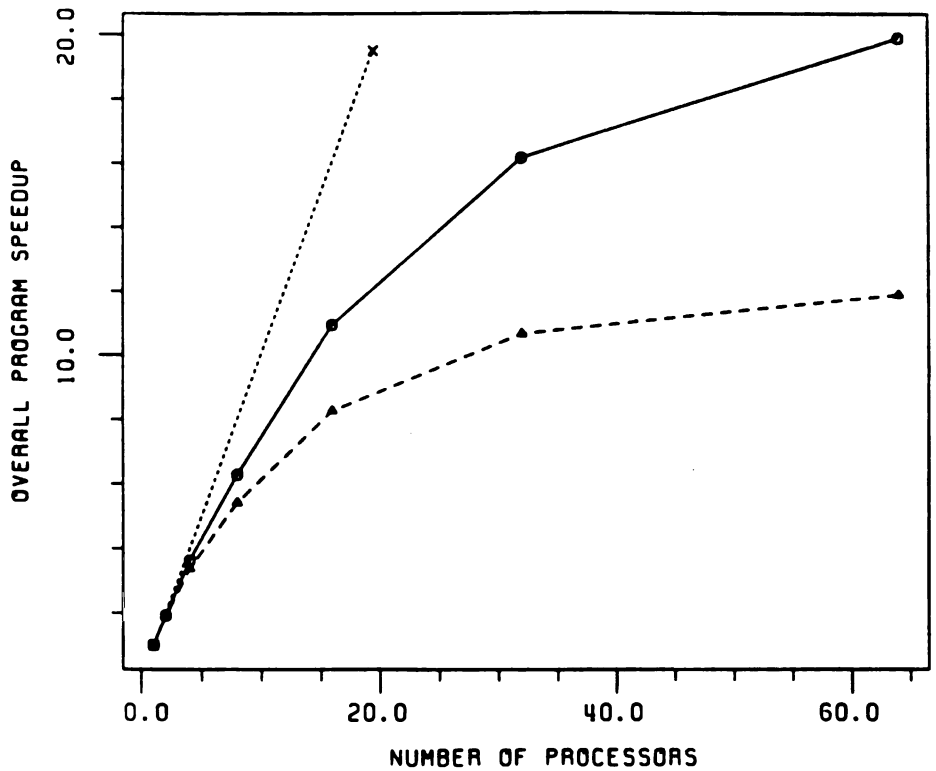


Figure 7. Overall program speedup vs. number of processors when SHAKEP is used (solid line) or sequential SHAKE is used (long broken lines). The line ending at point X is a yardstick of perfectly linear speedup given for comparison. The data used to generate this graph was an average taken over all shaken test systems described in Tables III, IV, and VI.

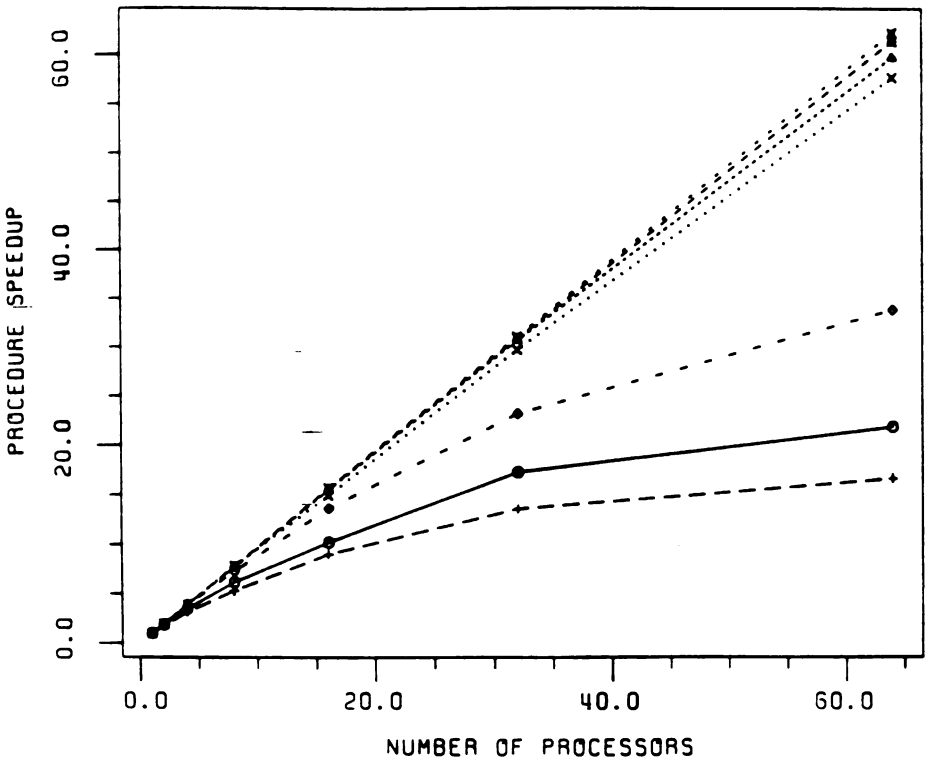


Figure 8. Performance of individual parallelized procedures. Starting from the top of the plot, the symbol key is as follows: Dihedral and 1-4 evaluations, crossed circle; angle evaluation, asterisk; nonbonded evaluation (ignoring force regathering communications), triangle; bond evaluation, X; nonbonded evaluation including force regathering communication, diamond; pairlist generation plus load balancing and load-balancing communications, circle; SHAKEP and SHAKEP internodal communications, plus sign. The data used to generate this graph was an average taken over the four test systems described in Tables III, IV, and VI.

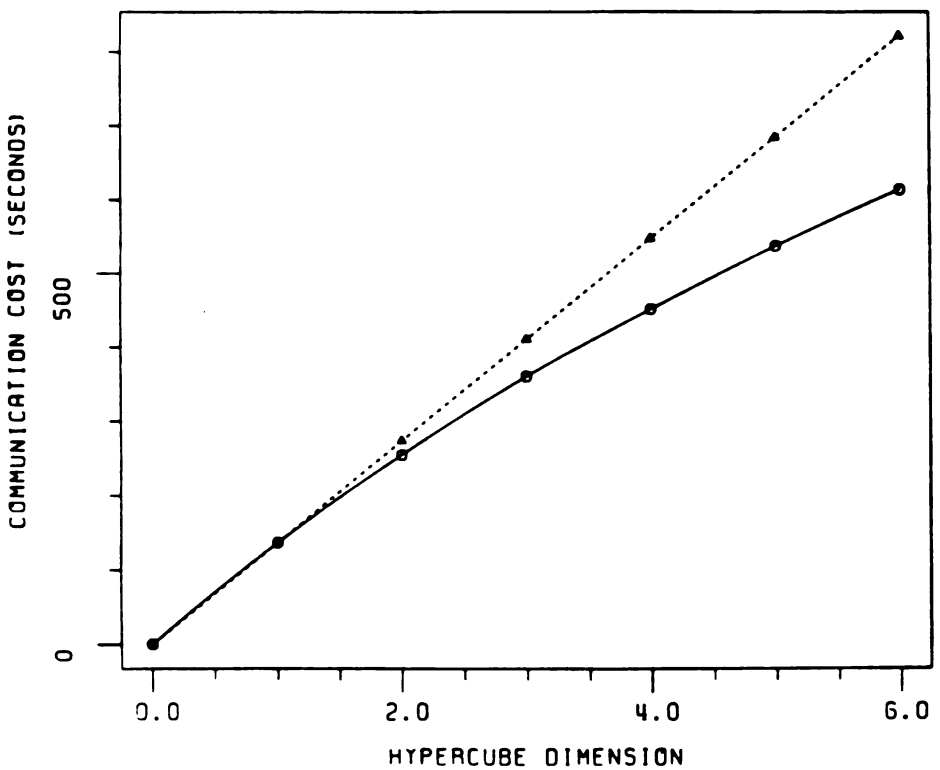


Figure 9. Cost of global summation of force array vs. hypercube dimension. Actual communication times for the crambin/water simulation are indicated by circles along the solid line. If communication accumulated linearly, the broken line would result.

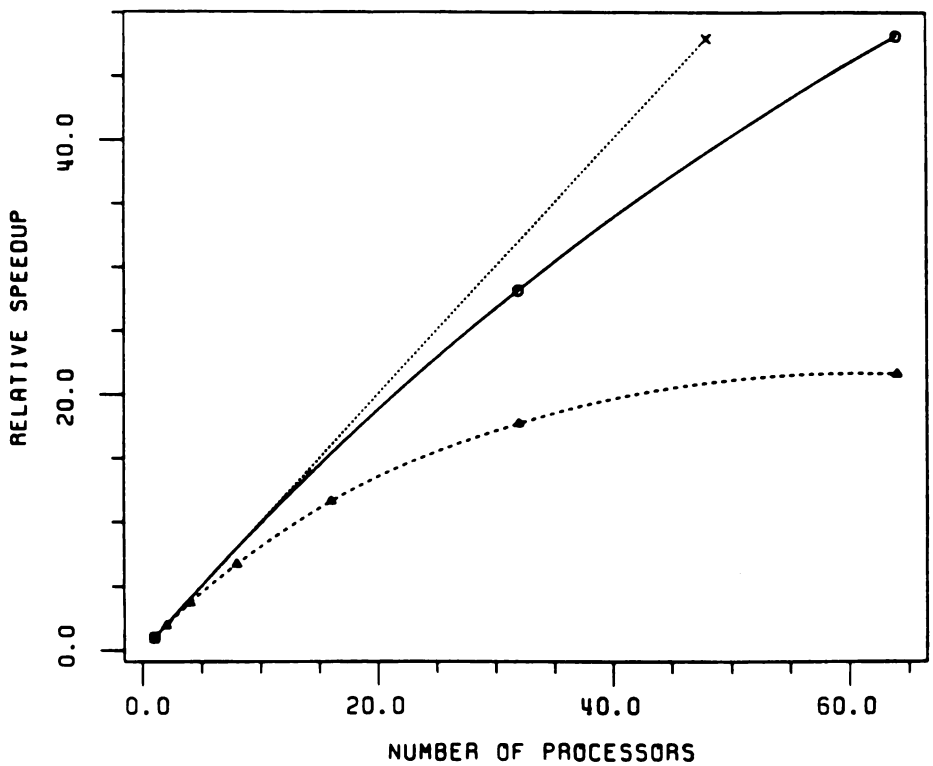


Figure 10. Overall program speedup vs. number of processors. Solid line with circles, result when a decrease in the communication/computation ratio is simulated by running systems with an extended nonbonded cutoff (see Table VII and the text). Beyond 64 processors, the interprocessor communications do not dominate; a greater speedup would result with added processors. Broken line with triangles, results for systems with a typical "short" nonbonded cutoff. Beyond 64 processors, interprocessor communications enforce a monotonic limit where little or no further speedup is obtained with added processors. A linear yardstick is included for the sake of comparison. The data used to generate this graph was an average taken over the four test systems described in Tables III, IV, and VI.

CHAPTER IV

Free Energy Perturbation Calculations On Parallel Computers: Demonstrations Of Scalable Linear Speedup

The text of this dissertation chapter III has been submitted for publication to The Journal of Computational Chemistry in July, 1993. The coauthor listed in the JCC article, Peter A. Kollman, directed and supervised the research which forms the basis of this chapter.

Permission to preprint has been granted from John Wiley & Sons, Inc.

Reprinted with permission from: S. E. DeBolt, P. A. Kollman, *Journal of Computational Chemistry*, Submitted (1993). Copyright 1993 John Wiley & Sons, Inc.

Free Energy Perturbation Calculations On Parallel Computers: Demonstrations Of Scalable Linear Speedup

ABSTRACT

A coarse grain parallel implementation of the free energy perturbation (FEP) module of the AMBER molecular dynamics program is described and then demonstrated using five different molecular systems. The difference in the free energy of (aqueous) solvation is calculated for two monovalent cations $\Delta\Delta G_{aq}(Li^{+} \rightarrow Cs^{+})$, and for the zero-sum ethane-to-ethane´ perturbation $\Delta\Delta G_{aq}(CH_3\text{-methyl-X} \rightarrow \text{X-methyl-}CH_3)$, where X is a ghost methyl. The difference in binding free energy for a docked HIV-1 protease inhibitor into its ethylene mimetic is examined by mutating its fifth peptide bond, ΔG(CO-NH→CH=CH). A potassium ion (K^{+}) is driven outward from the center of mass of ionophore salinomycin (SAL^{-}) in a Potential of Mean Force calculation $\Delta G_{MeOH}(SAL^{-}*K^{+})$ carried out in methanol solvent. Parallel speedup obtained is linearly proportional to the number of parallel processors applied. Finally, the difference in free energy of solvation of phenol versus benzene, $\Delta\Delta G_{oct}$(phenol→benzene), is determined in water-saturated octanol, and then expressed in terms of relative partition coefficients, $\Delta\text{Log}(P_{o/w})$. Because no interprocessor communication is required, this approach is scalable and applicable in general for any parallel architecture or network of machines. FEP calculations run on the nCUBE/2 using *50 or 100* parallel processors

were completed in clock-times *equivalent to or twice as fast* as a Cray Y-MP. There is discussion and analysis regarding the difficulty of ensuring adequate system equilibrium when a gradual configurational reorientation follows the mutation of the Hamiltonian. The results of a successful protocol for overcoming this equilibration problem are presented. The types of molecular perturbations for which this method is expected to perform most efficiently are described.

INTRODUCTION

In order to progress in their studies, scientists naturally need to have their questions answered as rapidly as possible. Fortunately for chemists seeking answers that rely on the results of computation-laden simulations, multiprocessor computers promise ever increasing calculation speeds as more processors are added. A major goal in adapting computational problems to parallel processing is to achieve "linear speedup" (speedup proportional to the number of processors applied, ideally with no monotonic limit) when many processors are used. There is a growing wealth of parallel computer power which stands waiting to be harnessed to computational chemistry software. But the efficient adaptation of chemical modeling algorithms for massive parallelism can prove to be a formidable and time consuming task. This point is exemplified by the difficulties encountered in recent attempts to parallelize molecular dynamics (MD) [1a,b].

The difficulty in distributing the work of many discrete MD time-steps over multiple processors lays in the fact that successive steps in a time-evolved trajectory are dependent on the results of preceeding steps. Typically, the algorithmic tasks in the calculation of molecular interactions during *each* individual MD time-step are broken down into subtasks and "farmed out" over the number of available processors (N_p), with appropriate data-decomposition and load-balancing (see Figure 1A). Then, for distributed memory architectures, the large data arrays (coordinates or forces, and velocities) are updated and globally regathered to all processors. This occurs at each time point during the evolution of an MD trajectory. The resulting dilemma is a communication bottleneck in the case of distributed memory message-passing architectures, or a delay

due to memory access contentions in shared memory machines, with the problem growing in significance as the number of processors increases. The speedup obtainable from parallelism eventually approaches a monotonic limit as added processors cause the time spent in data management to overshadow that devoted to computation. (We obtained efficient MD speedup on the nCUBE2 using 16 processors or less, even for relatively small sized systems, and further nonlinear speed increases when using up to 64 processors[1a].) Thus, either a complete redesign of the MD algorithm (e.g. dividing coordinate space among processors), or further enhancement of parallel machines' capabilities to manage data is required in order to take advantage of massive parallelism. The newest parallel computers offer an order of magnitude more rapid interprocessor communications and data retrieval [1c]. (We expect these new machines to run our parallel MD very efficiently even when employing many more processors than reported in reference 1a.)

But the current generation of parallel computers can serve quite well for "coarsely grained" tasks, where the overall task subdivides into N identical time-independent procedures which can be computed simultaneously rather than successively. Given multiprocessors which individually compute at the same speed as a given sequential machine, it is possible to complete a parallel calculation in time τ_{seq}/N_P (where τ_{seq} is the calculation time on the sequential machine). Alternatively, when completeness in statistical sampling of a distribution is an issue, it is possible to increase the accuracy of a calculation. We describe here an approach that takes advantage of the underlying coarse grain character of free energy perturbation calculations. First a brief review of

the FEP method is presented. This provides a background for understanding how time dependent reorientations of a system (which are due directly to perturbation of the Hamiltonian) may influence the calculated free energy value and potentially inhibit parallelization. We discuss a stratagem for overcoming this time dependent aspect of FEP calculations. The results of sequential and parallel FEP computations for five different molecular systems are then compared with regard to their comparative processing speeds and numerical agreement. Based on these results, the summary outlines the types of model systems which are the best candidates for the effective application of this coarse grained parallel FEP approach.

FREE ENERGY PERTURBATION BACKGROUND

From classical statistical mechanics, it can be shown that the free energy difference between two states A and B is given by:[2]

$$\Delta G = G_B - G_A = -RT\ln\langle e^{-(H_B(\mathbf{x})-H_A(\mathbf{x}))/RT}\rangle_A \quad . \tag{1}$$

Here, G_A and G_B are the respective free energies of states A and B, $H_A(\mathbf{x})$ and $H_B(\mathbf{x})$ are the potential energies using the Hamiltonians appropriate for states A and B, both evaluated for the same configuration $\mathbf{x}$ of the system, R is the gas constant and T is the temperature. $\langle\rangle_A$ refers to the average of the enclosed quantity for the ensemble of states appropriate for state A. Free energy calculations based on master equation (1) are termed Free Energy Perturbation (FEP).

Molecular dynamics can be used to generate an ensemble of states to evaluate the requisite average. A difficulty in implementing Equation (1) is that the ensemble evolves according to the potential energy function corresponding to A, but to correctly estimate the average we must sample not only those configurations favorable for state A, but also those favorable for state B. If states A and B are too dissimilar, this may not occur. This is the "orthogonality problem."

Fortunately, free energy is a state function. So we can calculate the total free energy between any pair of physical endpoints as the sum of the free energies between a series of non-physical intermediates. In practice, a coupling parameter λ is introduced into the Hamiltonian (potential energy function). The potential energy function used here is of the standard form:[3]

$$H_{Total} = \sum_{bonds} K_b\,(b - b_{eq})^2 + \sum_{angles} K_\theta\,(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}\,[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left\{ \varepsilon_{ij} \left[\left[\frac{r_{ij}^*}{r_{ij}} \right]^{12} - 2 \left[\frac{r_{ij}^*}{r_{ij}} \right]^{6} \right] + q_i q_j / \varepsilon r_{ij} \right\} \quad . \tag{2}$$

The λ dependence is introduced as

$$H_{Total} = \lambda H_A + (1-\lambda) H_B \quad , \tag{3}$$

where H_A = H(λ=0) and H_B = H(λ=1). In the Amber program, this progressive mutation of the Hamiltonian as λ changes from 0→1 is carried out as described in Pearlman[4].

A total free energy change can then be calculated as the sum of free energy changes

between these more similar non-physical "λ" intermediates:

$$\Delta G = \sum_{i=0}^{N-1} -RT\ln\langle e^{-(H(\mathbf{x},\lambda_{i+1})-H(\mathbf{x},\lambda_i))/RT}\rangle_{\lambda_i} \quad . \tag{4}$$

where $\lambda_0 = 0.0$ and $\lambda_N = 1.0$. Each interval in λ, $\delta\lambda_i=\lambda_{i+1}-\lambda_i$ is typically referred to as a "window." These intervals are typically of fixed width, but this is not a requirement[4].

An alternative to the FEP method described above is Thermodynamic Integration (TI). This method is based on a different master equation, also readily derivable from statistical mechanics:

$$\Delta G = \int_0^1 \langle \partial H(\mathbf{x},\lambda)/\partial\lambda\rangle_\lambda d\lambda \quad . \tag{5}$$

The variables are defined as before. Equation (5) is applied by evaluating the integrand at a set of intermediate values of λ, and then using a standard procedure to estimate the integral from these points. In Amber/Gibbs, the well-known simple trapezoidal rule[5] is used by default to make the estimate. One must ensure that enough points of the integrand are evaluated to define the integral. This will depend on how quickly and smoothly the integral is changing.

In carrying out a free energy calculation, we perform NEQUIL steps of equilibration and then NCOLL steps of data collection at every value of λ, i.e. over all "windows". The equilibration phase allows the system to relax to reflect the abrupt change in H(λ) that occurs each time we change windows. Sometimes the number of windows N is defined to be so large (e.g. ≥20000) that the changes in H(λ) with each subsequent

window are very small. It has sometimes been assumed (but not proven) that in such a case the system can remain in constant equilibration as λ changes, and that we can calculate the free energy setting NEQUIL=0 and NCOLL=1. This limiting-case approach of both FEP and TI has been termed "slow growth".

In a standard, sequential, free energy simulation, the chain of calculations at different values of λ is carried out in series, that is, λ=0,λ_1,λ_2,.... There is, in fact, no requirement that this order be followed. But since such an approach starts equilibration for any λ_i state with the configuration for the closest previous λ_{i-1} state, it is expected that equilibration will be most efficient when the calculation is carried out in this manner. As we will discuss below, for free energy calculations carried out in a parallel framework, an alternative approach where the starting configuration for state λ_i is not necessarily λ_{i-1} may be overall more efficient.

Physical states A and B in the above formalism can refer not only to compositional changes (changes in atom types, e.g. a change in the nature of a substituent), but also to conformational changes[6,7], such as translation along a vector. A free energy calculation carried out for a conformational change is termed a Potential of Mean Force (PMF) simulation; the result from such a calculation is a profile of the free energy as a function of the conformational variable being changed. Within Amber/Gibbs,[8] such calculations have been implemented by constraining an internal variable of interest during molecular dynamics using a holonomic constraint[7]. This makes the internal coordinate a function of λ_i. $H(\lambda_{i+1})$ in Equation (1) is then evaluated by modifying the internal coordinate to its appropriate value for λ_{i+1} before evaluating the energy. (For example, for a

distance restraint this would correspond to a rigid body translation along the distance vector).

One difficulty with free energy simulations is that--in the general case--we still have no reliable measure of the error associated with such a calculation. Instead, two estimates are frequently reported. One, the superposition error, results when in FEP we calculate the free energies both "forward" ($\lambda_i \rightarrow \lambda_{i+1}$) and "backwards" ($\lambda_i \rightarrow \lambda_{i-1}$) at each point of the λ pathway. At the end of the simulation this will provide two correlated estimates of the total free energy for the simulation. It has been shown that such errors are *at best* a test to ensure the window spacings are small enough to avoid the "orthogonality problem." Small values of such errors usually indicate the window spacing is sufficiently small. These errors are completely unreliable as an estimate of the total error in the calculation[7].

The second error estimate typically determined is the "hysteresis error", which is the difference between two separate free energy simulations, one in the forward (0→1) and one in the reverse (1→0) direction. Convergence of these values (accounting for sign) is a necessary *but not sufficient* criterion for convergence[7]. The difference between such values provides a crude lower bound on the error in a calculation.

In practice, the degree of accuracy and precision in the calculated value of ΔG are controlled by several factors: the time spent equilibrating at the endpoints λ_0 and λ_1, the RAM time (for parallel computations), the time spent equilibrating and collecting data in each window, and the size of the $\Delta\lambda$ interval spanned by each window.

PARALLEL STRATEGY

FEP calculations break down naturally for perfect coarse grain parallelism. The total free energy change for a process ΔG_{total} can be broken down into N summable free energy increments or "windows", ΔG_W. The calculation divides along the multiple lambda-points or windows, with each independent lambda-point λ_K assignable to an individual processor, as illustrated in Figure 1B. No interprocessor communication of large data arrays is required as each processor independently computes ΔG_W at its assigned λ_K, The parallel speedup obtained relative to a single processor is thus directly proportional to the number of processors applied. For as many ΔG-windows as are desired, speedup is linear without monotonic limit.

The configuration of a molecular system, or more precisely the character and distribution of states sampled in window W_K, is dependent on $H(\lambda_K)$ and thus on the numerical value of the mutation coefficient λ_K. So far in this discussion, we have assumed that the ΔG_W value calculated at each W_K is theoretically independent of the results of other windows $W_{I \neq K}$. However, in practice the starting configuration for W_K is prepared during the equilibration and sampling of the preceding window W_{K-1} at λ_{K-1}. Some finite period of simulation time, the "equilibrium phase" in FEP calculations, is then required to relax the system at λ_K. Often, the required period of equilibration is short.

But in some systems, relaxation toward λ_K-equilibrium partially occurs during the course of the perturbation trajectory, as a gradual configurational reorientation in

response to successive changes in lambda, $\lambda_{J<K}$. There is the possibility of long-time λ-dependent structural relaxations in mutated proteins, in the arrangement of the surrounding solvent media, or possibly in the radical reorientation of a mutated enzyme-bound ligand. This is not a problem for FEP calculations on traditional sequential machines, as the windows are treated in ordered succession, each window at λ_J preparing the system configuration for the next λ_{J+1}. But a slow component of lambda-dependent configurational relaxation may inhibit the simultaneous processing of multiple windows in the parallel treatment. This would reduce the parallel efficiency since each lambda-window's equilibration phase must be long enough to ensure that slow reorientations are essentially completed prior to the data collection phase. In fact, a requirement for slow sequential equilibration could defeat coarse-grain parallelism.

This potential equilibration problem is equivalent to the general difficulty described above for parallelization of molecular dynamics. Alternatively, the entire equilibration phase *could* be carried out via sequential processing over the $\Delta\lambda$-span. Only the data collection phase of each W_K (see Figure 1B) would remain perfectly parallel. But the best overall parallel speedup of FEP achievable in a case of non-simultaneous equilibrations would only be near a factor of two or three at best. More precisely, the best speedup achievable would be $(S_{collect} + S_{equil}) / S_{equil}$, where S is the number of MD steps of equilibration or data collection.

What is needed for simultaneous processing of λ-windows is a method to prepare semi-equilibrated starting states at each λ-point. Parallel speedup proportional to N_P will be obtained only when the time spent equilibrating the simultaneous parallel λ-

windows is less than or equal to the per-window equilibration time spent during sequential processing. The method must reduce or eliminate the influence that $\Delta\lambda$-mediated gradual configurational reorientations have on the calculated free energy. Three possible courses for the resolution of this potential problem are explored:

1) Dynamics is simultaneously initiated in all λ-windows from a *single starting configuration*, the initial-endpoint state. Although there are cases where this simplistic approach may actually be preferred, it is essentially an experimental "control" (see Figure 2A).

2) The perturbation trajectory is broken into a *series of M blocks* of N windows, where each λ-window within a block is initiated from a single starting configuration, and each succeeding block uses the configuration of the last window of the preceding block as its initial starting structure (see Figure 2B).

3) A *rapidly accessible manifold* (RAM) of starting states is generated by driving the system through the perturbation trajectory (across the complete λ-span) via a quick mutation employing the traditional sequential FEP approach. At specified intervals during the RAM, coordinates and velocities are output to a file, generating starting states at N_P different values of λ. The hope is that any gradual λ-dependent configurational reorientation will be resolved during the RAM to the degree that only a short equilibration phase at each lambda-window is then required. Naturally, the free energy calculated during a RAM run is expected to be inaccurate (see Figure 2C).

COMPUTATIONAL DETAILS

All parallel calculations were carried out in 64-bit precision using a version of the AMBER/Gibbs4[8] program modified to facilitate coarse-grain parallelism on the nCUBE/2 v3.1 running the distributed Fortran77. Sequential calculations used the standard issue AMBER/Gibbs4 on either a (single head of a) Cray Y-MP8/864 or a Hewlett-Packard 730. The program was compiler-optimized for each machine, taking full advantage of vectorization on the Cray Y-MP (running CF77_5.0.4.2 Fortran).

The task of submitting N_P parallel jobs, one for each lambda point, and managing multiple file-set input/output is relegated to a command script. This ensures the generality of the program for any multiprocessor setup. Special communications between the executing image, the submission command script, and the operating directory are instituted during submission so that redundant files are eliminated as they are generated. Large input files are handed off from one processor to the next as the multiple submission of N_P jobs progresses. This avoids duplication and accumulation of large files during submission and execution, thereby maintaining acceptable run-time memory requirements. Upon job completion, a command script is used to coalese the multiple output files into a more convienent single file format.

An MD stepsize of 2.0 femtoseconds was used for all calculations except those involving ethane where 1.0 fsec steps were taken, and the HIV-1 protease where 1.5 fsec steps were taken. All molecular bonds were constrained to their equilibrium lengths with holonomic constraints via SHAKE[9], within a tolerance of 0.0001 Å or finer; in the HIV-1 protease/inhibitor studies only the hydrogens were contrained.

Simulations were carried out with periodic boundary conditions, at constant pressure (1 bar) and temperature, with separate temperature coupling for the solute and solvent[10]. For the HIV-1 protease a non-periodic water "cap" of 20.0 Å radius was centered between residues ASP25 and ASP125 of the active site. Nonbonded pair lists were updated every 20 MD steps, except for the calculations in octanol/water which were updated every 30 MD steps. Each system was minimized and MD relaxed at the appropriate perturbation endpoint, i.e. λ_1 and/or λ_0, on the Cray Y-MP. This could have been accomplished via fine-grained parallel MD, but we were able to saturate the available processors of the parallel machine with FEP calculations. Further details regarding equilibration and collection times for each FEP run can be found in the tables.

For the aqueous $Li^+ \rightarrow Cs^+$ perturbation the periodic box contained a cation and 189 SPC waters[11] and averaged approximately 18.0 Å per side. The temperature was maintained at 300 Kelvin, coupled to a heat bath with a coupling relaxation time τ_T of 0.1 psec. The pressure coupling τ_P was also set to 0.1 psec. An 8.0 Å residue-based nonbonded cutoff was used for atomic interactions. For the perturbation, the λ_1 state refers to the solvated lithium cation, the λ_0 state to the solvated cesium cation. Each end-state was relaxed for 50 psec prior to the perturbation calculations.

For the aqueous ethane-to-ethane′ zero-sum perturbation the periodic box contained an ethane solute and 211 SPC waters. It averaged approximately 18.5 Å per side. This system was first described by Pearlman[13]. The temperature was maintained at 300 Kelvin, with a τ_T of 0.1 psec; τ_P was 0.1 psec. An 8.0 Å residue-based

nonbonded cutoff was used. The end-state was relaxed for 55 psec prior to the FEP calculations.

The HIV-1 protease/inhibitor system is described by Wlodawer[13], the specific JG365 inhibitor by Rich[14]. The molecular mechanical parameters used here are as in Cieplak[15]. The system consists of the dimeric enzyme, the short peptide inhibitor JG365 docked in the active site, 207 cap waters, and 95 crystal waters. The fifth peptide bond is "perturbed" into its ethylene mimetic, i.e. the CO-NH group mutates into a CH=CH group. The temperature was maintained at 298 Kelvin, with a τ_T of 0.4 psec. An 8.0 Å residue-based nonbonded cutoff was used. The λ_1 and λ_0 end-states were each relaxed for 50.0 psec prior to the TI/FEP calculations.

For the Sal^-*K^+ PMF calculation, one salinomycin ionphore and one potassium cation were solvated in a periodic box of 472 methanol solvent molecules. Salinomycin is a 122 atom noncyclic polyether carboxylic-acid ionophore containing several hydroxyl groups[16]. It forms alkaline-ion complexes that effect the transport of the ions across cell membranes via the "carrier" mechanism[17]. It has antimicrobial activity against gram-positive bacteria, mycobacteria, and fungi, and is widely used as an effective antibiotic in poultry and ruminants[18].

The box averaged approximately 33.3 Å per side. The temperature was maintained at 310 Kelvin (metabolic process temperature), with a τ_T of 0.1 psec; τ_P was 0.3 psec. A 10.0 Å residue-based nonbonded cutoff was used. The parameters for the potassium ion were those of Åqvist[19]; those of the ionophore and the methanol solvent

have been described in other publications[20,21]. The λ_1 state refers to the ion tightly bound state, λ_0 refers to the state where the ion has been driven furthest from the center of mass of the ionophore. The surrounding methanol solvents were allowed to relax around the kinetically frozen tightly bound λ_1 end-state Sal^-*K^+ molecule for 25 psec. All degrees of freedom of this system were then further relaxed for 300 psec prior to the perturbation calculations.

For the phenol to benzene perturbation the aromatic solute is solvated in a periodic box of water-saturated octanol solvent (98 octanols, 26 SPC waters), averaging approximately 30.2 Å per side. The phenol and benzene parameters were taken from Jorgensen[22], the octanol parameters are a modification of those presented by Jorgensen[23] as described in DeBolt[24]. The temperature was maintained at 310 Kelvin to encourage reorientation within the octanol media, with a τ_T of 0.2 psec; τ_P was 0.2 psec. A 10.0 Å residue-based nonbonded cutoff was used. The λ_1 state refers to the solvated phenol, the λ_0 state to the solvated benzene. The phenol hydroxyl group is mutated into a benzene hydrogen. Each end-state was relaxed for 200 psec prior to the perturbation calculations.

RESULTS AND DISCUSSION

Five progressively more challenging test cases are considered here. The results for the first three of the molecular systems demonstrate linear and about 1:1 (speedup:N_p) proportionality for equilibration and data collection phases of FEP in parallel. In these three cases ($Li^+ \rightarrow Cs^+$, ethane→ethane′, and the HIV protease-inhibitor mutation), a RAM and then short post-RAM equilibration for each λ-window is sufficient to prepare the system for data collection. MD driven sampling of phase space was apparently sufficient to produce $\Delta G_{parallel}$ values that agreed well with $\Delta G_{sequential}$ when runs of the same computation length are compared.

In the fourth system, the SAL^-*K^+ PMF, the decomplexation trajectories decribed by sequential and parallel processing are not identical. It is not possible to achieve complete phase-space sampling by either mode of computation. Therefore, sequential versus parallel PMF free energy profiles diverge as their trajectories diverge. We regard these two different trajectories as a subset of the many possible equally legitimate decomplexation routes.

In the fifth system, a phenol to benzene perturbation in water-saturated octanol, the speedup results were not 1:1 proportional relative to a sequential run. More extensive equilibration per window was required when this computation was done in parallel. There is an extemely slow λ-dependent component of configurational reorientation during equilibration. Because of octanol's inherently slow diffusional relaxation time, the extent of λ-dependent reconfiguration during the quick RAM was limited, in turn affecting the value of ΔG calculated. As an alternative approach, a relatively long

sequential run was used as an "extended RAM" to generate starting structures. Parallel FEP was then used to produce a more accurate and precise value for the perturbation free energy.

Perturbation of $Li^+ \rightarrow Cs^+$ (aqueous)

This is essentially a van der Waals cavity perturbation, with some rearrangement in the structure of the surrounding solvating waters. Experimental data for the $\Delta\Delta G$ of solvation for these ions was available[25]. The parameters used here in the AMBER potential energy function[3] are the same as those used in previous FEP calculations[19,26,27] which are in good agreement with the experimental data.

Several studies have been carried out with the Li^+/Cs^+ system. First, there is an investigation of the time required to equilibrate, τ_{eq}, a window at λ_K, where the λ_K's are taken progressively further from the starting state, λ_1, *without* the benefit of a RAM. We find the time spent equilibrating the λ_K windows is reduced when a FEP trajectory is divided into M blocks of N (parallel) windows. Second, the amount of time required to equilibrate a window *after* a RAM is examined. After a quick RAM and only a short per-window equilibration phase, the Li^+/Cs^+ system is well prepared for data collection. And finally, $\Delta\Delta G$ values generated via the sequential versus parallel methods are compared.

What happens when the initial end-point, λ_1, is used as the starting structure for all perturbation windows spanning the range $\lambda_1 \rightarrow \lambda_0$? Are windows at λ_K, far from the λ_1

starting state, poorly equilibrated? How much equilibration/relaxation is needed before data collection can begin? Figures 3 through 6 present plots of various quantities which are helpful in determining when an "equilibrated" state has been reached. The total energy of the system, the potential energy, the kinetic energy, and the overall volume of the constant pressure system are plotted with respect to time. The kinetic energy is the least relevant meter of equilibration, since a temperature bath is imposed on the system. The abscissa represents the time since initiation of MD at λ_K, starting from the λ_1-equilibrated starting state in all cases. Figure 3 is a "control" showing the type of oscillatory behavior produced by an equilibrated system; MD is initiated at $\lambda_{1.0}$ after having already pre-equilibrated at $\lambda_{1.0}$ for excess time.

Compare the range and frequency of oscillations (the ordinate) in Figure 3 ($\lambda_{1.0}$) with Figures 4, 5, and 6 which show the results of MD at $\lambda_{0.7}$, $\lambda_{0.5}$, and $\lambda_{0.15}$, respectively. A relatively fast relaxation occurs as indicated by the energy plots, within about half a picosecond at $\lambda_{0.7}$, a picosecond at $\lambda_{0.5}$, and about 4 psec at $\lambda_{0.15}$. Once the fast relaxation is completed, all λ_K have a range of fluctuations spanning about 50-60 kcal/mol in the total energy, about 80-90 kcal/mol in the potential energy, about 60-80 kcal/mol in the kinetic energy, and about 300-400 cubic Å in the volume. Of course the center of fluctuations is different when plots at different λ_K are compared. It is uncertain whether some slow component of relaxation is indicated at $\lambda_{0.7}$ and $\lambda_{0.5}$, but clearly this occurs at $\lambda_{0.15}$ between about 1.5 and 4.0 psec along the time scale in Figure 6. At $\lambda_{0.15}$ a volumetric *catastrophe* is recorded between 0.0 and 4.0 psec; this effect is also noticable to a lesser extent at $\lambda_{0.5}$. The drastic volumetric response is due to suddenly

inserting a cesium ion into a solvent cavity prearranged for a lithium ion. The system reacts by overexpanding at first, but gradually recuperates over a time proportional to the extent of the disorganization caused by the repulsive forces.

Figure 7 is a plot of the time taken to equilibrate, τ_{eq}, versus λ_K. The plotting-points were taken from a series of trajectories (not shown) like those presented in Figures 3-6. Note (Figure 7) the strongly concave upward (fitted) exponential curve and the linear region near the initial starting state at λ_1. This plot shows that when λ_K is far from the λ_1 starting state,

$$\tau_{eq}(\frac{1}{m}(1-\lambda_K)) < \frac{1}{m}\tau_{eq}(1-\lambda_K) \quad . \tag{6}$$

More simply, let us consider λ_0 as the starting state. Then

$$\tau_{eq}(\frac{1}{m}\lambda_K) < \frac{1}{m}\tau_{eq}(\lambda_K) \quad , \tag{7}$$

i.e. (for a λ_0 starting state) equilibrating the window at $\lambda_{0.5}$ requires less than 1/2th the time it takes to equilibrate the window at $\lambda_{1.0}$, or equilibrating the window at $\lambda_{0.1}$ requires less than 1/10th the time it takes to equilibrate the window at $\lambda_{1.0}$. This means that if the calculation is run in M blocks of span λ/M, where all N_P are applied to the N windows within each block, and each succeeding block uses the preceding block's last window as a starting configuration, it will take less equilibrium time to complete the FEP than when using a single end-point λ as the starting structure for all windows. An appropriately "blocked" calculation should be over an approximately linear region of

the τ_{eq}-versus-λ curve. By imbedding some sequential character into the otherwise parallel calculation, the concave upward curve is "bent toward linearity" with lowered slope. The curve could be made even flatter by applying the RAM method over the N_P windows within a block. "Blocking" a FEP trajectory will be worthwhile only when the full N_P can be usefully employed within a block. Such an instance is exemplified by the PMF calculation discussed below.

Because configurational reorientation is fast in the Li^+/Cs^+ calculation, effective parallelism can be achieved via the RAM method alone. In a RAM run, the goal is to "bend" the τ_{eq}-versus-λ_K curve to nearly zero slope. Each window should then require only brief equilibration. Figures 8, 9, and 10 present the results of equilibrating at λ_0 (the final endpoint window in the $\lambda_1 \rightarrow \lambda_0$ perturbation trajectory) after various length RAM runs of 8.0, 1.0, and 0.2 picoseconds respectively. The last of the manifold of states generated by these RAM runs was used as the starting state for each equilibration in Figures 8-10. Note the fluctuation widths in each of these plots are the same as those in the sections of previous plots (Figures 3-6) which describe equilibrated states. Volumetric catastrophe is entirely absent, no fast relaxation is evident, and it is uncertain whether any slow relaxations are occurring. Apparently, a RAM as rapid as 0.2 psec (100 steps total, with λ changing by 0.01 at each step) is sufficient to prepare even the window most distant from the λ_1-equilibrated end-state.

The RAM provides a set of N starting configurations so that each of N λ-windows can simultaneously enter into equilibration and then the data collection. When considering RAM'd parallel performance, the actual run-time for job completion scales as

$$\tau_{job} \ \alpha \ \frac{1}{N_P} \tag{8}$$

including both the equilibration and FEP collection phases. For FEP on a sequential machine:

$$\tau_{seq} = N_w(\tau_{eq} + \tau_{coll}) = N_w\tau_w \tag{9}$$

where τ_{seq} is the total time to complete a FEP job on a sequential machine, N_w is the number of λ-windows, τ_{eq} is the time spent in equilibration, τ_{coll} is the time spent collecting perturbation data. The factor ($\tau_{eq} + \tau_{coll}$) is abbreviated to τ_w, the net time spent computing per each window. For parallel computing, if we assume that $N_P = N_w$, and if the processors of the sequential and parallel machines are equal in speed, then for a parallel machine:

$$\tau_{par} = \tau_{RAM} + \frac{\tau_{seq}}{N_P} = \tau_{RAM} + \frac{N_w\tau_w}{N_P} = \tau_{RAM} + 1\tau_w \tag{10}$$

All of the (sequential) RAM runs were done on the Y-MP, so their relative contribution τ_{RAM} to τ_{par} is rather small. In actuality though, for the machines compared here, the parallel processors are much slower; a single nCUBE/2 processor runs AMBER's FEP module many times slower than the Cray Y-MP. The processor used in the Cray Y-MP runs the AMBER4 FEP program "Gibbs" fifty times faster than an individual processor of the nCUBE/2. This holds for all molecular systems treated in this paper except the HIV protease/inhibitor study for which the Y-MP runs "Gibbs" about thirty-eight times

faster than an nCUBE/2 processor.

Table I reports difference in free energy of (aqueous) solvation values, $\Delta\Delta G_{aq}(Li^+ \rightarrow Cs^+)$, calculated during sequential and parallel runs of varying duration. The Table presents four sets of data where sequential and parallel results are compared within the framework of identical equilibration and data collection times. $\Delta\Delta G$ of solvation is 50.3 kcal/mol as determined by experiment (by adding together several experimentally measured $\Delta\Delta G$ values)[25]; 54.5 kcal/mol from Åqvist's calculations[19], 56.3 from previous FEP calculations[27], and 56.4±0.6 is taken as the most converged value from this study (from 100 psec sequential/standard-window, FEP not reported in the Table).

The first two datasets present the $\Delta\Delta G$ for the $Li^+ \rightarrow Cs^+$ cation perturbations under sequential/slowgrowth and RAM/parallel/standard-window FEP regimes where no explicit per-window equilibration phase is carried out prior to data collection. Note firstly that the $\Delta\Delta G$ values are in reasonable agreement whether the λ-windows were evaluated sequentially or in parallel, especially in view of the reported errors. The trend toward $\Delta\Delta G$ of lower magnitude with increasing collection time is followed by both sets. Although most of the $\Delta\Delta G$ values in these two sets are within error of the converged value 56.4, the numerical agreement becomes acceptable (accurate *and* precise) after about 25 picoseconds of data collection. The sequential/slowgrowth (no-equil) run converges more rapidly but notice that at 50 psec collection time the parallel (no-equil) result is closer to 56.4. The error magnitudes are also similar, although the superposition errors of the parallel (no-equil) run are notably larger than for the

sequential. This may be due to the independence of each parallel λ-window (except for possible residual correlations from the RAM), whereas the adjacent sequential windows are strongly phase-space correlated due to their having decended from the specific trajectory followed during immediately preceding windows.

In the next two datasets of Table I, sequential/standard-window and RAM/parallel/standard-window, an equilibration of 25 picoseconds duration is spread over the windows, i.e. each of 50 windows undergoes 0.5 psec of dynamic relaxation prior to data collection. Again, the ΔΔG values and the associated errors are similar for sequential versus parallel regimes, and much more accurate now even with a small statistical sample (e.g. 0.1 psec collection). If any slow λ-dependent configurational relaxation is occurring, it does not affect the free energy numbers once minimal per-window equilibration is carried out. Overall then, considering this simple aqueous-cation FEP test case within the framework of some given equilibration and data collection times, we arrive at essentially comparable ΔΔG(solvation) values and errors, regardless of whether the calculation is carried out sequentially or in parallel.

As for sequential versus parallel performance, recall the Cray Y-MP runs the FEP program "Gibbs" fifty times faster than an individual processor of the nCUBE/2 for this molecular system. The 50 psec collection time (25 psec equilibrated) sequential/standard-window runs took 0.89 cpu-hours *net* on a single head Y-MP and 0.87 cpu-hours *per window* on the nCUBE/2. The RAM took 0.002 cpu-hours. Because all of the parallel runs reported in Table I use fifty simultaneous nCUBE processors, the parallel and sequential FEP computations were completed in the same net

run-time. This is 1:1 linear parallel speedup, and the method is scalable without limit so long as more λ-windows are needed.

Perturbation of X-methyl-CH_3 → CH_3-methyl-X (aqueous)

Figure 11 illustrates the "zero-sum" perturbation of ethane-to-ethane′. This system has the advantageous property that the net free energy change calculated should be exactly zero and is thus a good "calibrator" of methods. There is no guesswork about when convergence has been reached or inhibited by slow internal relaxations in the molecular system. Also, there is likely to be a greater necessity for rearrangement of solvent waters during the annihilation and creation of hydrophobic van der Waals cavities in this dual-methyl perturbation than in the $Li^+ \rightarrow Cs^+$ perturbation study. Our concern was to determine if slower internal relaxations (with respect to changes in λ) in this system will inhibit parallelization. Incidentally, a parallel run using the λ_1 endpoint starting configuration for all windows was attempted. It failed due to the catastrophic repulsive forces created by atomic overlaps between the "growing" methyl group and nearby solvents.

The RAM method was used to generate the results presented in Table II. The results for three different equilibration/collection time sets are listed; each parallel 50-window run used the 0.1 psec RAM states. There is marked numerical agreement between parallel and sequential runs, even with only (5 psec divided by 50 windows equals) 0.1 psec equilibration per independent window. Parallel and sequential approaches converge toward the correct zero value at the same rate with respect to increasing amounts of equilibration and data collection. Again the parallel calculations

were completed in the same run-time as the sequential runs.

Perturbation of Enzyme-bound HIV Protease Inhibitor

Because the binding modes of the peptidic inhibitors JG365 and its ethylenic mimetic are not radically dissimilar, the coarse-grain RAM/parallel approach is successfully applied. Table III reports the perturbation free energy for the process CO-NH $\rightarrow$ CH=CH for the inhibitor's fifth peptide bond. Note that these calculated ΔG values do not represent the overall difference in binding free energy for these two inhibitors; the difference in their solvation free energies was not calculated here. The overall $\Delta\Delta G_{binding}$ including the solvation component has been reported elsewhere as 7.2 kcal/mol[15].

The free energy values in Table III, both parallel and sequential, are the results of thermodynamic integration (TI) calculations, each using fifty windows per run. The first 45 psec parallel run was preceded by a 3 psec RAM to prepare initial starting states. Succeeding parallel runs (read across Table III from left to right) use the prior parallel run as their equilibration phase. The first 45 psec parallel run (ΔG(average) = 20.8±1.8) compares reasonably well with the sequential run (ΔG(average) = 22.5±1.7), in agreement within overlapping errors. After the second and third 45 psec parallel runs (ΔG(average) = 21.8±2.8 and 22.4±2.5), the parallel/sequential agreement is somewhat closer. In view of the hysteresis error-widths and the fact that internal enzyme motions and rearrangements can occur over timescales longer than the length of our simulations here, we would not claim that sampling is complete for either the sequential or parallel runs. The two methods generate similar but non-identical ensembles. The comparison

of sequential versus parallel free energy is offered only as a qualitative indicator of agreement in this case where there is no absolute yardstick (no known "correct" value) to measure directly against.

Each sequential HIV/JG365 FEP took about 12.0 hours net on a Cray Y-MP. The parallel calculations required only 9.1 hours per window. Because all fifty windows were calculated simultaneously, the parallel calculation completed in a shorter run-time (even when adding in the 0.7 hour RAM) than did the sequential runs. Due to the increased size of this system (nearly 4000 atoms) relative to the others studied here, parallel computational efficiency is increased; fifty nCUBE processors equivalenced to 1.3 Cray Y-MP's.

PMF of Salinomycin*K $\rightarrow$ Sal$^-$ •••• K$^+$ (in methanol)

In a PMF calculation the free energy as a function of some reaction coordinate $\Delta G(R)$ is computed. In this test simulation, a bound potassium cation is driven from the center of mass of the salinomycin ionophore through a distance of one angstrom. This short distance is adequate to demonstrate the parallel approach. The simulation is fragmented into a series of 100 moves, each of 0.01 Å, giving 100 λ-windows. Each individual point (each λ-window) along the reaction coordinate must be thoroughly equilibrated in a PMF calculation, regardless of whether the windows are treated sequentially or in parallel. Although sequential processing may be better for pre-equilibrating the windows, we find that a RAM generates a set of starting states which serve the purpose well.

Table IV reports sequential versus parallel ΔG(R) values. For the parallel run the equilibration time is gradually lengthened (from left to right in table IV) by restarting successive calculations from previously completed cycles. The time consuming sequential calculation was done only at the 260 psec total time level (TEQUIL=0.6/TCOLL=2.0 psec per window). 121.6 cpu-hours (five day's worth) were required on our local Hewlett Packard 730 workstation to complete the 1.0 Å PMF. A ten-window benchmark on the Cray Y-MP took 2.81 hours, which extrapolates to 28 Y-MP hours per angstrom. It would have cost 1680 allocation units to do the entire 1.0 Å PMF on the Y-MP at the San Diego Supercomputing Center. 100 nCUBE processors compute the 100-window 260 psec 1.0 Å PMF in 13.8 cpu-hours, twice as fast as the Y-MP, nearly nine-times as fast as the HP730, and actually completing in a bit longer than a half day of real-time.

Considering the first 0.5 Å "leg" of the PMF, the free energy values reported in Table IV compare well for the parallel and sequential runs in the first 260 picoseconds (labelled TEQUIL/TCOLL 0.6/2.0). No hysteresis error is available since the PMF was performed only in the $\lambda_1 \rightarrow \lambda_0$ (decomplexation) direction. The magnitude of the error at 260 psec might be estimated by subtracting the more converged ΔG(avg,0.0→0.5Å)=3.5 kcal/mol at about 1.5 nanoseconds (TEQUIL/TCOLL 16/2) from the value at 260 psec ΔG=4.5. This gives about 1.0 kcal/mol, supporting the interpretation that the parallel and sequential ΔG(avg,0.0→0.5)=4.4 values agree; they are close in absolute terms and the same within the estimated error. Apparently then, for this first 0.5 Å leg of the PMF, the parallel and sequential approaches reach compar-

able values within 260 psec (300 equilibration steps and 1000 collection steps per window).

Note the convenience and efficiency with which the accuracy of the PMF can be increased via the coarse-grain parallel approach. "Restart" configurations and velocities are written out for each parallel window at the termination of a FEP run. These restart files can serve as per-window starting states for a following parallel run where greater accuracy is the goal. No new equilibration is required to extend sampling; the run's entire trajectory can be devoted to data collection (200 psec per run in Table IV). The system continues to equilibrate during these successive data collection cycles. The results of all of these successive data collection cycles can then be pieced together and logarithmically-averaged to give a more accurate free energy from the broader sample. The data were not averaged that way here because we were interested in determining the amount of equilibration required for accuracy, since this is the factor that might potentially inhibit parallelism. We should note that with some effort the source code could be modified so the sequential method could also read/write restart files on a per-window basis; but this feature just "falls out" of the parallel method.

In the second leg of the PMF, the values of $\Delta G(avg,0.5\rightarrow1.0\text{Å})$ for the parallel and sequential runs differ by about 2 kcal/mol (2.7 versus 4.7) in the first 260 psec. For the parallel run this value stays fairly constant near 2.7 kcal/mol or thereabouts, even as the equilibration time accumulates across the table. Does the sequential trajectory steer into a region of configurational space that significantly differs from that carved out by the RAM? We looked (via computer graphics[28]) at configurations between 0.5 and 1.0

Å for sequential and parallel runs. A slightly different ionophore/ion configuration results between 0.5 and 1.0 Å due to an intra-ionophore hydrogen-bond exchange which occurs during RAM/parallel processing, but not during the more gradual sequential run.

This intra-ionophore hydrogen-bond exchange opens the pseudo-macrocyclic ring. In the closed circular form of salinomycin the "tail-end" hydroxyl group forms a (donor) hydrogen bond with the "head-end" carboxylate group; both oxygen containing groups participate as close K^+ ligands. In the open (or half-open) form this tail-end hydroxyl forms a new donor hydrogen bond with a mid-chain hydroxyl oxygen, leaving the carboxylate group and potassium cation tightly bound to each other and the solvent molecules; the tail-end hydroxyl no longer functions as a close liganding group. Thus the K^+ liganding oxygen groups assume a different orientation in the RAM/parallel trajectory than in the sequential trajectory. Naturally, the free energy profiles of the parallel and sequential simulations differ because their reaction coordinate trajectories (and end-states) are non-identical.

Because of the ionophore's very gradual conformational response to changes in λ during sequential processing, optimal non-covalent binding of the ion is maintained, even as the ion's distance from the ionophore's center of mass increases. This PMF reaction coordinate is somewhat artificial owing to the constraint of moving the ion along a vector emanating from the center of mass of the ionophore. The RAM more rapidly extracts the ion, before a conformational rearrangement of the ionophore that is most optimal for ion binding can fully occur. Once opened, the pseudo-ring does not reform during parallel equilibration. We will not attempt to judge which dynamic tra-

jectory is more "real". This highly flexible ion/ionophore/methanol system exemplifies the difficulty often experienced in obtaining complete sampling of complex systems, whether processing is sequential or parallel. Such multiple trajectories might even result from different starting structures, regardless of the methodology used. In fact, the ion may exit from a complexed state by any number of probable pathways. Note that the 1.0 Å trajectory generated by the RAM/parallel method results in a *lower free energy barrier* for the release process, indicating this trajectory is a more probable one. An argument could be made that if the goal *were* to determine a probability distribution for binding/release trajectories, the RAM/parallel method could be used by employing multiple starting structures. For these reasons, the multiple trajectory result should not be construed as a drawback or failing of the RAM method in this case.

Actually, there is a net preference for $SAL^{-}*K^{+}$ *binding* in methanol[29], a factor which may increase the difficulty of determining the "real" equilibrium binding/release trajectory(s) in this medium. Other investigators have presented data supporting the supposition that the pseudo-ring *does* break open during cation release[20]. We suspect that the intra-ionophore hydrogen-bond exchange may also occur (later) along the sequentially processed PMF trajectory. The exchange may occur when the ion has been pushed far enough away from the ionophore's center of mass yet clings preferentially to the carboxylate group. At that point the sequential- and parallel-generated system configurations would be more similar and the end-state free energies more comparable.

If system relaxation were allowed at the end of the first 0.5 Å RAM, would the hydrogen-bond exchange occur during the following RAM between 0.5 and 1.0 Å ?

Maybe. It would depend again on the starting structure (at 0.5 Å) and the solvent configuration as well. We tried this (a la the parallel "blocking" technique described previously) from one particular relaxed 0.5 Å (initially RAM'd) starting structure. As we expected, the ionophore remained hydrogen-bond-closed-circular. We can conclude that if this PMF study been carried out using the "blocking" technique, with 0.5 Å, 50 window blocks, the resultant parallel trajectory would have more closely resembled the sequentially processed trajectory. Such a parallel strategy might be followed in a case where details of the dynamic aspects of some specific slowly evolving trajectory are important.

Of course, the RAM/parallel PMF could be continued out to say 10 Å by scaling up to 1000 processors. This would give a net speedup to twenty times that of a Cray Y-MP even using the fifty-times slower nCUBE processors. A future parallel machine with 1000 YMP-speed processors could carry out a 10 Å PMF calculation in a couple of hours.

Perturbation of Phenol → Benzene (in octanol/water)

This calculation challenges the coarse-grain parallel methodology to the extent that the efficiency of parallelization is significantly reduced. There is a very slow relaxation time (hundreds of picoseconds) for reorientation of the solvent structure as λ changes. Phenol alligns to optimally hydrogen bond with polar groups (clustered hydroxyls) in the solvent media, whereas benzene optimally locates in a hydrophobic zone (octanol tails) of the media. Thus, a diffusional reorientation of the phenol/benzene and surrounding media must occur during the perturbation. Several

nanoseconds of data sampling could be required to cover enough states to render a fully converged free energy.

The sequential approach facilitates this calculation more readily since the gradual $\Delta\lambda$-dependent reconfiguration of the system must occur only once, spread over the windows of a sequential trajectory, instead of (at worst) N times, within each window of a parallel calculation. More time spent equilibrating per parallel window can alleviate the problem, while sacrificing parallel efficiency. A longer RAM time would serve the same purpose. For this system a sequential simulation of inadequate accuracy could serve to generate moderately well equilibrated initial starting states for a parallel calculation with a mission of higher accuracy. A 100 psec RAM time was chosen.

Table V reports the results of sequential versus parallel FEP runs, the difference in free energy and hysteresis error as a function of the number of steps of equilibration and collection. In view of the slowly relaxing solvent media, we do not expect the 400 picosecond sequential ΔG value of 3.6 kcal/mol to be completely converged. This expectation is born out given the relatively large hysteresis error, $\pm$ 1.0 kcal/mol. In Table V, parallel runs labeled 1/4, 4/4, 8/4, and 12/4 were started from the 100 psec "quick RAM" listed as the 0/1 sequential run. Parallel run 4S/52 was started from the "extended RAM" listed as the 2/2 sequential run. It is encouraging that the parallel "quick RAM" difference free energies (2.7, 3.1, 3.3, 3.4) are moving in the direction of the sequential value (3.6 kcal/mol). The "quick RAM" parallel results compare best after the fourth 400 picosecond run.

In the 100-window 100-processor "quick RAM" case, the parallel machine actually completes a 400 psec job in half the clock-time taken by the sequential Y-MP. Four 400 psec runs on the parallel machine, plus the RAM time, sums to a bit more than twice the run-time of the sequential machine to arrive at comparable numbers for this system. Based on the convergence trend indicated in Table V, we would expect the parallel/sequential free energies and hysteresis to compare better after six parallel runs. Thus, we anticipate that three times the run-time of the Y-MP would be required to bring about close parallel/sequential agreement. While this may not appear to be a cause for celebration, when we consider the actual dollar cost efficiency of using parallel versus sequential cpu cycles, even in this difficult case the parallel machine still "wins". Beginning from the "extended RAM", parallel computation was continued for 2.6 nanoseconds. At this point, sufficient accuracy and precision in the ΔG(phenol→benzene) value (3.8 ± 0.4) was obtained for our purposes.

This is apparently the first time the octanol solvation contribution to an octanol/water partition coefficient $P_{o/w}$ has been calculated using explicitly simulated water-saturated octanol media. We also calculated the difference in *aqueous* solvation free energies for phenol versus benzene (not reported in Table V). The aqueous (SPC water model) phenol→benzene perturbation was done as a fifty-window FEP calculation, 100 psec (40/60 equil/collect) in each of forward and reverse directions (after 20 psec end-state relaxations), at a constant temperature of 310K, using a 10 Å nonbonded cutoff, 2.0 fsec stepsize, updating the pairlist each 20 MD steps. The result was ($*DG_{aq}$(phenol→benzene) = 4.89±0.20 kcal/mol. This compares well with the value

obtained by Jorgensen for the similar TIP3P water model, 5.2 ± 0.2 [22]. Subtracting our most converged value for $\Delta\Delta G_{oct}$(phenol→benzene) gives 4.89 − 3.64 = 1.09±0.55 kcal/mol for the difference in solvation free energies of phenol versus benzene as solute in water-saturated octanol versus pure water. Converting this number through the equation

$$\Delta\Delta G_{oct/wat}(solvation) = -2.303\, R\, T\, \Delta Log(P_{o/w}) \qquad (11)$$

where $P_{o/w}$ is the octanol-water partition coefficient, gives $\Delta Log(P_{o/w})$ = 0.77±0.39 for for the relative partition coefficients of benzene versus phenol. From experiment, $\Delta Log(P_{o/w})$ benzene-phenol is 0.67 kcal/mol[30].

The experimental octanol/water partition coefficient is situated well within the calculated range. The trend apparent (from left to right) in table V is one of increasing magnitude in the free energy and decreasing associated error. As further MD sampling continues, the λ-hybrids of the phenol/benzene solute have the opportunity to more optimally locate and orient within the heterogenous binary media. ΔG_{oct} may increase to a number slightly greater than 3.8 at full convergence. A slight increase in ΔG_{oct} would give a decrease in the calculated $\Delta Log(P_{o/w})$. If this trend proves out, a more completely converged MD-calculated $\Delta Log(P_{o/w})$ would be brought into even closer agreement with experiment, with reduced relative error. Further "explicit" octanol/water/solute studies are in progress[24].

SUMMARY AND CONCLUSIONS

FEP calculations were carried out for five molecular systems, three of them demonstrating parallel speedup about 1:1 linearly proportional to the number of parallel processors applied. We have shown that enzyme/ligand binding free energies can be efficiently calculated with parallel computers. Parallel and sequentially processed PMF simulations gave different (albeit equally legitimate) trajectories. The parallel efficiency was somewhat lower for the most challenging molecular system due to an extremely slow λ-dependent relaxation component. An extended parallel FEP calculation of the relative octanol/water partition coefficients of phenol and benzene gave excellent agreement with experiment.

We have also discussed the "equilibration dilemma", the possible inhibition of efficient parallelism when a perturbing system undergoes configurational reorientation following the successive changes in λ. Two methods are offered as useful ways to overcome the equilibration dilemma for non-drastic cases: the "blocking" method (breaking an MD/FEP trajectory into several smaller blocks each of which uses a preceeding block's end-state as its pre-equilibrated initiation state) and the "RAM" method (generating a Rapidly Accessible Manifold of starting states for independent processing at each value of λ by periodically writing out structures and velocities during a very quick perturbation). The following criteria should be assessed when deciding whether a particular molecular system is a good candidate for coarse grain parallel FEP.

The coarse grain parallel approach is most efficient when applied to systems for which one of the following is true:

1) No *slow* configurational reorientation occurs as a direct consequence of changes in λ. Any relaxation is fast and adequately completed during a short equilibration phase at each λ-point in each independent parallel window. The calculation of $\Delta\Delta G_{aqueous}(Li^+ \rightarrow Cs^+)$ is in this catagory.

2) Any configurational λ-dependence is adequately relaxed during a RAM plus a short equilibration phase. The zero-sum aqueous ethane → ethane´ calculation is a case of this type.

3) A λ-dependent configurational change can be "driven", as in a coordinate-driven potential of mean force (PMF) calculation. The PMF reorientation can be driven during a RAM, after which each λ-window is efficiently equilibrated and data-sampled in coarse-grain parallel. Such a calculation efficiently divides into λ-"blocks". The decomplexation PMF of SAL^-*K^+ is an example of this case as well as of case 4 directly following.

4) There is a diversity of many configuration states which must be sampled during the data collection phase in order to get a representative boltzman distribution and thus reliable free energy. And this broad sampling is required at each λ-point, apart from relaxations which occur with respect to changes in λ. The HIV protease/inhibitor FEP provides such an example.

5) It is desired that a system will *not* drift from a pre-established starting state with changes in λ, e.g. limiting the divergence of a protein from a known crystal structure. This might reduce the need for structural restraints which are sometimes used in enzyme/ligand and DNA/ligand calculations.

The coarse grain parallel approach is expected to be least efficient when applied to systems for which the following is true:

6) A very slow reorientation of system components occurs as a direct consequence of the changes in the value of λ. In this case the molecular configuration at each succeeding λ-point would be nearest "equilibrium" when the system has been prepared via gradual relaxation during preceeding λ-points. This type of system is best suited for sequential processing. Cost efficient coarse grain parallel processing may be possible if the slow reorientation is completed during a RAM of adequately long duration, and/or when longer per-window equilibration times are used. $\Delta\Delta G_{octanol}$(phenol→benzene) is a calculation of this type.

A move to high-performance parallel supercomputers is currently being touted by U.S. federal science officials as the best way to reach the scientific goals known as the Grand Challenges. We believe the FEP method presented here is a first glimpse of the exciting uses for parallel supercomputers in the modeling of chemical interactions. Let us consider the speed of FEP calculations in the near future. Computer manufacturers will soon be marketing parallel machines with individual processor speeds approaching that of a Cray Y-MP or better. Bearing in mind the rate at which computer hardware decreases in price, even as chip performance increases, we can easily foresee stringing

together 100-1000 YMP-equivalent chips in parallel. With continued improvement of interprocessor communication and data retrieval rates, fine-grain parallel MD will be useful for equilibrating molecular systems at the λ_1 and λ_0 FEP endpoints[1a]. And this study has shown that many FEP calculations can be carried out very efficiently in coarse-grain parallel. Next, a combination fine-grain/coarse-grain approach to FEP calculations, where each independent λ-window may employ multiple processors, will yield yet another order of magnitude increase in speed.

We are lead to the following conclusion: *free energy perturbation calculations involving macromolecular systems will soon take seconds to minutes on parallel supercomputers rather than the hours to days required on current-day sequential machines.* This promises to be of great value to computational chemists in modeling drug/receptor binding, in peptide and protein design, and in general where there is a question of how the molecular level interactions of various interacting groups mediate the free energy differences in biochemical systems.

REFERENCES FOR CHAPTER IV

(1) (a) S.E. DeBolt, P.A. Kollman, *J. Comput. Chem.*, **14**, 312 (1993). (b) W. Smith *Theo. Chim. Acta*, **84**, 385 (1993). *Comput. Phys. Comm.*, **67**, 392 (1992). *Comput. Phys. Comm.*, **62**, 229 (1991). H. Schrieber, O. Steinhauser, P. Schuster, *Parallel Computing*, **18**, 557 (1992). M.A. Shifman, A. Windemuth, K. Schulten, P.L. Miller, *Comput. and Biomed. Res.*, **25**, 168 (1992). H. Sato, Y. Tanaka, T. Yao, *Fujitsu Scientific & Tech. J.*, **28**, 98 (1992). J.F. Janak, P.C. Pattnaik, *J. Comput. Chem.*, **13**, 1098 (1992). J.E. Mertz, D. J. Tobias, C.L. Brooks III, U.C. Singh, *J. Comput. Chem.*, **12**, 1270 (1991). R.D. Skeel, *J. Comput. Chem.*, **12**, 175 (1991). T.W. Clark, J.A. McCammon, *Computers Chem.*, **14**, 219 (1990). G.C. Fox, M.A. Johnson, G.A. Lyzenga, S.W. Otto, J.K. Salmon, D.W. Walker, in *Solving Problems On Concurrent Processors. General Techniques And Regular Problems*, **1**, chapts. 9 and 16, Prentice-Hall Inc., Englewood Cliffs, New Jersey, ISBN 0-13-823022-6, (1988). D. Fincham, *Molecular Simulation*, **1**, 1 (1987). H.L. Nguyen, H. Khanmohammadbaigi, E. Clementi, *J. Comput. Chem.*, **6**, 634 (1985). (c) Machines such as: nCUBE/3, Intel Paragon, Cray T3D, Fujitsu AP-1000, Kendall Square Research KSR-1, and others.

(2) D.L. Beveridge, F.M. Di Capua, *Annu Rev. Biophys. Biphys. Chem.*, **18**, 431 (1989).

(3) S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A Case, *J. Comp. Chem.*, **7**, 230 (1986). U.C. Singh, P.K. Weiner, J.W. Caldwell, P.A. Kollman, **AMBER (UCSF)**, version 3.0, Department of Pharmaceutical Chemistry, University of California, San Francisco, (1986). In the AMBER potential function H_{Total} is the potential energy of the system; K_b and b_{eq} are the bond stretching constant and equilibrium bond distance; K_Ω and Ω_{eq} are the bond angle stretching constant and equilibrium bond angle; V_n, n, and γ are the torsional force constant, periodicity of the torsional term, and the phase angle; r_{ij}^* and ε_{ij} are the (Lennard-Jones) nonbond interaction parameters for particles i and j; r_{ij} is the interatomic distance between centers i and j; q_i and q_j are the atomic partial charges on i and j; ε is the effective dielectric constant.

(4) D.A. Pearlman, P.A. Kollman, *J. Chem. Phys.*, **90**, 2460 (1989).

(5) W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes*, Cambridge University Press, New York (1989).

(6) D.J. Tobias, C.L. Brooks, *J. Chem. Phys.*, **89**, 5115 (1988).

(7) D.A. Pearlman, P.A. Kollman, *J. Chem. Phys.*, **94**, 4532 (1991).

(8) D.A Pearlman, D.A. Case, J.A. Caldwell, G.L. Seibel, U.C. Singh, P. Weiner, P.A. Kollman, *AMBER 4.0*, University of California, San Francisco (1991).

(9) W.F. van Gunsteren, H.J.C. Berendsen, *Mol. Phys.*, **34**, 1311 (1977). J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, *J. Comput. Phys.*, **34**, 327 (1977). G. Ciccotti, M. Ferrario, J.P. Ryckaert, *Mol. Phys.*, **47**, 1253 (1982).

(10) H.J.C. Berendsen, J.P.M. Postma, W. F. van Gunsteren, A. DiNola, J.R. Haak, *J. Chem. Phys.*, **81**, 3684 (1984).

(11) H.J.C. Berendsen, J.P.M. Postma, W. van Gunsteren, J. Hermans, in *Intermolecular Forces*, 331-342; B. Pullman, ed., Reidel: Dordrecht, The Netherlands (1981).

(12) D.A. Pearlman, P.A. Kollman, *J. Chem. Phys.*, **94**, 4532 (1993).

(13) A. Wlodawer, M. Miller, M. Jaskolski, B.K. Sathyanarayana, E. Baldwin, I.T. Weber, L.M. Selk, L. Clawson, J. Schneider, S.B.H. Kent, *Science*, **245**, 616 (1989).

(14) D.H. Rich, J. Green, M.V. Toth, G. Marshall, S.B.H. Kent, *J.Med.Chem.*, **33**, 1285 (1990).

(15) P. Cieplak, P.A. Kollman, *J. Computer-Aided Mol. Design*, **7**, paper in press (1993).

(16) H. Kinashi, N. Otake, H. Yonehara, S. Sato, Y. Saito, *Tetrahedron Let.*, **49**, 4955 (1973).

(17) B.C. Pressman, *Federation. Proc.*, **27**, *1283 (1968).*

(18) Y. Miyazaki, M. Shibuya, H. Sugawara, O. Kawaguchi, C. Hirose, J. Nagatsu, S. Esumi, *J. Antibiot.*, **27**, *814 (1974).*

(19) J. Aqvist, *J. Phys. Chem.*, **94**, 8021 (1990).

(20) B. Caughey, G. R. Painter, A. F. Drake, W. A. Gibbons, *Biochim. Biophys. Acta*, **854**, 109 (1986); and M. J. O. Anteunis, N. A. Rodios, *Bull. Soc. Belg*, **90**, 715 (1981).

(21) S.E. DeBolt, P.A. Kollman, *J. Amer. Chem. Soc.*, **112**, 7515 (1990).

(22) W.L. Jorgensen, T.B. Nguyen, *J. Comp. Chem.*, **14**, 195 (1993).

(23) W.L. Jorgensen, *J. Phys. Chem.*, **90**, 1276 (1986).

(24) S.E. DeBolt, P.A. Kollman, manuscript in progress.

(25) M.A. Burgess, *Metal Ions In Solution*, Ellis Horwood: Chichester, England (1978).

(26) G. King, A. Warshel, *J. Chem. Phys.*, **91**, 3647 (1989).

(27) C.I. Bayly, P.A. Kollman, *J. Amer. Chem. Soc.*, submitted for publication (1993).

(28) T.E. Ferrin *et.al.*, "The MIDAS Display System", *J. Mol. Graphics*, **6**, 13 (1988).

(29) B. Caughey, G.R. Painter, A.F. Drake, W.A. Gibbons, *Biochim. Biophys. Acta*, **854**, 109 (1986).

(30) C. Hansch, A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley-Interscience, New York (1979).

Table I. The difference in solvation free energy $\Delta\Delta G_{aq}(Li^+ \rightarrow Cs^+)$ as a function of total collection time.*

TOTAL COLLECTION TIME (psec)	0.1	0.2	1.0	10.0	25.0	50.0
Sequential / Slowgrowth (no equil)						
ΔΔG(average)	74.7	61.0	54.2	57.2	-	57.0
hysteresis error	52.8	24.1	10.3	1.5	-	0.1
superposition error	7.4	2.4	0.4	0.1	-	0.0
Parallel / Standard-Window (no equil)						
ΔΔG(average)	76.8	73.9	63.4	58.1	56.3	56.6
hysteresis error	50.8	46.1	18.8	0.3	1.4	0.3
superposition error	7.7	5.7	9.9	8.3	4.7	1.6
Sequential / Standard-Window (equil 25 psec)						
ΔΔG(average)	55.7	-	56.4	57.1	56.6	56.5
hysteresis error	5.0	-	0.4	0.7	0.2	0.1
superposition error	4.4	-	3.7	1.4	0.6	0.4
Parallel / Standard-Window (equil 25 psec)						
ΔΔG(average)	55.8	-	55.1	55.2	56.6	56.7
hysteresis error	0.7	-	0.7	0.6	0.4	0.4
superposition error	4.4	-	3.8	0.8	0.3	0.5

* All standard-window FEP runs used 50 lambda-windows. The TOTAL COLLECTION TIME for each calculation was divided evenly over all 50 windows. The term "no equil" indicates that there was no equilibration prior to data collection (which is always the case for slow-growth runs). ΔΔG(average) is the average of the absolute values of the forward $\lambda_{1\rightarrow 0}$ and reverse $\lambda_{0\rightarrow 1}$ runs, in kcal/mol. The hysteresis and superposition errors are defined in the text Background section. Sequential runs were on a Cray Y-MP, parallel runs on an nCUBE/2. Parallel runs were preceded by the 0.1 psec RAM on a Cray Y-MP reported above under sequential standard-window (no-equil).

Table II. The difference in solvation free energy $\Delta\Delta G_{aq}$(X-methyl-CH_3 $\rightarrow$ CH_3-methyl-X) as a function of per-window equilibration and collection times.*

TEQUIL/TCOLL TIMES (psec/w)	0.1/0.1	0.5/0.5	1.0/2.0
Parallel / Standard-Window			
$\Delta\Delta G$(average)	0.35	0.06	0.04
Sequential / Standard-Window			
$\Delta\Delta G$(average)	0.31	0.07	0.04

* All standard-window FEP runs used 50 lambda-windows. X represents a ghosted methyl group. TEQUIL is the per-window equilibration time; TCOLL is per-window collection time. No hysteresis error is reported as FEP was run only in the $\lambda_1 \rightarrow \lambda_0$ direction due to the fact that the forward and reverse runs are essentially identical for zero-sum calculations. $\Delta\Delta G$(average) is the average over the absolute values of adjacent window pairs $\lambda_K+\Delta\lambda$ and $\lambda_{K+1}-\Delta\lambda$ from double-wide sampling. The difference between $\Delta\Delta G$(average) and 0.0 provides an accurate estimate of the error in the simulation. Sequential runs were on a Cray Y-MP, parallel runs on an nCUBE/2. Parallel runs were preceded by a 0.1 psec RAM on a Cray YMP; $\Delta\Delta G$(average) for the RAM was 2.3 kcal/mol.

Table III. Free energy (TI) of HIV-1 protease inhibitor JG365 ΔG(CO-NH → CH=CH) as a function of per-window equilibration and collection times.*

TEQUIL/TCOLL TIMES (psec/w)	0.3/0.6	0.9/0.6	1.5/0.6
Parallel / TI			
ΔG(average)	20.8	21.8	22.4
hysteresis error	1.8	2.8	2.5
ΔG(forward)	22.5	24.6	24.9
ΔG(reverse)	19.0	18.9	19.9
Sequential / TI			
ΔG(average)	22.5		
hysteresis error	1.7		
ΔG(forward)	24.2		
ΔG(reverse)	20.8		

* All runs were via thermodynamic integration (TI) using 50 lambda-windows. TEQUIL is the per-window equilibration time; TCOLL is per-window collection time. ΔG(forward) and ΔG(reverse) are the absolute values in the $\lambda_{1\to0}$ and $\lambda_{0\to1}$ directions. ΔG(average) is the average of forward and reverse runs. Sequential runs were on a Cray Y-MP, parallel runs on an nCUBE/2. Parallel runs were preceded by a 3.0 psec RAM on a Cray Y-MP; ΔG(average) for the RAM was 20.0 ± 0.6 kcal/mol.

Table IV. Potential of mean force $\Delta G_{MeOH}(Sal^*K \rightarrow Sal^-...K^+)$ as a function of per-window equilibration and collection times.*

TEQUIL/TCOLL (psec/w)	0.6/2.0	2.6/1.4	4/2	6/2	8/2	10/2	12/2	14/2	16/2
Parallel PMF / Standard-Window									
ΔG(avg, 0.0→0.5Å)	4.5	4.4	4.2	3.8	3.8	3.8	3.5	3.6	3.5
ΔG(avg, 0.5→1.0Å)	2.8	3.0	2.8	2.7	2.8	2.7	2.5	2.6	2.7
ΔG(avg, 0.0→1.0Å)	7.3	7.4	7.0	6.6	6.6	6.5	6.0	6.2	6.2
Sequential PMF / Standard-Window									
ΔG(avg, 0.0→0.5Å)	4.4								
ΔG(avg, 0.5→1.0Å)	4.7								
ΔG(avg, 0.0→1.0Å)	9.1								

* FEP runs used a total 100 lambda-windows to complete 1.0Å PMF, (50 windows to complete the first 0.5Å PMF). TEQUIL and TCOLL are per-window values as in Table II. ΔG(avg) is the average over the absolute values of $\lambda_K+\Delta\lambda$ and $\lambda_{K+1}-\Delta\lambda$ adjacent window pairs, in kcal/mol. Superposition error was between 0.05 and 0.20 for all runs. No hysteresis error is reported for this unidirectional PMF calculation. Sequential runs were on HP730, parallel runs on nCUBE/2. The first parallel run (0.6 TEQUIL / 2.0 TCOLL) was preceded by a RAM for which ΔG(avg, 0.0→1.0Å)=7.7 (20 psec RAM on Y-MP), and ΔG(avg, 0.0→0.5Å)=4.6 kcal/mol (10 psec RAM on Y-MP). Successive parallel runs used the successively accumulating parallel data collections as their MD equilibration.

Table V. The difference in solvation free energy ΔG_{Oct}(phenol → benzene) as a function of per-window equilibration and collection times.[a]

TEQUIL/TCOLL TIMES (psec/w)	1/4	4/4	8/4	12/4	4S/52[b]
Parallel / Standard-Window					
ΔG(average)	2.7	3.1	3.3	3.4	3.8
hysteresis error	1.5	1.5	1.4	1.3	0.4
ΔG(forward)	1.2	1.5	1.9	2.2	3.4
ΔG(reverse)	4.2	4.6	4.6	4.7	4.1

TEQUIL/TCOLL TIMES (psec)	0/1	2/2
Sequential / Standard-Window		
ΔG(average)	2.6	3.6
hysteresis error	1.5	1.0
ΔG(forward)	1.1	2.7
ΔG(reverse)	4.2	4.6

[a] Standard-window FEP runs used 100 λ-windows. 4 psec/window thus gives a production run of 400 psec total. TEQUIL and TCOLL are per-window values as in Table II. ΔG(average) is the average over the absolute values of forward $\lambda_{1\to0}$ and reverse $\lambda_{0\to1}$ runs, in kcal/mol. Only hysteresis error is reported; superposition error was insignificant. Sequential runs were on Cray Y-MP, parallel runs on nCUBE/2. The first parallel run was preceded by the 100 psec sequential RAM reported above (TEQUIL=0/TCOLL=1); the remaining parallel runs used the successively accumulating parallel data collections as their MD equilibration.

[b] The denotation "4S" means 4 psec/window done sequentially; the run listed as 2/2 under the sequential/standard-window heading was used as a long RAM for this slowly configuring system. Then 50 parallel windows were started from coordinates saved at the appropriate intervals along the 2/2 sequential trajectory, and FEP was carried out in parallel over 52 psec per window for a total of 50 × 52 = 2.6 nanosecond to get a converged and precise value for ΔG.

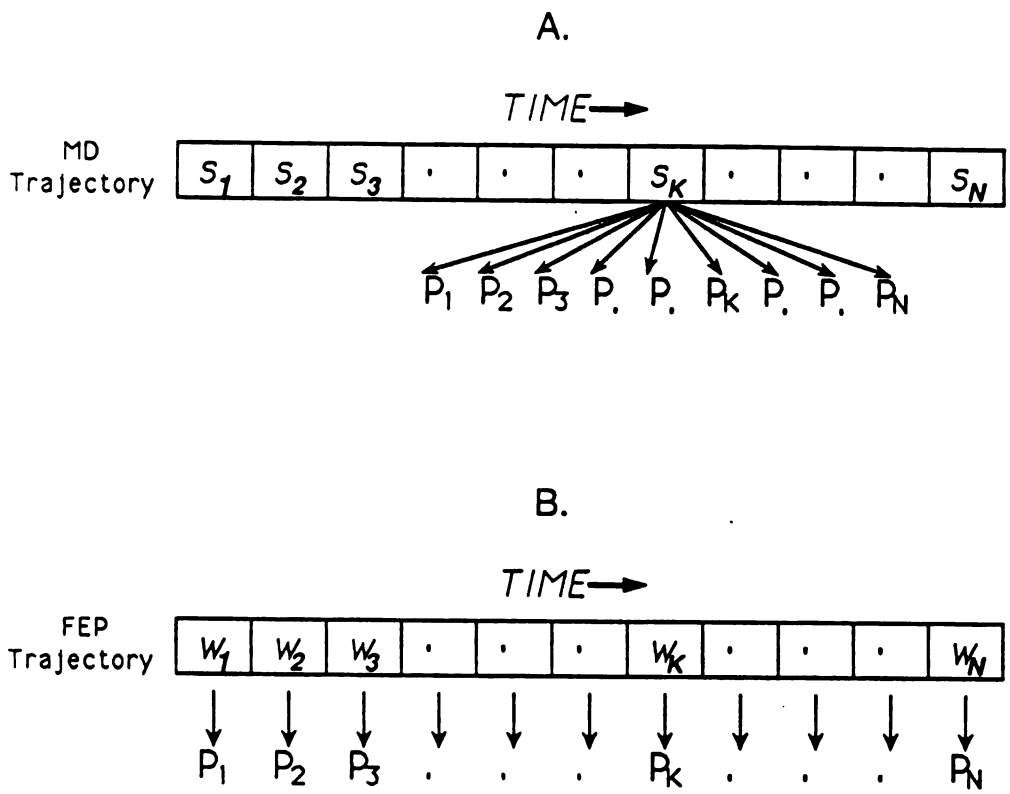


Figure 1. (A) Schematic of individual MD time-steps (S_1, S_2,....) in a trajectory. Each time-step must be decomposed into subtasks which are distributed over multiple processors (P_1, P_2,....). No work on the S_Kth step can proceed until the S_{K-1}th step is completed. The bars separating each MD step represent the data regathering stage which can result in an interprocessor communications bottleneck. (B) Schematic of individual FEP windows (W_1, W_2,....). The work in each lambda-window is accomplished independently of other windows (no waiting), one window to a processor in this illustration. No interprocessor communication.

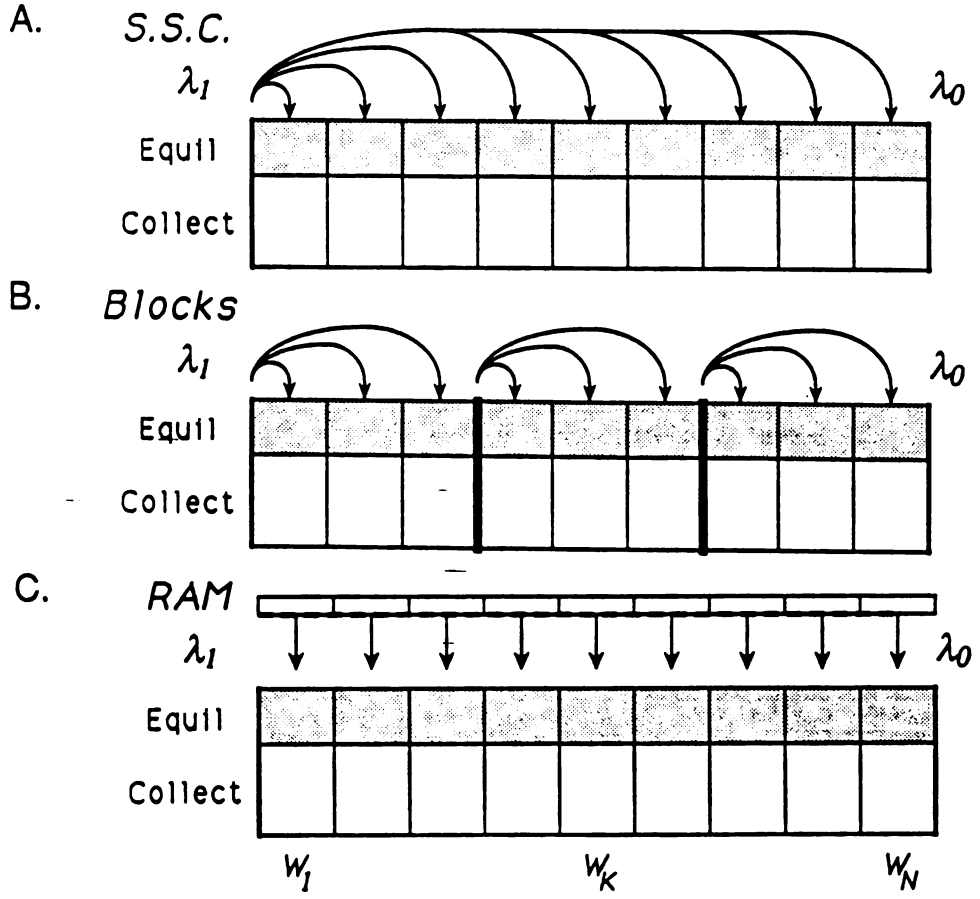


Figure 2. Three schemes for generating starting states to facilitate simultaneous equilibration and collection of FEP data. Boxes from left to right represent independent lambda-windows $\lambda_1 \rightarrow \lambda_0$, with each window's equilibration phase shaded. (A) In the single starting configuration (SSC) approach, the λ_1-endpoint equilibrated state is used as the initial configuration for all N windows. (B) For purposes of illustration the block method is presented as if N_p=3 total processors were available. Three blocks are shown, divided by heavy bar lines, with all three processors operating in a block. The equilibrated endpoint configuration of the final window in a block serves as a better starting state for the windows of the succeeding block than would the λ_1 state. (C) A rapidly accessible manifold (RAM) of starting states is generated when a FEP Hamiltonian function is mutated quickly during MD. $\lambda_1 \rightarrow \lambda_0$ is traversed in a time which is not expected to give an accurate ΔG value, but sufficient to prepare the states so volumetric and energetic catastrophe are avoided. Also, some degree of lambda-dependent configurational reorientation will be completed during the RAM, prior to each window's equilibration phase.

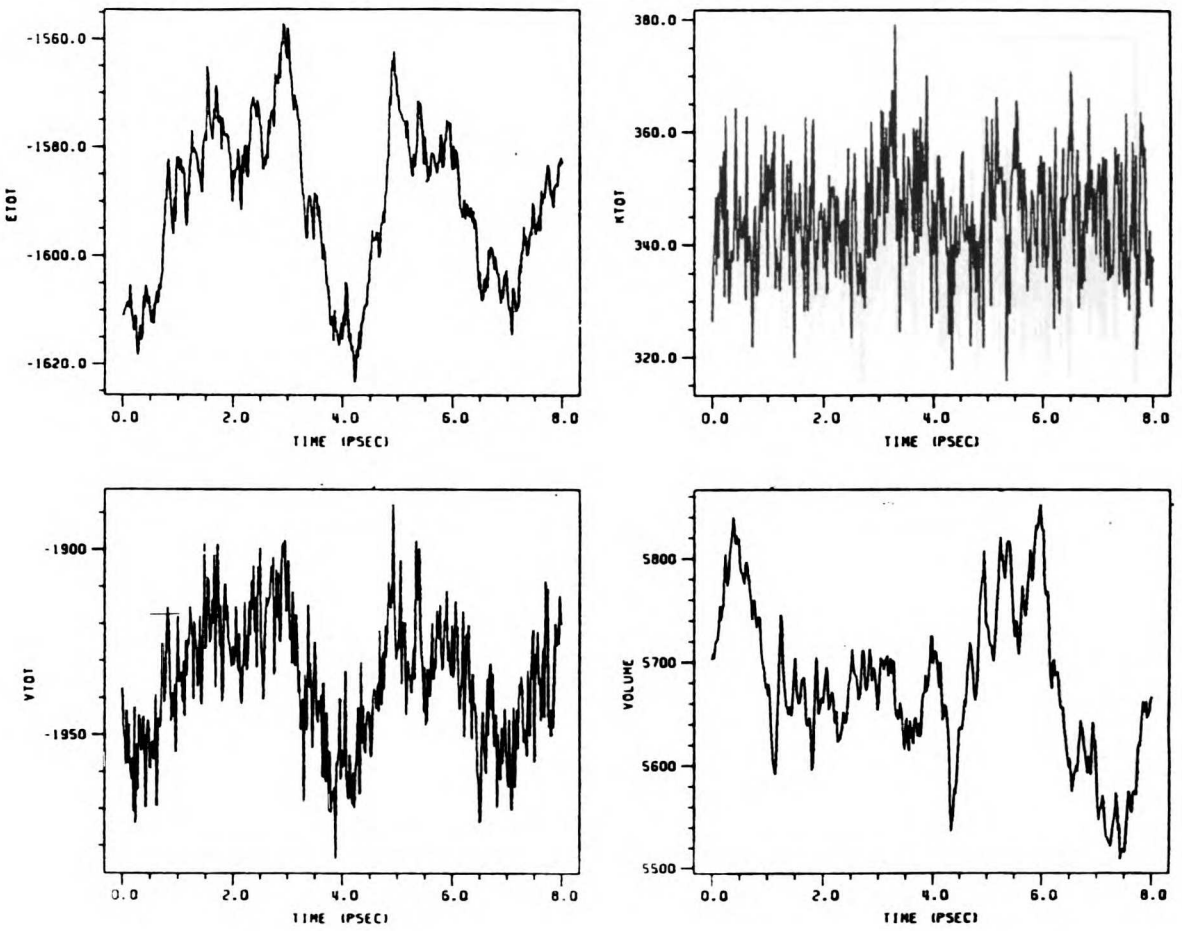


Figure 3. Equilibrium fluctuations of quantities versus time for MD at lambda = 1.0 after pre-equilibrating for excess time at lambda = 1.0. ETOT is the total energy of the aqueous lithium/cesium system. VTOT is the potential energy contribution. KTOT is the kinetic energy. VOLUME is the overall volume of the periodic system.

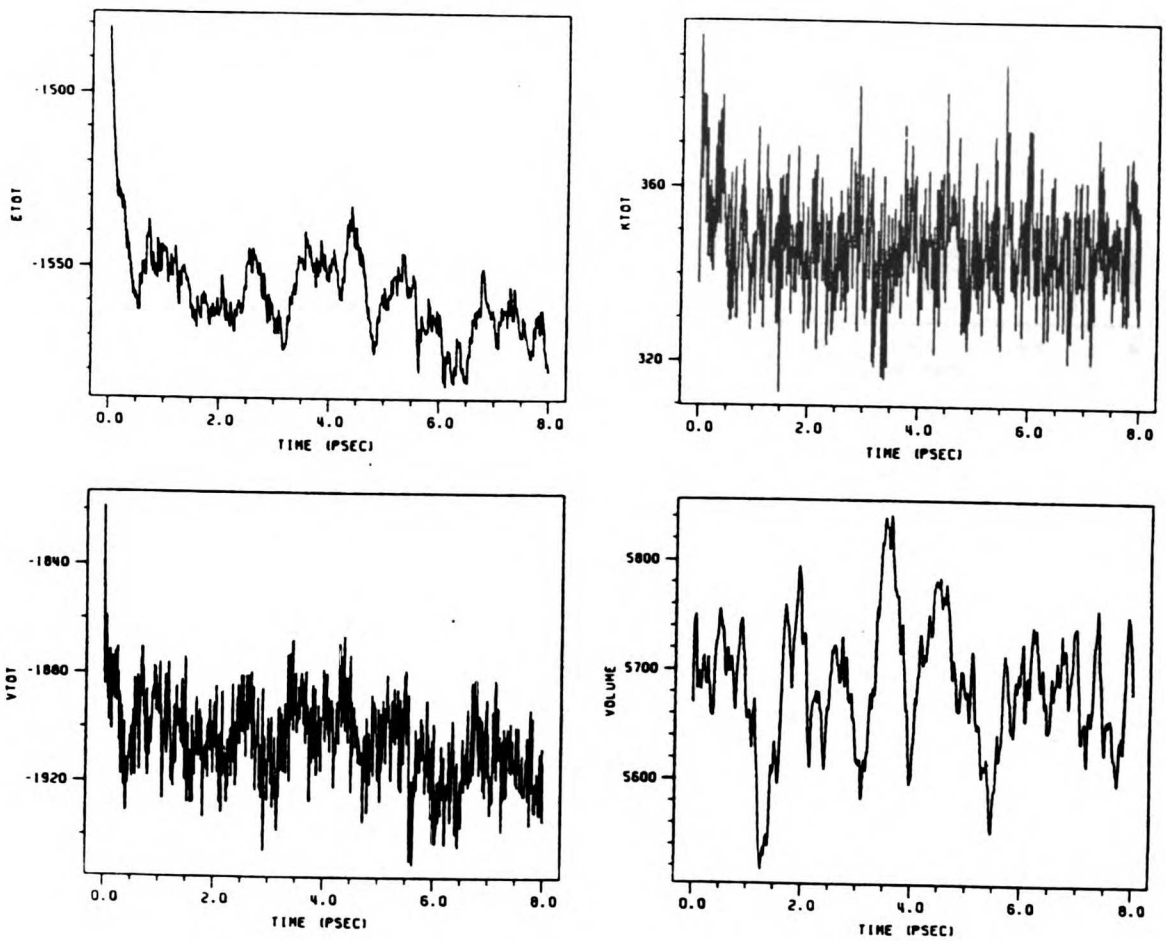


Figure 4. Relaxing fluctuations of quantities (as in fig. 3) for MD at lambda = 0.7 using the pre-equilibrated $\lambda_{1.0}$ configuration as the starting structure.

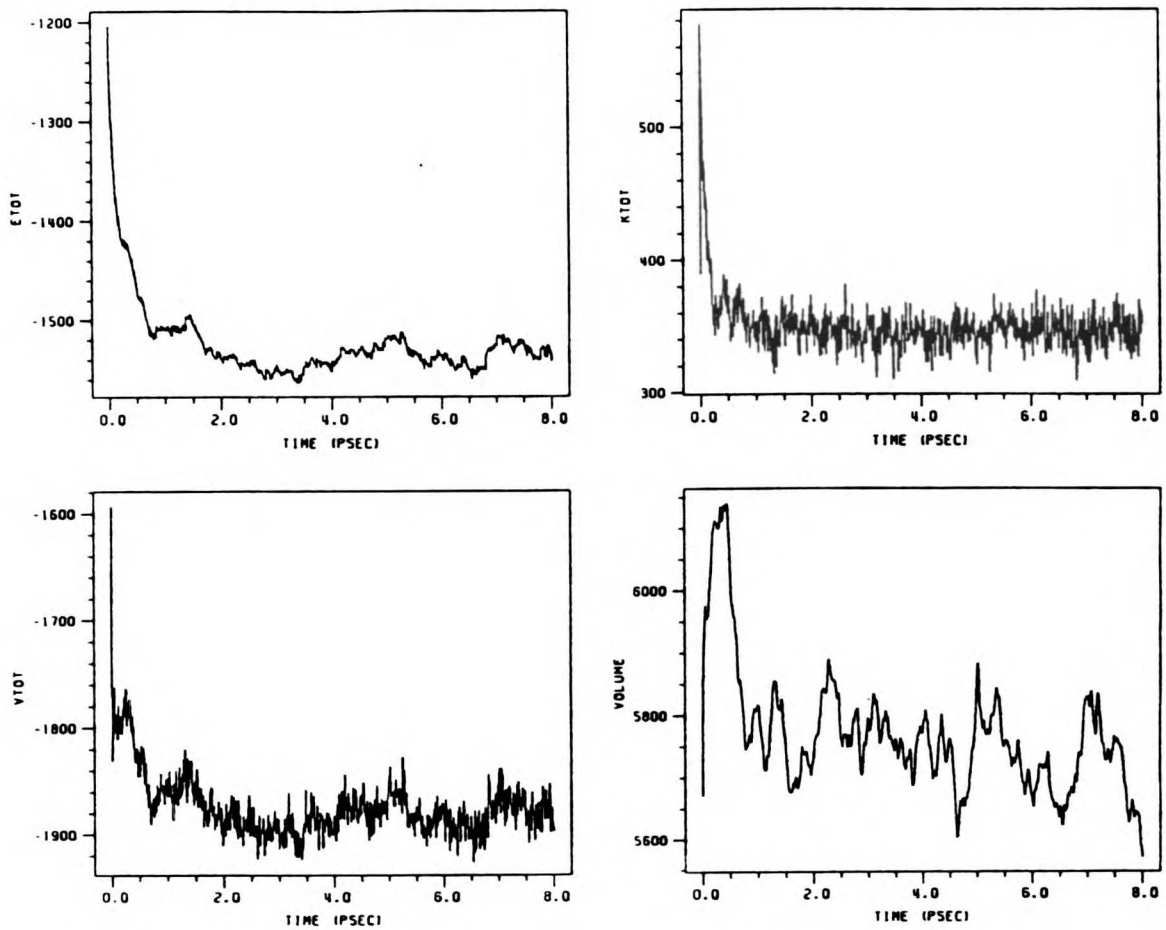


Figure 5. Relaxing fluctuations of quantities (as in fig. 3) for MD at lambda = 0.5 using the pre-equilibrated $\lambda_{1.0}$ configuration as the starting structure.

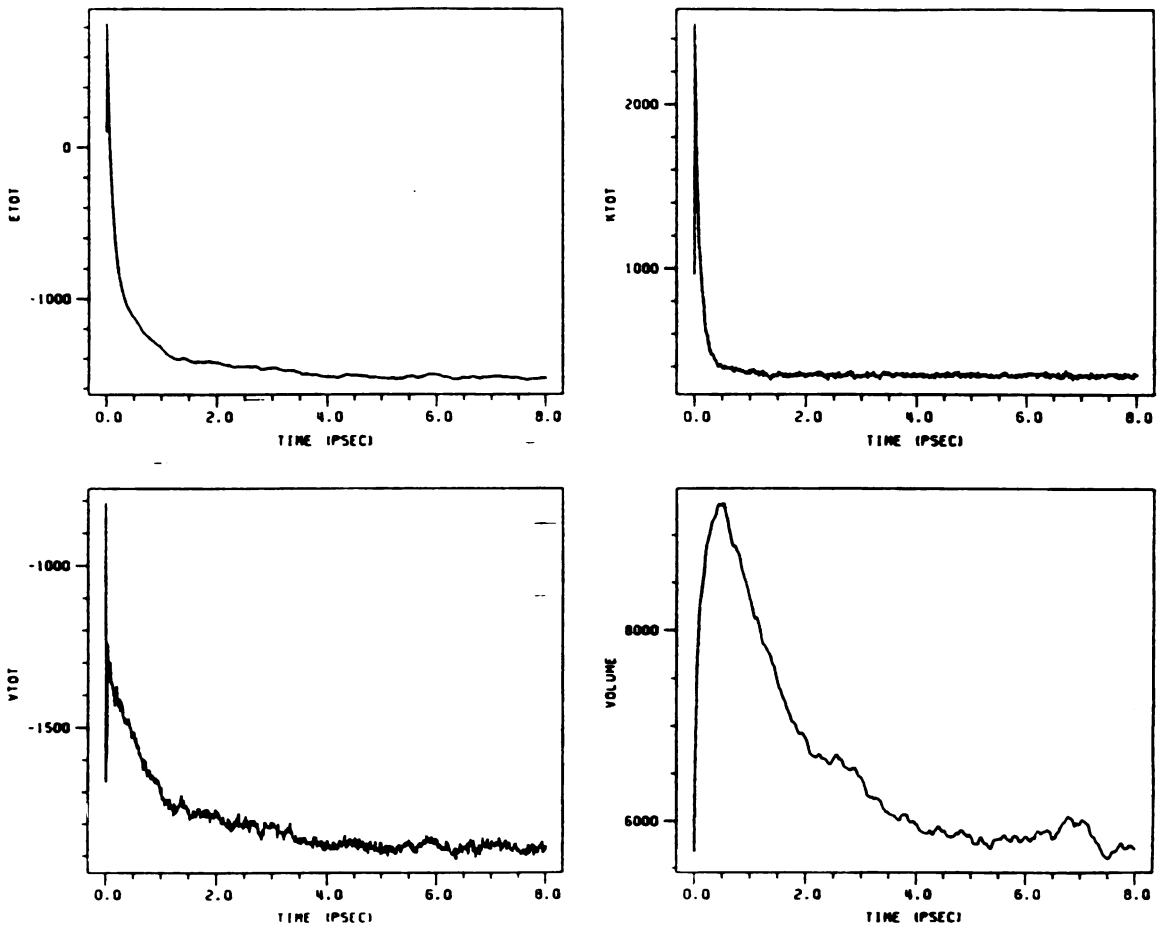


Figure 6. Relaxing fluctuations of quantities (as in fig. 3) for MD at lambda = 0.15 using the pre-equilibrated $\lambda_{1.0}$ configuration as the starting structure.

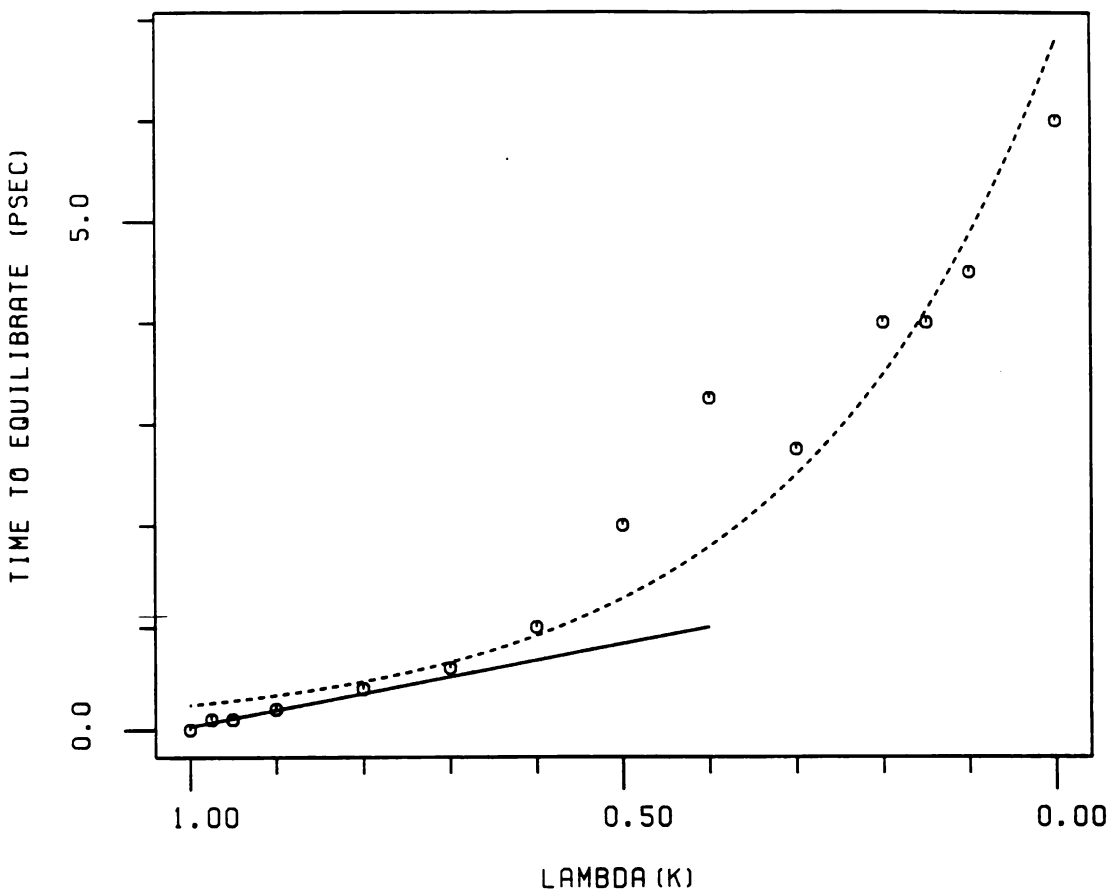


Figure 7. Plot of time to equilibrate (psec) versus λ_K for K in the range 1→0. Each point represents the length of time that MD was run at λ_K to relax the aqueous lithium/cesium system when the equilibrated λ_1 state is taken as the starting state. The dashed line is an exponential curve fitted to the points (excluding the point at $\lambda_{0.6}$). The solid line simply emphasizes the linear region near the λ_1-endpoint.

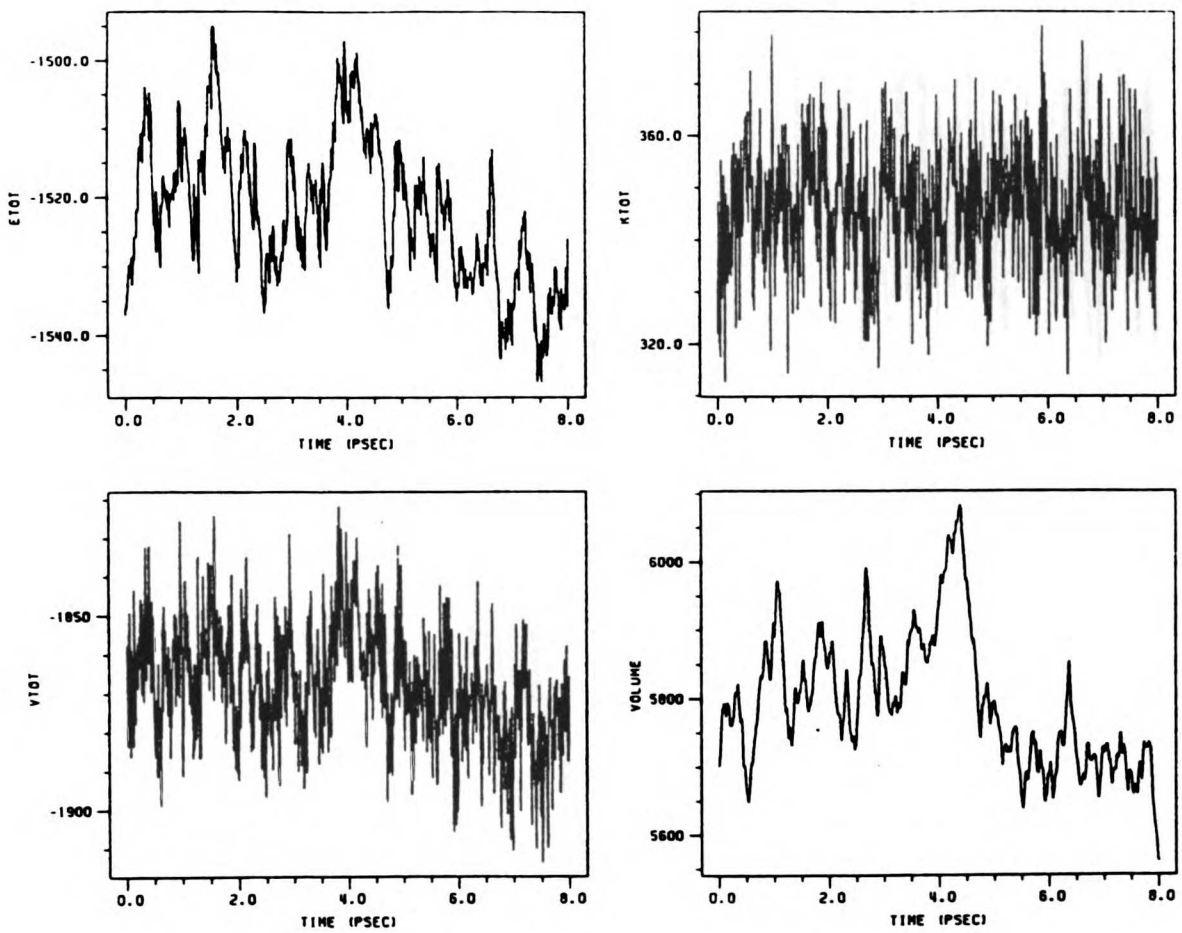


Figure 8. Fluctuations in MD derived quantities at lambda = 0.0 observed for 8.0 picoseconds. The starting state was prepared via an 8.0 psec RAM (4000 MD steps) in the ($\lambda_1 \rightarrow \lambda_0$) direction.

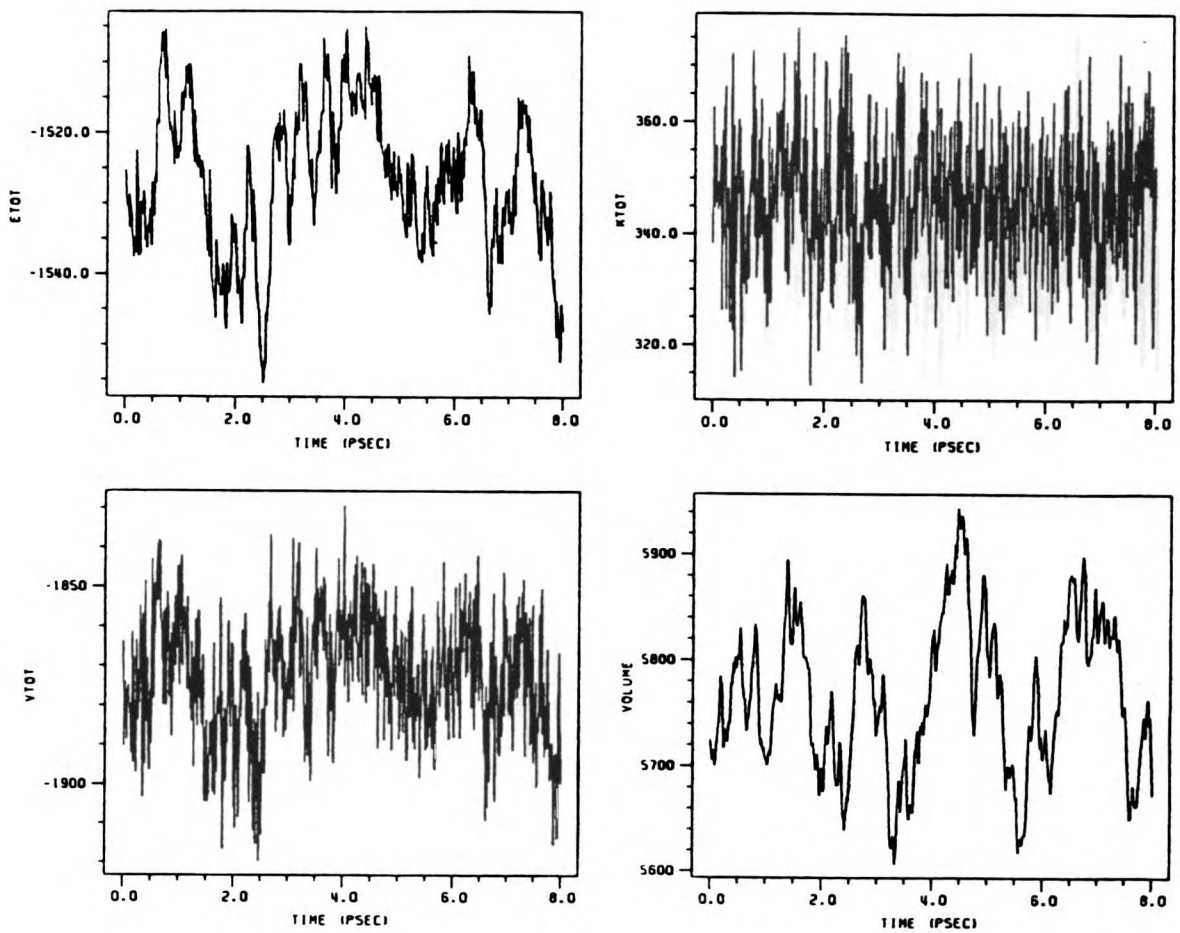


Figure 9. Fluctuations in MD derived quantities at lambda = 0.0 from a starting state prepared via a 1.0 psec RAM (500 MD steps $\lambda_{1\rightarrow 0}$).

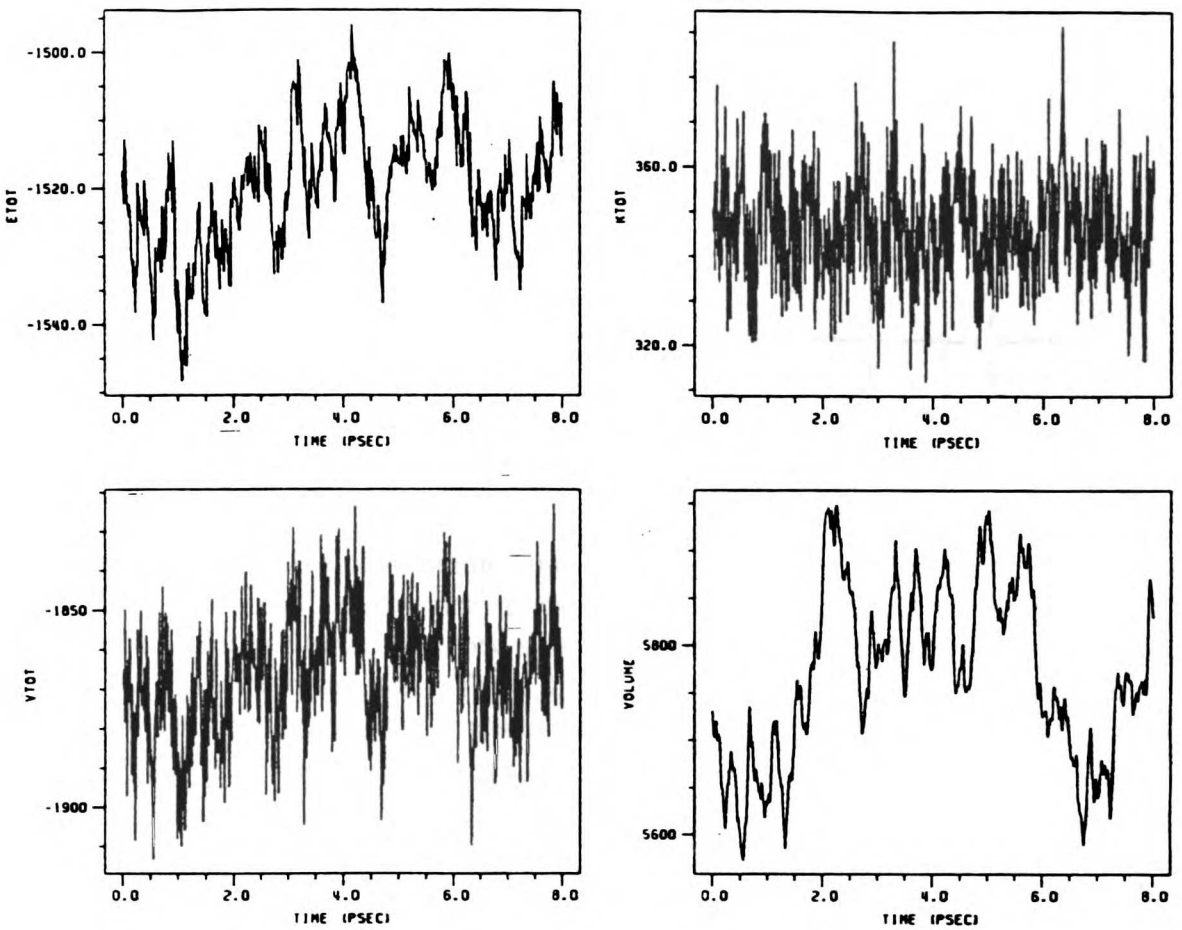


Figure 10. Fluctuations in MD derived quantities at lambda = 0.0 from a starting state prepared via a 0.2 psec RAM (100 MD steps $\lambda_{1\rightarrow 0}$).

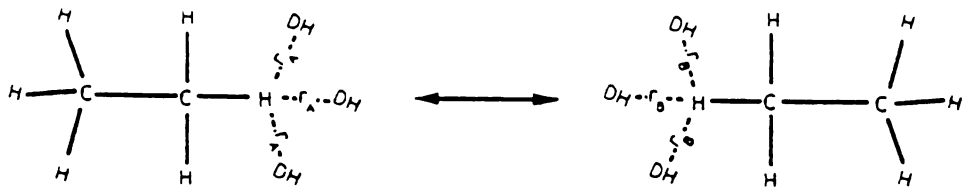


Figure 11. The "zero-sum" perturbation of ethane to ethane´. The dummy atoms (DH) have no nonbonded interactions with the system. Target values for the valence angles and torrsions in one state are the same as those for the corresponding atoms in the other state. Thus, the beginning and final states are identical, except that the lengths of the bonds to the dummy atoms in state A are 0.2 Å (r_A = 0.2 Å), while in state B the bonds to the dummy atoms are 1.09Å long (r_B = 1.09 Å). The free energy for this change should be $\Delta G = 0$, regardless of the potential function or methodology used.

APPENDIX

Fortran source code is given for three subroutines faciliting the *'setup'* for MD simulations in arbitrary solvents.

Gsol.for is a routine to immerse a solute molecule into a bath of arbitrary solvent. It is called by edit.for, standard AMBER, except that the call to gsol is present as an option in the main menu.

Grobox builds a solvent box of specified size around a given solute.

Redsol reads in a general solvent molecule from a standard link.bin file.

```
      SUBROUTINE GSOL
c
c     WRITTEN BY:  Stephen E. DeBolt
c                  September 1988
c                  UCSF Pharm Chem
c
C     ----- ROUTINE TO IMMERSE THE SOLUTE MOLECULE INTO A BATH
C          OF ANY GENERAL SOLVENT AND PUT A CUBIC BOX AROUND IT -----
c
c          GSOL IS A HIGHLY MODIFIED AND EXTENDED VERSION OF BSOL.FOR,
c          an AMBER subroutine which was hardcoded to solvate solutes with
c          TIP3P waters only.  The authorship of bsol.for was P.K. Weiner,
c          U.C. Singh, P.A. Kollman.
C
C
      parameter    (mslres=13850 , mapsr=50 , mapsm=50)
      integer      soltop
      real         cms(1:3)
      character*20  solnam
c
c     ----- SOLBOX 3-D array holds the solvent coords as function of
c          (num_solvent_res,num_atom_per_solvent_res,xyz_coords) ,
c          dimensioned to do up to 27, 8x8x8 boxes of 512 residues,
c          or a total of 13850 solvent residues, where a solvent may
c          have a total of maxsize 50 atoms per entire molecule. ------
c
      real         solbox(1:mslres,1:mapsr,1:3)
c
c     ----- COMMONS FOR THE SOLUTE TOPOLOGY INFORMATION, ETC -----
c
      COMMON/BOXPAR/IFBOX,IPTRES,NSPM,NSPSOL,BETA,BOX(3),NSP(9000)
      COMMON/INFOM/NATOMS,NRES,NBOND,NTHETA,NPHI,NBUILD
      COMMON/MISCL/LSTDRW,NTOTRS,ICONN,NA0,NA1,ITYPCO,MLAST,ISYMDU
      COMMON/GENL/NTOTAT,NMOL,NUMMOL,NAMRES,IBOND(200),ITHETA(200),
     +          IPHI(200),NATML(200),MOLTYP(200),NFRESM(200)
      COMMON/INFOR/LABRES(9000),ITRES(9000),IBRES(9000),
     +          IPRES(9000)
      COMMON/ZMAT/R(30000),THETA(30000),PHI(30000)
      COMMON/INFOA/CHRG(30000),IGRAPH(30000),ISYMBL(30000),
     +        ITREE(30000),JOIN(30000),IROTAT(30000),NATN(30000)
      COMMON/BOND/IB(50000),JB(50000)
      COMMON/ANGL/IT(50000),JT(50000),KT(50000)
      COMMON/DIHED/IP(50000),JP(50000),KP(50000),LP(50000)
      COMMON/NONBON/NEXT,NUMEX(30000),NATEX(80000)
      COMMON/IOFILE/IN,IOUT,IOUTB,IWI,IWN,IWA,IWO,INPO,IWP
```

```
      COMMON/WORKA/MIXC(50000),IWORK(50000)
      COMMON/RUNHED/ITITL(20),ITITL1(20)
c
c     ----- COMMONS FOR THE SOLVENT TOPOLOGY INFORMATION -----
c
      COMMON/INFOMs/NBONDs,NTHETs,NPHIs,ISYMDs
      COMMON/CROSSs/IRCsSS,NCRsSS,NCRsSF,IACsSS(30),JACsSS(30),
     +          ICROSs(30),JCROSs(30)
      COMMON/GENLS/Napsm,Nrpsm,NMOLs,IBONDs(50),ITHETs(50),
     +          IPHIs(50),NATMLs(50),MLTYPs(10),NFRSMs(50)
      COMMON/INFORS/LABRSs(9),ITRESs(9),IBRESs(9),IPRESs(9)
      COMMON/ZMATS/Rs(50),THETAs(50),PHIs(50)
      COMMON/INFOAS/CHRGs(50),IGRPHs(50),ISYMBs(50),ITREEs(50),
     +          JOINs(50),IROTAs(50)
      COMMON/BONDS/IBs(50),JBs(50)
      COMMON/ANGLS/ITs(50),JTs(50),KTs(50)
      COMMON/DIHEDS/IPs(50),JPs(50),KPs(50),LPs(50)
      COMMON/NONBNS/NEXTs,NUMEXs(50),NATEXs(500)
C
      DATA   KCION/4HCIO /
C
c     ----- WRITE OUT SUBROUTINE HEADER INFORMATION ------------------
      write(6,'()')
      write(6,'()')
      write(6,'(''----------------------------------------------'',
     +          ''-------------------------------'')')
      write(6,'()')
      write(6,'(10x,''GENERAL SOLVENTS SUBROUTINES '',
     +            ''ARE BEING CALLED.'')')
      write(6,'(16x,''WRITTEN BY:  Stephen E. DeBolt'')')
      write(6,'(16x,''           Sept 1988, UCSF(PAK)'')')
      write(6,'()')
      write(6,'(''----------------------------------------------'',
     +          ''-------------------------------'')')
      write(6,'()')
      write(6,'()')
      soltop = 36
      NMOLW = NMOL+1
      IFBOX = 1
      MAXRES = mslres
C
C     ----- PUT EACH MOLECULE DEFINED BY LINK AS ONE SUB-MOLECULE -----
c     ----- these submolecs will be the "solute" which gets solvated --
C
      NSPM = 0
```

```
      JFIN = 1
      DO 140 I = 1,NMOL
      JRES = NFRESM(I+1)-NFRESM(I)
      JEND = JFIN+JRES-1
      NSPM = NSPM+1
      NSP(NSPM) = IPRES(JEND+1)-IPRES(JFIN)
      JFIN = JEND+1
  140 CONTINUE
      IPTRES = NFRESM(NMOL+1)-1
C         IF ANY FOR GATHERING INTO THE BOX -----
C
      NSPSOL = 0
      DO 160 I = 1,NMOL
      JRES = NFRESM(I)
      IF(LABRES(JRES).EQ.KCION) GO TO 180
      NSPSOL = NSPSOL+1
  160 CONTINUE
  180 CONTINUE
      NSPSOL = NSPSOL+1
C
      open(soltop,file='slnktp',status='old',form='unformatted')
      call redsol(soltop)
      write(iout,'(/,2x,''NAPSM= '',I6,4X,''NRPSM= '',I6)')
     +   napsm,nrpsm
      if (napsm .ge. mapsm) then
        write(6,'(/,/,2x,''---ERROR--- TOO MANY ATOMS PER '',
     +          ''SOLVENT MOLECULE'')')
      STOP '...AN ERROR HAS OCCURRED...'
      endif
C
C     ----- DISCHK = DISTANCE ABOUT SOLUTE that CONTACT WITH the
C               SOLVENT IS CHECKED AND LIMITED -----
C
      READ(IN,'(f10.5,i5)') dischk,NCUBE
      IF(NCUBE.LE.0) NCUBE = 3
      WRITE (IOUT,1070) dischk
      dischk = dischk * dischk
C
      IN6 = 35
      isol = 0
      XCM = 0.0
      YCM = 0.0
      ZCM = 0.0
      AWT = 0.0
C
```

```
      READ(IN,9118) CUTX,CUTY,CUTZ
      WRITE(IOUT,1050) CUTX,CUTY,CUTZ
      NORIGT = NTOTAT
C
C     ----- ORIENT THE SOLUTE ALONG THE PRINCIPAL AXES -----
C
      CALL MOVECM(NORIGT,R,THETA,PHI,1)
C
C     ----- READ IN THE SOLVENT COORDINATES FROM A "MAKE_BOX"
C           CONFIGURATION, AN MD PRE-EQUILIBRATED BOX, OR A
C           MONTE CARLO PRE-EQUILIBRATED BOX.  -----
C
      open(IN6,file='boxfil',status='old',form='formatted')
      CALL grobox(IN6,Nsolvt,natpsm,nrspsm,NCUBE,BOXL,solbox,
     +            solnam,IOUT)
      CALL CLOSC(IN6,0)
      WRITE(IOUT,1110)
      WRITE(IOUT,1100) solnam
      WRITE(IOUT,1110)
      Ncube3 = Ncube*Ncube*Ncube
      WRITE (IOUT,1010) Nsolvt*Ncube3
      WRITE(IOUT,1190) NCUBE, solnam
      if ((natpsm .ne. napsm) .or. (nrspsm .ne. nrpsm)) then
       write(6,'(///,2x,''-------- ERROR --------'',/,
     +      2x,''MISMATCHED SOLVENT TOPOL AND COORDS'')')
      write(6,'(2x,''natpsm='',i6,2x,''nrspsm='',i6,///)')
     +          natpsm,nrspsm
      stop '...AN ERROR HAS OCCURRED...'
      endif
c
C     ----- GET THE MAXIMA AND MINMA OF THE COORDINATES -----
C
      XMIN = GCMIN(NORIGT,R)-CUTX
      YMIN = GCMIN(NORIGT,THETA)-CUTY
      ZMIN = GCMIN(NORIGT,PHI)-CUTZ
      XMAX = GCMAX(NORIGT,R)+CUTX
      YMAX = GCMAX(NORIGT,THETA)+CUTY
      ZMAX = GCMAX(NORIGT,PHI)+CUTZ
      write(6,'(5x,''RESULT OF USER DEFINED SOLVENT SHELL CUTOFF'')')
      write(6,'(7x,''XYZ-MINS:'',3(2X,F8.3))') XMIN,YMIN,ZMIN
      write(6,'(7x,''XYZ-MAXS:'',3(2X,F8.3))') XMAX,YMAX,ZMAX
C
      BOX(1) = XMAX-XMIN
      BOX(2) = YMAX-YMIN
      BOX(3) = ZMAX-ZMIN
```

```
      BETA = 90.0E0
      if((box(1).gt.boxl).OR.(box(2).gt.boxl).OR.(box(3).gt.boxl))
     + write(6,'(/,2x,''User Specified CUTOFF for Solvent Shell''
     + ,/,''lays outside bounds of periodic box generated.'')')
      IF(BOX(1).GT.BOXL) WRITE(6,9408) boxl
      IF(BOX(2).GT.BOXL) WRITE(6,9418) boxl
      IF(BOX(3).GT.BOXL) WRITE(6,9428) boxl
      IF(BOX(1).GT.BOXL) BOX(1) = BOXL
      IF(BOX(2).GT.BOXL) BOX(2) = BOXL
      IF(BOX(3).GT.BOXL) BOX(3) = BOXL
C
C     ----- CALCULATE THE CENTER OF POSITION OF THE SOLVENT. -----
C
      avsolx = 0.0e0
      avsoly = 0.0e0
      avsolz = 0.0e0
      DO 120 NC = 1, Ncube3
       mnsols = (NC-1) * nsolvt
       do 118 j = 1+mnsols, mnsols+nsolvt
        do 116 k = 1, napsm
         avsolx = avsolx + solbox(J,k,1)
         avsoly = avsoly + solbox(J,k,2)
         avsolz = avsolz + solbox(J,k,3)
  116    continue
  118   continue
  120 CONTINUE
      Rna = FLOAT(Nsolvt*napsm)
      write(6,'(1x,''NUM ATOMS IN ONE BOX = '',f8.3)') rna
      CMS(1) = avsolx/(Rna*Ncube3)
      CMS(2) = avsoly/(Rna*Ncube3)
      CMS(3) = avsolz/(Rna*Ncube3)
      write(6,'(1x,''COM OF THE SOLVENT BEF TRANSLN TO ORIG'')')
      write(6,'(2x,3(2X,f8.3))') cms(1),cms(2),cms(3)
C
C     ---- MAIN DO LOOP. ----
C
      DO 74 Imol = 1, Nsolvt*Ncube3
C
      do 101 inat=1,napsm
       do 100 iq = 1, 3
        solbox(Imol,inat,iq) = solbox(Imol,inat,iq) - CMS(iq)
  100  continue
  101 continue
C
C     ----- SKIP SOLVENT MOLEC IF IT'S OUTSIDE USER SPECIF'D SHELL -----
```

```
C
      DO 105 n=1,napsm
      IF((solbox(Imol,n,1).GT.XMAX
     +            .or. solbox(Imol,n,1).LT.XMIN)
     +   .OR. (solbox(Imol,n,2).GT.YMAX
     +            .or. solbox(Imol,n,2).LT.YMIN)
     +   .OR. (solbox(Imol,n,3).GT.ZMAX
     +            .or. solbox(Imol,n,3).LT.ZMIN)) then
        GO TO 74
      ENDIF
C
C     ----- SKIP THE SOLVENT IF IT IS BUMPING INTO THE SOLUTE. ------
C     ----- CHECK SOLUTE-SOLVENT DISTS ABOUT ALL SOLUTE ATOMS, ------
C     ----- AGAINST ALL SOLVENT ATOMS OF ALL SOLVENT MOLECULES. -----
C
      do 104 J = 1,NORIGT
        RSQ = (solbox(Imol,n,1)-R(J))**2
     +       +(solbox(Imol,n,2)-THETA(J))**2
     +       +(solbox(Imol,n,3)-PHI(J))**2
        IF(RSQ .LT. dischk) then
          goto 74
        ENDIF
  104   continue
  105 CONTINUE
C
      Isol = Isol + 1
C
C     ---- INCLUDE (retain) THIS SOLVENT'S COORDS IN THE SYSTEM ----
C
      DO 5 K = 1,napsm
        R(NTOTAT+K) = solbox(Imol,k,1)
        THETA(NTOTAT+K) = solbox(Imol,k,2)
        PHI(NTOTAT+K) = solbox(Imol,k,3)
    5 CONTINUE
c
c     ---- INCLUDE THIS SOLVENT'S TOPOL IN THE SYSTEM TOPOL ARRAYS ----
c
      if (isol .ne. 1) then
        join(ntotat+1) = ipres(ntotrs+1)-1
      else
        join(ntotat+1) = ntotat
      endif
      DO 10 j=2,NAPSM
        join(ntotat+j) = joinS(j) + ntotat
   10 CONTINUE
```

```
      DO 20 j=1,nbondS
        IB(nbond+j) = IBs(j) + ntotat
        JB(nbond+j) = JBs(j) + ntotat
   20 CONTINUE
      do 24 j=1,nthetS
        IT(ntheta+j) = ITs(j) + ntotat
        JT(ntheta+j) = JTs(j) + ntotat
        KT(ntheta+j) = KTs(j) + ntotat
   24 continue
      do 26 j=1,nphiS
        IP(nphi+j) = ABS(IPs(j)) + ntotat
        JP(nphi+j) = ABS(JPs(j)) + ntotat
        KP(nphi+j) = ABS(KPs(j)) + ntotat
        LP(nphi+j) = ABS(LPs(j)) + ntotat
        if (IPs(j).lt.0) IP(nphi+j) = -IP(nphi+j)
        if (JPs(j).lt.0) JP(nphi+j) = -JP(nphi+j)
        if (KPs(j).lt.0) KP(nphi+j) = -KP(nphi+j)
        if (LPs(j).lt.0) LP(nphi+j) = -LP(nphi+j)
   26 continue
      jsum=0
      do 30 j=1,napsm
        ntotpj = ntotat + j
        isymbl(ntotpj) = isymbS(j)
        igraph(ntotpj) = igrphS(j)
        itree(ntotpj)  = itreeS(j)
        irotat(ntotpj) = irotaS(j)+ntotat
        chrg(ntotpj)   = chrgS(j)
        numex(ntotpj)  = numexS(j)
c
c       do 25 k=1,numexS(j)
c         natex(next+jsum+k) = natexS(jsum+k) + ntotat
c   25  continue
c       jsum = jsum + numexS(j)
c
   30 continue
        do 35 k=1,nextS
          if (natexS(k) .eq. 0) then
            natex(next+k) = 0
          else
            natex(next+k) = ntotat + natexS(k)
          endif
   35 continue
      next = next + nextS
      natex(next+1) = 0
c
```

```
c    ----- MAKE EACH OF THE SOLVENT'S RESIDUES A UNIQUE RESIDUE
c         IN THE SOLUTE/SOLVENT SYSTEM & UPDATE NUM OF SUBMOLECS -----
c
     do 40 m=1,nrpsm
      NTOTRS = NTOTRS+1
      ntotat = ntotat + ipresS(m+1)-ipresS(m)
      nbond = nbond + ibresS(m+1) - ibresS(m)
      IF(NTOTRS.GE.MAXRES) GO TO 900
      LABRES(NTOTRS) = labrsS(m)
      ITRES(NTOTRS) = itresS(m)
      IBRES(NTOTRS+1) = NBOND+1
      IPRES(NTOTRS+1) = NTOTAT+1
   40 continue
     ntheta = ntheta + nthetS
     nphi  =  nphi + nphiS
     NSPM = NSPM+1
     NSP(NSPM) = napsm
   74 CONTINUE
C
C    ----- UPDATE THE (SYSTEM) MOLECULE INFORMATION ARRAY -----
C    ----- CONSIDERING ALL THE ADDED SOLVENTS AS 1 MOLEC. -----
C
     NFRESM(NMOLW+1) = NTOTRS+1
     write(6,'(/,1x,i7,'' RESIDUES TOTAL (SOLUTE+SOLVENTS) '')')
    +                    NTOTRS
     IBOND(NMOLW+1) = NBOND+1
     ITHETA(NMOLW+1) = NTHETA+1
     IPHI(NMOLW+1) = NPHI+1
     NATML(NMOLW) = NTOTAT
     MOLTYP(NMOLW) = mltypS(1)
     NMOL = NMOL+1
C
C    ----- PUT THE BOX BACK INTO THE POSITIVE QUADRANT -----
C
     XMIN = GCMIN(NTOTAT,R)
     YMIN = GCMIN(NTOTAT,THETA)
     ZMIN = GCMIN(NTOTAT,PHI)
     DO 82 I = 1,NTOTAT
     R(I) = R(I)-XMIN
     THETA(I) = THETA(I)-YMIN
     PHI(I) = PHI(I)-ZMIN
   82 CONTINUE
C
     CGTOT = 0.
     DO 55 I = 1,NTOTAT
```

```
      IF(ISYMBL(I).EQ.ISYMDU) GO TO 55
      CGTOT = CGTOT+CHRG(I)
   55 CONTINUE
      WRITE(IOUT,1120) Isol*nrspsm, Isol
      WRITE(IOUT,270) CGTOT
      WRITE(IOUT,9438) BOX(1),BOX(2),BOX(3)
      WRITE(IOUT,9448) NSPM
C     WRITE(IOUT,9458) (NSP(I),I=1,NSPM)
      WRITE(IOUT,9468) IPTRES,NSPSOL
      RETURN
  900 CONTINUE
      WRITE(IOUT,9508) NTOTRS
  270 FORMAT(/,t2,'TOTAL CHARGE ON MOLECULE  = ',F9.3,/)
 1010 FORMAT (/,t2,i6,' SOLVENT MOLECULES ARE INITIALLY CONSIDERED')
 1040 FORMAT(3I5)
 1050 FORMAT(/T2,'CUTOFF CRITERION FOR DISCARDING THE SOLVENTS',
     +     /T2,'CUTX =',F10.4,2X,'CUTY =',F10.4,2X,'CUTZ =',F10.4,/)
 1070 FORMAT (/,t2,'ANY SOLVENT WITH ANY ATOM DISTANCE LESS THAN ',F6.2,
     +      ' ANGSTROMS',/,t2,'FROM THE SOLUTE WILL BE DISCARDED')
 1100 FORMAT (t2,'IMMERSING THE SOLUTE INTO A BATH OF ',a20)
 1110 FORMAT (t2,'****************************************',
     +       '**************')
 1120 FORMAT (t2,I7,' RESIDUES, ',I7,
     +       ' MOLECULES ARE SOLVATING THE SOLUTE')
 1130 FORMAT (/,t2,I4,' SOLUTE ATOMS FOUND FOR CENTER OF MASS ',
     +      'CALCULATION')
 1140 FORMAT (/,t2,'NUMBER OF MOLECULES CONSIDERED FROM CENTER',
     +      ' OF MASS ROUTINE = ',I4,A20)
 1150 FORMAT (/,t2,'RESIDUES FOUND FOR COM CALCULATION:',//,
     +      (6X,10(1X,I4)))
 1190 format(/t2,'Number of solvent cubes to be ',
     +     'generated =',2x,I2,A20)
 9118 FORMAT(5F10.5)
 9408 FORMAT(t2,'WARNING: BOXX IS RESET TO THE default SIZE = ',f10.3)
 9418 FORMAT(t2,'WARNING: BOXY IS RESET TO THE default SIZE = ',f10.3)
 9428 FORMAT(t2,'WARNING: BOXZ IS RESET TO THE default SIZE = ',f10.3)
 9438 format(/t2,'box X =',F10.2,' box Y =',F10.2,' box Z =',F10.2)
 9448 FORMAT(/T2,'NUMBER OF SUB-MOLECULES FOR DYNAMICS =',I6,/)
 9458 FORMAT(T2,10I5)
 9468 FORMAT(/T2,'The INDEX to the last solute res =',I5,
     +     /T2,'THE SUBMOLEC INDEX to the 1st SOLVENT MOLEC =',I5)
 9508 FORMAT(t2,'TOO MANY RES, number of residues exceeded',i7)
      STOP
      END
c
```

```
c
c
      SUBROUTINE GROBOX(NF,nsols,natpsm,nrspsm,NCUBE,BOXL,
     +                  solbox,solnam,IOUT)
c
c     WRITTEN BY:  Stephen E. DeBolt
c                  September 1988
c                  UCSF Pharm Chem
C
C     ----- ROUTINE TO GROW A SOLVENT BOX OF "ANY" DIMENSION BASED
C           ON THE STARTING SOLVENT CUBE. THAT IS, IT GROWS AN
C           N**3 SIZE CLUSTER OF (MAX OF 13850 SOL_RES's) BOXES -----
c
c     ----- SOLBOX 3-D array holds the solvent coords as function of
c           (num_solvent_res,num_atom_per_solvent_res,xyz_coords) ,
c           dimensioned to do up to an 8x8x8 box of 512 (49 atom) res's. ------
c
      parameter    (mslres=13850 , mapsr=50 , mapsm=50)
      real         solbox(1:mslres,1:mapsr,1:3)
      real         Boxshf(3)
      integer      Q
      character*76  ititlt
      character*20  solnam
      character*8   crdfmt
c
C
C     ----- READ THE INITIAL STARTING SOLVENT BOX FIRST -----
C
      read(nf,'(a76)') ititlt
      backspace(nf)
      read(nf,'(30x,a20)') solnam
      write(6,'(/,9x,''----- GENERAL SOLVENT COORDINATE '',
     +          ''FILE TITLE -----'')')
      write(6,'(4x,a76,/)') ititlt
      READ(NF,9018) nsols,natbox,nrspsm,boxsid,crdfmt
      natpsm = natbox/nsols
      nsolvt = nsols*NCUBE**3
      BOXL = boxsid*NCUBE
      IF ((nsolvt.LT.mslres) .AND. (natpsm.LT.mapsm)) THEN
C
       if (crdfmt .eq. '(6f13.8)') then
        READ(NF,9028) (((solbox(I,J,K),K=1,3),J=1,natpsm),I=1,nsols)
       else
       if (crdfmt .eq. '(6f12.7)') then
        READ(NF,9029) (((solbox(I,J,K),K=1,3),J=1,natpsm),I=1,nsols)
```

```
      else
      if (crdfmt .eq. '(10f8.3)') then
       READ(NF,9030) (((solbox(I,J,K),K=1,3),J=1,natpsm),I=1,nsols)
      else
       write(6,'(/,/,2x,''--------ERROR READING SOLBOX.-------'')')
       write(6,'(4x,''CHECK SOLBOX FMT HEADER = 6f13.8 or '',
     +           ''6f12.7 or 10f8.3 only.'',/)')
      endif
      endif
      endif
C
C     ----- NOW GENERATE THE BIG CUBE -----
C
      M = 0
      DO 100 I = 1,NCUBE
      DO 120 K = 1,NCUBE
      DO 140 J = 1,NCUBE
      BOXshf(1) = boxsid*(I-1)
      BOXshf(2) = boxsid*(K-1)
      BOXshf(3) = boxsid*(J-1)
      NS = 0
      mnsols = M * nsols
      DO 180 MNS=(1+mnsols),(mnsols+nsols)
       NS = NS + 1
       do 170 NA=1,natpsm
        do 160 Q=1,3
         solbox(MNS,na,q) = solbox(NS,na,q) + boxshf(q)
  160   continue
c      write(6,'(2x,''solbox(mns,na,q)= '',3f8.3)')
c     +           (solbox(mns,na,q),q=1,3)
  170  continue
  180 CONTINUE
      M = M + 1
  140 CONTINUE
  120 CONTINUE
  100 CONTINUE
      RETURN
C
      ELSE
       write(6,'(/,2x,''--- ERROR OF MAJOR PROPORTION.  '',
     +      ''CLEAN OUT YOUR DESK.  YOU'RE FIRED! ---'',/)')
       write(6,'(/,2x,'' TOO MANY SOLVENT RESIDUES:'',i8,
     +           '' OR TOO MANY ATOMS PER SOLVENT MOL:'',
     +           i8,/,/)') nsolvt,natpsm
       write(IOUT,9008) mslres,mapsr
```

```
      STOP '...YOU HAVE SCREWED UP AGAIN...'
     ENDIF
C
 9028 FORMAT(6F13.8)
 9029 FORMAT(6F12.7)
 9030 FORMAT(10f8.3)
 9018 FORMAT(3I6,F12.7,2x,a8)
 9008 FORMAT(t2,'ERROR:  requested cube is too big. ',/,t2,
    + 'max number of solvents is ',I8,/,t2,'max number of ',
    + 'atoms per solvent res is ',I8,/,t2,
    + 'reduce Ncube or size of res's.',/,/)
     END
c
c
c
     SUBROUTINE REDSOL(nf)
c
c    MODIFIED BY: Stephen E. DeBolt
c              September 1988
c              UCSF Pharm Chem
c
c    TO READ IN A "GENERAL SOLVENT" LINK.BIN FILE.  This is a modified
c    version of subroutine "READIN" which reads in a solute.
c    This version facilitates the reading in of a general solvent molecule.
c    Complete array overwrite checking has been installed as well.
c    The original authorship of readin was P.K. Weiner, U.C. Singh, and
c    P.A. Kollman.
C
     COMMON/INFOMs/NBONDs,NTHETs,NPHIs,ISYMDs
     COMMON/CROSSs/IRCsSS,NCRsSS,NCRsSF,IACsSS(30),JACsSS(30),
    +         ICROSs(30),JCROSs(30)
     COMMON/GENLS/Napsm,Nrpsm,NMOLs,IBONDs(2),ITHETs(2),
    +         IPHIs(2),NATMLs(2),MLTYPs(2),NFRSMs(2)
     COMMON/INFORS/LABRSs(9),ITRESs(9),IBRESs(9),IPRESs(9)
     COMMON/ZMATS/Rs(50),THETAs(50),PHIs(50)
     COMMON/INFOAS/CHRGs(50),IGRPHs(50),ISYMBs(50),ITREEs(50),
    +         JOINs(50),IROTAs(50)
     COMMON/BONDS/IBs(50),JBs(50)
     COMMON/ANGLS/ITs(50),JTs(50),KTs(50)
     COMMON/DIHEDS/IPs(50),JPs(50),KPs(50),LPs(50)
     COMMON/NONBNS/NEXTs,NUMEXs(50),NATEXs(51)
     COMMON/RUNHDS/ititls(19)
C
     READ(nf)      (ITITLs(I),I = 1,19)
     write(6,'(/,1x,''GENERAL SOLVENT TOPOLOGY FILE INFORMATION: '')')
```

```
      write(6,'(2x,19a4)') (ititls(i),i = 1,19)
      READ(nf)     Napsm,NEXTs,NMOLs,nrpsm,NBONDs,Nthets,NPHIs,INDXs,
     +          ISMDUs,IRCsSS,IRCsSS,NCRsSS
      write(6,'(5x,'' Napsm NEXTs NMOLs nrpsm NBONDs Nthets'')')
      write(6,'(2x,6(1x,i6),/)')
     +          Napsm,NEXTs,NMOLs,nrpsm,NBONDs,Nthets
      write(6,'(5x,'' NPHIs INDXs ISMDUs IRCsSS IRCsSS NCRsS'')')
      write(6,'(2x,2(1x,i6),3x,a4,3(2x,i6),/)')
     +          NPHIs,INDXs,ISMDUs,IRCsSS,IRCsSS,NCRsSS
      NMOLPs = NMOLs+1
      if (nmolps .gt. 2) then
       write(6,'(/,2x,''*ERROR* THE GENERAL SOLVENTS CODE'',
     +           '' RECOGNIZES ONLY 1 'SOLVENT MOLECULE'',
     +           '' TYPE'',/)')
       STOP '...AN ARRAY OVERWRITE ERROR AS OCCURRED...'
      endif
c
      NRESPs = Nrpsm+1
      if (nresps .gt. 9) then
       write(6,'(/,2x,''*ERROR* THE GENERAL SOLVENTS CODE'',
     +           '' ALLOWS A MAX OF 8 RESIDUES PER''
     +           '' INDIVIDUAL SOLVENT MOLECULE'',/)')
       STOP '...AN ARRAY OVERWRITE ERROR HAS OCCURRED...'
      endif
c
      if ((nbonds.gt.49).OR.(nthets.gt.49).OR.(nphis.gt.49)) then
       write(6,'(/,2x,''*ERROR* THE GENERAL SOLVENTS CODE'',
     +           '' RECOGNIZES A MAX OF 49 (of each of),''
     +           '' BONDS, ANGLES, & DIHEDRALS, PER''
     +           '' INDIVIDUAL SOLVENT MOLECULE'',/)')
       STOP '...AN ARRAY OVERWRITE ERROR HAS OCCURRED...'
      endif
c
      if (nexts .gt. 49) then
       write(6,'(/,2x,''*ERROR* THE GENERAL SOLVENTS CODE'',
     +           '' RECOGNIZES A MAX OF 49 EXCLUDED ATOMS''
     +           '' PER INDIVIDUAL SOLVENT MOLECULE'',/)')
       STOP '...AN ARRAY OVERWRITE ERROR HAS OCCURRED...'
      endif
c
      IF(NCRsSS.GT.0) then
       if (NCRsSS .le. 29) then
        READ(nf)    (ICROSs(I),JCROSs(I),I = 1,NCRsSS)
       else
        write(6,'(2x,''ERROR , only 29 crosslinks allowed'')')
```

```
   STOP '..AN ARRAY OVERWRITE ERROR HAS OCCURRED...'
  endif
 ENDIF
 READ(nf) (IBONDs(I),ITHETs(I),IPHIs(I),NFRSMs(I),NATMLs(I),
+        MLTYPs(I),I = 1,NMOLPs)
 write(6,'(2x,''ibonds ithets iphis nfrsms natmls mltyps'')')
 write(6,'(6(i5,2x),/)') (IBONDs(I),ITHETs(I),IPHIs(I),
+        NFRSMs(I),NATMLs(I),MLTYPs(I),I = 1,NMOLPs)
 READ(nf)  (LABRSs(I),ITRESs(I),I = 1,Nrpsm)
 write(6,'(2X,''LABRSs= '',19a4)')
+        (LABRSs(I),ITRESs(I),I = 1,Nrpsm)
 READ(nf)  (IBRESs(I),IPRESs(I),I = 1,NRESPs)
 READ(nf)  (IBs(I),JBs(I),I = 1,NBONDs)
 READ(nf)  (ITs(I),JTs(I),KTs(I),I = 1,NTHETs)
 READ(nf)  (IPs(I),JPs(I),KPs(I),LPs(I),I = 1,NPHIs)
 READ(nf)  (IROTAs(I),JOINs(I),NUMEXs(I),I = 1,Napsm)
 READ(nf)  (NATEXs(I),I = 1,NEXTs)
 READ(nf)  (ITREEs(I),ISYMBs(I),IGRPHs(I),I = 1,Napsm)
 READ(nf)  (CHRGs(I),I = 1,Napsm)
 READ(nf)  (Rs(I),THETAs(I),PHIs(I),I = 1,Napsm)
 RETURN
 END
```

UC San Francisco LIBRARY
UNIVERSITY OF CALIFORNIA

For reference

Not to be taken from the room.

617328

3 1378 00617 3283

UC San Francisco LIBRARY

UNIVERSITY OF CALIFORNIA