Computer Simulation of Chemical and Biomolecular Systems
Reprinted from Vol. 482
ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Computer Simulations of Macromolecular Dynamics: Models for Vibrational Spectroscopy and X-Ray Refinement[a]

RONALD M. LEVY
Department of Chemistry
Rutgers University
New Brunswick, New Jersey 08903

INTRODUCTION

The use of computer simulations to study the internal dynamics of proteins and nucleic acids has attracted a great deal of attention in recent years. These simulations constitute the most detailed theoretical approach available for studying internal motions and structural flexibility of proteins.[1,2] It is essential that there be a continuing effort to develop procedures for comparing the results of simulations with a wide variety of experimental measurements. This has been a central focus of our work during the past few years. Such studies are necessary if the methodology is to be reliably used to study properties that are only indirectly accessible to experiment. Equally important, these studies lead to deeper insights into the relationship between experimental measurements and underlying molecular processes. Computer simulations, therefore, provide a fundamental connection between experimental probes of dynamics and analytical theories that are used to interpret experiments. In this paper we review recent work concerned with the use of computer simulations for the interpretation of experimental probes of macromolecular structure and dynamics.

Molecular dynamics simulations using simple potential functions have been demonstrated to provide an accurate description of structural and dynamic properties of both unassociated and complex liquids. Packing considerations play a major role in determining the molecular properties of both liquids and proteins; because packing can be described by simple and reliable potential functions there is a strong foundation for simulation studies of protein dynamics. However, many factors distinguish molecular dynamics simulations of proteins from liquid-state simulations so that it is difficult to use experience gained from molecular dynamics simulations of liquids to estimate the precision inherent in the protein simulations. For liquid simulations the basic system contains at least 100 identical molecules so that it it possible to take advantage of considerable statistical averaging in the calculation of quantities for comparison with experiment. For protein molecular dynamics simulations in contrast, the computational effort required to evaluate the large number of interatomic interactions within a single protein molecule limits the simulated system to one or at most a very small number of (macro) molecules. The highly anisotropic nature of the protein interior and

[a]This work was supported in part by the National Institutes of Health, the Petroleum Research Fund of the American Chemical Society, and the Alfred P. Sloan Foundation.

intrinsic interest in extracting site-specific information further complicates the computational problem.

Additional features of macromolecular simulations that make them different from and more difficult than simulations of liquids and solids include the difficulty in obtaining exact results for comparison with trajectory averages, the slow convergence of the macromolecular simulations, and the problem of treating quantum effects for large systems which lack a high degree of symmetry. Despite these difficulties, computer simulations of biological macromolecules are playing an increasingly important role in biophysical chemistry. One important use involves the modeling of biochemical processes via computer simulation, e.g., the binding of a substrate to an enzyme.[3,4] Another major area of research concerns the relationship between computer simulations and experiment.[5] There are four aspects of this area of research that have been stressed: (1) the use of experimental results to refine empirical potentials; (2) the use of simulations as a testing ground for models used to interpret experiment; (3) the use of simulations to better understand the molecular information contained in experiment; and (4) the use of simulations to suggest new experiments. The development of methods for analyzing NMR relaxation and fluorescence depolarization experiments on proteins and nucleic acids using the results of molecular dynamics computer simulations have been reviewed recently.[5] In this paper, we discuss the development of new methods for simulating vibrational spectra using detailed molecular simulations. In connection with this work we have developed new and more powerful algorithms for performing quantum Monte Carlo simulations,[6,7] and these methods are reviewed. In the final section of this paper we review recent work concerned with the use of molecular dynamics simulations of proteins and nucleic acids to analyze restrained parameters least-squares X-ray refinement models for macromolecules.[8,9]

VIBRATIONAL SPECTROSCOPY

Vibrational spectroscopy has played a very important role in the development of potential functions for molecular-mechanics studies of proteins. Bond length, bond angle, and torsional force constants which appear in the energy expressions are heavily "parameterized" from infrared and Raman studies of small model compounds. Considerable information concerning molecular structure, and potential surfaces, is contained in high-resolution FTIR and resonance Raman studies of polypeptides and proteins. One approach to the interpretation of vibrational spectra for biopolymers has been a harmonic analysis whereby spectra are fit by geometry and/or force constant changes. There are a number of reasons for developing other approaches. The consistent force field (CFF) type potentials used in computer simulations are meant to model the motions of the atoms over a large range of conformations and, implicitly, temperatures, without reparameterization.[10] It is also desirable to develop a formalism for interpreting vibrational spectra which takes into account the variation in the conformations of the chromophore and surroundings which occur due to thermal motions. Much of the interesting structural information that can be extracted from experiments involves understanding the origin of spectral shifts as a function of some experimentally adjustable parameter, often the temperature. If computer simulations

on realistic potential surfaces are to be useful for interpreting these experiments, it is important that methods be developed that incorporate these anharmonic effects on the spectra. Semiclassical trajectories provide one possible approach to this problem.[11,12] Here we present alternative approaches.

Quasiharmonic Method for Calculating Vibrational Spectra from Classical Simulations on Anharmonic Potentials

We have introduced a new method for calculating vibrational spectra from classical molecular dynamics or Monte Carlo simulations.[13] The method involves a quasiharmonic oscillator approximation in which a temperature-dependent quadratic Hamiltonian is parameterized from the results of a simulation on the complete (anharmonic) potential. The parameterization is accomplished by fitting the first and second moments of the coordinate and momentum distributions obtained from a simulation on the exact surface to a harmonic model. The model provides a method for partially incorporating anharmonicity in the evaluation of optical and thermodynamic properties and estimating quantum corrections to the classical simulations. Furthermore, the approximation can be systematically improved and is extremely useful for the development of new quantum computer simulation methods.

As an illustration of the method, we recently reported the results of a vibrational analysis of a small molecule (butane) with six internal degrees of freedom using the quasiharmonic oscillator model.[13] The empirical potential contained all the terms present in the potential for macromolecules, namely, bond stretching, bending, and torsional terms as well as nonbonded interactions. A novel aspect of the simulation procedure was the use of normal-mode eigenvectors as the independent coordinates for Monte Carlo sampling, which was demonstrated to substantially increase the convergence rate of the simulation. From a conventional normal mode analysis we extracted the frequencies of the model which ranged from 119 cm^{-1} for a pure torsional vibration to 1044 cm^{-1} for a mixed bond stretch-angle bend vibration. Classical simulations were performed on the complete surface at a series of temperatures between 5 K and 300 K. We demonstrated how anharmonic effects at higher temperatures can rotate the

TABLE 1. Quasiharmonic Frequencies (in cm^{-1}) Calculated from Monte Carlo Trajectories[a] on the Exact Potential Surface for *Trans* and *Gauche* Butane

Trans Butane			*Gauche* Butane
100 K	200 K	300 K	300 K
113 (5)[b]	102 (17)	91 (25)	99 (33)
407	406	407	419
435	437	436	602
899	911	892	857
1008	1014	1002	966
1045	1055	1043	1034

[a]Monte Carlo trajectories constructed with Q_k (mass-weighted Cartesian) as independent coordinates.

[b]Numbers in parentheses indicate percent deviation from harmonic normal-mode eigenvalues; only deviations greater than 1% indicated.

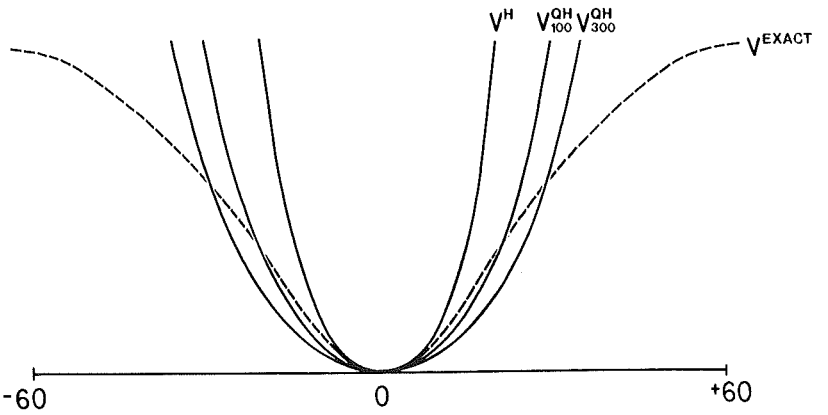

FIGURE 1. Schematic illustration of the exact potential V^{exact}, the harmonic approximation V^H, and quasiharmonic approximations at 100 K, V^{QH}, and 300 K, V^{QH}, for the torsional coordinate Q of *trans* butane. The anharmonicity of the exact potential results in the decreasing curvature of the quasiharmonic potentials with increasing temperature. (See Ref. 13.)

normal coordinates and shift the frequencies with respect to the harmonic values. For the lowest frequency mode (a torsion) increasing the temperature lowered the effective frequency and this was rationalized in terms of the shape of the quasiharmonic torsional potential. The quasiharmonic frequencies calculated from Monte Carlo trajectories on the anharmonic potential surface for *trans* and *gauche* butane are listed in TABLE 1. The effective frequency of the torsional mode is lowered by 25 cm^{-1} to 91 cm^{-1} at 300 K. The anharmonicity of the exact potential results in the decreasing curvature of the quasiharmonic potential and the lowering of the effective torsional frequency with temperature. FIGURE 1 shows a schematic illustration of the exact potential V_{exact}, the harmonic approximation V^H, and quasiharmonic approximations at 100 K and 300 K for the torsional coordinate Q of *trans* butane.

We have recently initiated a study of intramolecular vibrations in liquid water using the quasiharmonic analysis method.[14] For the initial simulations we are using a 3-point model[15,16] for the intermolecular water-water potential and a central force model[17] for the intramolecular vibrations. Since these potentials, in contrast to a more recent complete water potential,[18] were not parameterized to be used together, only qualitative conclusions can be drawn from this initial study. Of interest, the addition of intramolecular vibrations to the water model has a negligible effect on the calculated intermolecular pair correlation functions.[14] The important experimental observation is that on going from the gas phase to the liquid the OH stretching vibration is red-shifted by 300 cm^{-1}, while the HOH bending vibration is blue-shifted by 100 cm^{-1}. In the simplest form of quasiharmonic analysis the effective frequencies are constructed from the probability distributions for the vibrational displacements of the OH bond and HOH bond angle in the gas and liquid phases, respectively. In the gas phase these probability distributions were calculated by direct numerical integration of the partition function, while for the liquid, the distributions were constructed from a 5-ps simulation of 216 water molecules with periodic boundary conditions in the TVN

ensemble. The temperature-dependent effective frequency is then given by:

$$\omega_{eff}^2 = \frac{K_B T}{m_{eff} \langle (\Delta q)^2 \rangle}$$

The effective mass of the OH stretching and HOH bending vibrations were calculated to be 0.94 au and 0.48 au, respectively. For the OH stretching vibration, the gas-phase quasiharmonic frequency is 3970 cm^{-1}, the liquid-phase vibrational frequency is calculated to be 3792 cm^{-1}, resulting in a red-shift of ~200 cm^{-1}. In contrast, the HOH bending quasiharmonic frequency is blue-shifted by 75 cm^{-1} in solution compared to the gas-phase calculation (ω_{eff} = 1274 cm^{-1} gas phase, ω_{eff} = 1348 cm^{-1} liquid). It is very encouraging that the simplest quasiharmonic approximation qualitatively reproduces the directions of the water vibrational frequency shifts on going from the gas to the liquid state. Furthermore, it should be possible to use this simple approximation to estimate from classical water simulations the inhomogeneous contributions to the vibrational linewidths by constructing a distribution of quasiharmonic frequencies from a quasiharmonic analysis of individual water molecules in the simulation.

Path Integral Quantum Monte Carlo Simulations and Vibrational Spectra

The approach to the evaluation of vibrational spectra described above is based on classical simulations, for which quantum corrections are possible. The incorporation of quantum effects directly in simulations of large molecular systems is one of the most challenging and actively pursued research areas in theoretical chemistry today. The development of quantum simulation methods is particularly important in the area of molecular spectroscopy for which quantum effects are often important and where the goal is to use simulations to help understand the structural and dynamical origins of spectral lineshapes and changes in lineshapes with environmental variables, e.g., the temperature. Although it has long been known that path integrals provide one method for calculating quantum statistical-mechanical properties of polyatomic systems, only recently has attention been turned to the development of computationally tractable methods for evaluating path integrals.[19] In this section we introduce some new methods that we are developing for the rapid and accurate evaluation of path integrals and we outline the procedure we are using to construct vibrational lineshapes from the quantum simulations. We have developed and initially tested the methodology on one-dimensional systems, e.g., the quartic and morse oscillators and double-well potential. These methods are being developed because their generalization to large polyatomic systems is particularly promising, and this work is now in progress.

Moment Method for Evaluating Vibrational Lineshapes from Quantum Monte Carlo Path Integral Simulations

The direct evaluation of quantum time-correlation functions for anharmonic systems is extremely difficult. Both wave packet[20] and path integral methods[21] have been used to evaluate quantum time-correlation functions directly, but because of numerical difficulties associated with these techniques it seems unlikely that these

methods will be successfully applied to large molecules. Our approach to the evaluation of finite temperature anharmonic effects on vibrational lineshapes is derived from the fact that the moments of the vibrational lineshape spectrum can be expressed as functions of expectation values of positional x_i, momentum p_j and mixed $x_i p_j$ operators. The expectation values can be evaluated using extremely efficient techniques that we have developed to evaluate equilibrium discretized path integrals. The main points are summarized here. The infrared vibrational lineshape is given as the Fourier transform of the dipole moment correlation function:

$$I(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \langle u(0)u(t) \rangle \, e^{iwt} \, dt \tag{1}$$

By inverse Fourier transformation of Equation 1 and expansion of both sides in a Taylor series we obtain[22]

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} \left[\frac{d^n}{dt^n} \langle u(0)u(t) \rangle \right] = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \int \omega^n I(\omega) \, d\omega \tag{2}$$

Equating coefficients of powers of t,

$$\frac{d^n}{dt^n} \langle u(0)u(t) \rangle_{t=0} = (i)^n \int \omega^n I(\omega) d\omega \tag{3}$$

Thus, the *n*th vibrational spectral moment is equal to an equilibrium correlation function, the *n*th derivative of the dipole moment autocorrelation function evaluated at $t = 0$.[22] By using the repeated application of the Heisenberg equation of motion

$$\frac{du}{dt} = \frac{i}{\hbar} [H, u] \tag{4}$$

and substitution in the l.h.s. of Equation 3, we can express the *n*th vibrational spectral moment as an expectation value of nested commutators of H with the dipole moment operator:

$$\int \omega^n I(\omega) \, d\omega = \hbar^{-n} \langle u(0) \cdot [H, [H, \ldots [H, u]]] \rangle \tag{5}$$

The expectation values on the right-hand side of this equation depend only on the ensemble averages of position and momentum operators, which can be evaluated using a new quantum Monte Carlo method described below:

Path integral reference systems. Our approach to the evaluation of finite temperature anharmonic effects on vibrational lineshapes requires the evaluation via computer simulation of expectation values of the moments of positional x_i, momentum p_j, and mixed $x_i p_j$ operators. These expectation values are evaluated using path integral techniques. The quantum partition function Z in the coordinate representation is:

$$Z = \int dx_1 \langle x_1 | e^{-\beta H} | x_1 \rangle \tag{6}$$

To obtain the discretized path integral representation for Z we use the identity:

$$e^{-\beta H} = (e^{-\beta H/P})^P \tag{7}$$

and insert complete sets of states $I = \int |x_i\rangle \langle x_i| dx_i$ p times:

$$Z = \int dx_1 \ldots dx_p \langle x_1|e^{-\beta H/P}|x_2\rangle \langle x_2|e^{-\beta H/P}|x_3\rangle \ldots \langle x_P|e^{-\beta H/P}|x_1\rangle \quad (10)$$

a discretized path integral with p points. It is not possible in general to evaluate the matrix elements in the path integral for arbitrary H. We define the exact Hamiltonian H and a reference Hamiltonian H_0 by:

$$H = \frac{P^2}{2m} + V \quad (11a)$$

$$H_0 = \frac{P^2}{2m} + V_0 \quad (11b)$$

where V is the exact anharmonic potential and V_0 is a quadratic "reference" potential discussed below. Clearly:

$$H = H_0 + (V - V_0) = H_0 + V' \quad (12)$$

We wish to separate the matrix elements involving H in Equation 10 into matrix elements of H_0 and V. For p "large enough" we have:

$$\langle x_i|e^{-\beta(H_0+V')}|x_{i+1}\rangle \approx e^{-\beta V1(x_1)/2P} \langle x_1|e^{-\beta H_0/P}|x_{i+1}\rangle e^{-\beta V'(x_{i+1})/2P} \quad (13)$$

A lot has been accomplished in the separation, Eq. 13, because for quadratic Hamiltonians the matrix elements can be evaluated analytically. As $p \longrightarrow \infty$, Equation 13 becomes exact. For finite, and in particular small values of p, the error depends on the magnitude of the commutator of H_0 and V'. In a p point discretized path integral Monte Carlo simulation each quantal degree of freedom is simulated by p "classical" degrees of freedom. If we wish to apply discretized path integral methods to simulate polyatomic systems, it is essential that Equation 6 hold for small values of p. The approximation of the path integral by p discretized points for small p depends on the construction of an appropriate reference Hamiltonian H_0. We discuss below three choices for H_0:

(i) The free particle reference system

$$H_0 = \frac{P^2}{2m} \quad (14a)$$

$$V' = V \quad (14b)$$

This is the standard reference system for the evaluation of path integrals. The partition function in this reference system is

$$Z = \int dx_1 \ldots dx_P \prod_{i=1}^{P} e^{-(mP/2\hbar^2\beta)(x_{i+1} - x_i)^2} e^{-\beta V(x_i)/P} \quad (15)$$

As has been pointed out, this is equivalent to a classical configurational partition function for a polymer with p beads. Adjacent beads are connected to each other by

"springs," with spring constant $K = \{mp/\hbar^2\beta\}$ and each bead interacts with an external potential $V(x_i)$.[19]

(ii) Quasiharmonic path integral reference system

We have recently proposed the use of a temperature-dependent harmonic (quasiharmonic) reference system for the evaluation of discretized path integrals of anharmonic systems.[6] For the quasiharmonic reference system:

$$H_0 = \frac{P^2}{2m} + \omega_{\text{eff}}^2 x^2 \tag{16a}$$

$$V'(x) = V(x) - \omega_{\text{eff}}^2 x^2 \tag{16b}$$

The effective frequencies and the normal modes are obtained from the classical quasiharmonic normal mode analysis described in the previous section. The quasiharmonic potential is defined so that $V'(x)$ is small in the important regions of x space sampled at temperature T. This implies that Eq. 13 will be valid for small p and that for intramolecular vibrations, the quasiharmonic reference system is superior to the free-particle reference. The harmonic propagator is given by:

$$G_0(x, x', \beta/p) = \left\{\frac{m\omega}{2\pi\hbar \sinh(\beta\hbar\omega/p)}\right\} \cdot \exp\left\{\left[\frac{-m\omega}{2\sinh(\beta\hbar\omega/p)}\right]\left[(x^2 + x'^2)\cosh\left[\frac{\beta\hbar\omega}{p}\right] - 2xx'\right]\right\} \tag{17}$$

The p point discretized path integral expression for the partition function is:

$$Z = \int dx_1 \ldots dx_P \prod_{i=1}^{P} G_0(x_i, x_{i+1}, \beta/P)\, e^{-\beta V(x_i)/P} \tag{18}$$

Defining the integrand of Eq. 10 to be $K(X_1, \ldots Xp)$, the expectation value of an operator A in the quasiharmonic discretized path integral (QHDPI) formulation is given by:

$$\langle A \rangle = \frac{\int dx_1 \ldots dx_P\, A(x_i)\, K(x_1, \ldots x_P)}{Z} \tag{19}$$

The kernal $K(X_1, \ldots Xp)$ plays the role of the Monte Carlo weight function that $\exp(-\beta H)$ plays in classical simulations.

(iii) Variable-frequency quasiharmonic reference systems

The quasiharmonic reference system described above represents an optimized fixed-frequency reference for the evaluation of discretized path integrals. It is reasonable to attempt to construct a variable frequency harmonic reference system which (1) has an analytic form for the propagator from x_i to $_{i+1}$, and (2) minimizes the value of $V'(X)$ along the most important paths from X_i to x_{i+1}. In a forthcoming publication we discuss methods for constructing variable-frequency quadratic reference systems.[7] The reference Hamiltonian is given by:

$$H_0(x_i, x_{i+1})(x) = \frac{P^2}{2m} + \omega^2(x_i, x_{i+1})\, x^2 + bx + c \tag{20}$$

The notation is meant to suggest that the frequency is variable and depends on the

propagator matrix elements. We have shown that $\omega^2(x_i, x_{i+1})$ can be defined by the equation:

$$\int_{x_i(0)}^{x_{i+1}(\beta\hbar/P)} \{V_0(x(t)) - V(x(t))\}\, dt = 0 \qquad (21)$$

where the integral is along the classical path for the reference system between x_i at time o and x_{i+1} at time $t = \beta\hbar/p$. Notice that the path depends on the temperature. We now compare the results of path integral simulations for a model quartic oscillator Hamiltonian using the free particle, quasiharmonic, and variable harmonic reference systems.

TABLE 2a.

Reference System	Exact Quantum Result	Discretized Path Integral Quadrature Points 1	4	8
	Evaluation of $\langle X^2 \rangle$			
A. $\beta\hbar\omega = 10$				
Anharmonicity b = 0.05				
Basis set	0.445			
Free particle		0.093	0.262	0.358
Quasiharmonic		0.320	0.397	0.431
Variable harmonic		0.446		
B. $\beta\hbar\omega = 2$				
Anharmonicity b = 5				
Basis set	0.161			
Free particle		0.094	0.143	0.159
Quasiharmonic		0.140	0.155	0.159
Variable harmonic		0.165		
	Evaluation of $\langle X^4 \rangle$			
A. $\beta\hbar\omega = 10$				
b = 5				
Basis set	0.071			
Free particle		0.003	0.017	0.029
Quasiharmonic		0.010	0.023	0.034
Variable harmonic		0.068		

Path integral simulations of the quartic oscillator: results for position, momentum, and spectral moments. In TABLE 2a we compare the results for the evaluation of the expectation value of $\langle x^2 \rangle$ using the different reference systems. Considering first the result for $\beta\hbar\omega = 10$ and anharmonicity b = 0.05, the exact quantum value for the second moment is $\langle x^2 \rangle = 0.445$. The classical value (which is the result for a p = 1-point quadrature discretized path integral using the free-particle reference) is 0.093. For p = 1 the quasiharmonic path integral evaluation of $\langle x^2 \rangle$ is 0.320; when the variable harmonic reference is employed the result $\langle x^2 \rangle = 0.446$ is very close to the exact value. In contrast, even for p = 8 the free particle reference is in error by 20% ($\langle x^2 \rangle = 0.358$), while the quasiharmonic reference path integral evaluation yields an improved result ($\langle x^2 \rangle = 0.431$; the error is 4%). Similar results are shown in TABLE 2a for larger values of the anharmonicity and higher temperatures. Thus for the

TABLE 2b Evaluation of Higher Moments

Property	Exact Quantum	Variable Quadratic
$\langle X^8 \rangle$	1.45	1.15 (p = 1) 1.52 (p = 2)
$\langle P^2 \rangle$	0.680	0.597 (p = 1) 0.645 (p = 2)
$\langle P^2X^2 \rangle$	−0.198	−0.175 (p = 1) −0.192 (p = 2)

evaluation of $\langle x^2 \rangle$ using the variable harmonic reference system, a 1-point quadrative is very accurate. This means that the computational effort required to calculate quantum expectation values for this model system is only slightly greater than that required to evaluate classical ensemble averages. In TABLE 2b we show the results for the path integral evaluation of higher positional moments and momentum moments. It is clear that for p = 1 or 2, the approximate values obtained using the variable harmonic reference method to evaluate the path integral are close to the exact results. Equally promising results using this new path integral method have been obtained for the double-well and morse oscillator Hamiltonians. We are presently modifying the programs to treat polyatomic model systems.

In TABLE 2c we present the first and second moments of the vibrational spectrum of the quartic oscillator calculated by the moments method. The quantum results were obtained from path integral Monte Carlo simulations using the variable quadratic reference. For comparison, the average frequency and linewidth obtained from classical Monte Carlo evaluation of the moments is also listed. Spectral features for two values of the temperature ($\beta\hbar\omega = 5$ and 1.0) and two anharmonicities (b = 0.05, b = 1.0) are listed. As the anharmonicity or the temperature is increased, the oscillator frequency and linewidth increase for both the classical and quantum simulations. The important point is that the classical spectrum is shifted less and broadened more as the temperature and anharmonicity increase. For example, when $\beta\hbar\omega = 1$ and b = 0.05 are chosen, the classical spectrum is almost twice as broad as the quantum spectrum. The results presented in TABLE 2c demonstrate that for realistic values of the temperature and anharmonicity, quantum effects on the vibrational spectrum are important. However, the strong coupling of broadening with frequency shifts is a limitation of one-dimensional models for which the only broadening mechanism is anharmonicity in the vibrational degree of freedom. Significant population of vibrationally excited states is required for thermal broadening. In contrast, it is the coupling

TABLE 2c. Vibrational Spectrum of Quartic Oscillator by Moments Method

	Spectral Moment[a]	Quantum	Classical
A. $\beta\hbar\omega = 5$	$\langle \omega_1 \rangle$	1.12	1.05
b = 0.05	$\langle \omega^2 \rangle^{1/2}$	0.03	0.06
B: $\beta\hbar\omega = 5$	$\langle \omega \rangle$	1.96	1.4
b = 1.0	$\langle \omega^2 \rangle^{1/2}$	0.3	0.4
C. $\beta\hbar\omega = 1$	$\langle \omega \rangle$	1.19	1.12
b = 0.05	$\langle \omega^2 \rangle^{1/2}$	0.12	0.20

[a] $\omega_0 = 1$, $\hbar = 1$, m = 1.

of the vibrational degree of freedom to other modes, protein and/or liquid that is of primary interest with respect to an understanding of the structural information content of chromophore lineshapes. The coupling of a chromophore vibration to a bath results in the time and spatial modulation of the energy spacing of the first few vibrationally excited states. This broadening mechanism can give rise to line broadening without significant frequency shifting. Path integral computer simulations of models for chromophoric molecules coupled anharmonically to additional modes, including heat bath models, are presently under way.

We have described in this section methods for calculating vibrational lineshapes for anharmonic systems which incorporate quantum properties in a fundamental way. The methods have been demonstrated to be very powerful when applied to a variety of one-dimensional problems. However, for one-dimensional models alternative methods for calculating spectra are more direct. We are pursuing path integral approaches to the problem of calculating lineshapes because we believe that there numerical methods can be generalized to polyatomic systems and that they will be much more stable and accurate than alternative approaches. One particularly attractive aspect of our approach is the ability to combine quantum and classical Monte Carlo algorithms in a single simulation. For example the simulation of the vibrational spectrum of a chromophore with 10–20 degrees of freedom imbedded in a protein could be accomplished by combining quantum Monte Carlo methods for the chromophore coordinates with classical Monte Carlo trajectories for the protein atoms.

Restrained X-Ray Refinement of Nucleic Acids and Proteins

Crystallographic refinement of proteins and nucleic acids at high resolution is being used to obtain a wealth of information concerning the atomic fluctuations and the flexibility of these macromolecules. The information concerning atomic mobility is contained in the Debye-Waller temperature factors. In proteins, correlations between temperature factors and biological functions, like ligand access to the protein interior[23,24] or antigenic recognition,[25,26] have been noted. The availability of high-resolution X-ray structures of DNA oligomers has opened up a new era in the study of the structure and biological function of DNA and its interactions with proteins.[27,28] The goal is to determine the roles of DNA structure and flexibility in the specificity of DNA interactions with ligands. The correct interpretation of high-resolution crystallographic data depends upon a thorough understanding of the effects of the refinement model upon derived quantities such as temperature factors. Temperature-dependent crystallographic studies can, in principle, provide detailed information concerning the potential surface on which the atoms move.[23,29,30] For these studies, it is necessary to determine the accuracy of the refinement model over a wide range of temperatures.

Molecular dynamics simulations of proteins and nucleic acids provide a very powerful method for testing crystallographic refinement models. The simulations constitute the most detailed theoretical approach available for studying the internal motions of these macromolecules.[31] From the time evolution of the atomic positions, time-averaged X-ray intensities can be calculated and treated as data for crystallographic refinement. The final structure and temperature factors obtained from the refinement can then be compared with the "exact results" obtained directly from the trajectory. Kuriyan *et al.*[8] have recently carried out a detailed analysis of a protein

refinement model at a single temperature using a molecular dynamics simulation of myoglobin[32] to generate the X-ray data. In this review, we discuss the results of an analysis[9] of the temperature-dependent molecular dynamics and X-ray refinement of a Z-DNA hexamer 5BrdC-dG-5BrdC-dG-5BrdC-dG for which the experimental X-ray data are available and whose crystal structure has been refined to high resolution.[33]

Methods for Simulating Restrained X-Ray Refinement Data from Molecular Dynamics Trajectories

Molecular dynamics simulations were carried out on the 248 atom Z-DNA hexamer, using the AMBER nucleic acid force field[34] with a distance-dependent dielectric and excluding counterions. Although the model treats electrostatic effects only in a qualitative way, recent molecular dynamics simulations for both peptides[35] and nucleic acids[36,37] have demonstrated that for localized conformations sampled during short molecular dynamics simulations, average properties are not very sensitive to the electrostatic model; it is the packing and hydrogen-bonding terms that together with the vibrational potential (bond, bond angle and torsional stretching) dominate the calculated equilibrium and dynamical properties. The crystal structure of the Z-DNA hexamer was first energy-minimized with 200 conjugate gradient steps to relieve any initial strain in the structure before the molecular dynamics simulations were begun. The rms displacement between the crystal and energy-minimized coordinates is less than 0.1 Å. Simulations were performed at a series of temperatures, defined by the mean kinetic energy of the system, between 100 K and 300 K. For each temperature, 10 trajectories, each 2 psec in length, were calculated by solving simultaneously the classical equations of motion for the atoms of the helix. The use of multiple short trajectories instead of a single long trajectory has been found to be a more efficient method for sampling conformations for macromolecular systems containing many harmonic degrees of freedom.[38,39]

Crystallographic refinement is a procedure that iteratively improves the agreement between structure factors derived from X-ray intensities and those derived from a model structure. For macromolecular refinement, the limited diffraction data have to be complemented by additional information in order to improve the parameter-to-observation ratio. This additional information consists of restraints on bond lengths, bond angles, aromatic planes, chiralities, and temperature factors.

In the restrained refinement procedure[40] a function of the form:

$$\Phi = \sum_{Q} W_Q \left| |F_0(Q)| - |F_c(Q)| \right|^2 + \sum_{i} W_i \Delta_i^2 \qquad (22)$$

is minimized. W_Q is the weight assigned to the structure factors and it varies linearly with Q with coefficients adjusted so that low-resolution structures are weighted more strongly than high-resolution ones. $F_0(Q)$ and $F_c(Q)$ are, respectively, the observed and calculated structure factors. The second term in Equation 1 contains the stereochemical restraint information. Δ is the deviation of a restrained parameter (bond lengths, bond angles, volumes, nonbonded contacts, and temperature factors) from its ideal value and W_i is the weight assigned to the restraint. The form of Equation 1 is such that the weights W_i correspond to the inverse of the variance Δ^2 for each set of observations. The weights on the various classes of restraints in the simulated refinements described

below were between 0.02 and 0.04Å for distances, 0.01Å for planarity, 0.04Å for chirality, and 0.25Å for nonbonded contacts. The refinements were carried out with no and tight (less than $0.1Å^2$ at 165 K or less than $1.0Å^2$ at 300 K) restraints on the temperature factors.[9]

The structure factor F(Q) in X-ray crystallography is the Fourier transform of the electron density for the molecule:

$$F(Q) = \int dr\, \rho(r)\, e^{iQ\cdot r} \tag{23}$$

where $\rho(r)$ is the electron density at r. In a crystallography experiment the electron density varies with time due to thermal motion and the observed structure factor amplitude is the time average of Equation 2:

$$F_0(Q) = \langle F(Q) \rangle = \int dr\, \langle \rho(r) \rangle\, e^{iQ\cdot r} \tag{24}$$

In order to generate a set of calculated structure factors $F_c(Q)$ from a set of coordinates it is necessary to introduce a model for the time variation of the electron density. The usual assumptions in macromolecular crystallography include harmonic isotropic motion of the atoms, and, in addition, the molecular scattering factor is expressed as a superposition of atomic scattering factors. With these assumptions the calculated structure factor (Equation 1) is given by[42]:

$$F_c(Q) = \sum_{j=1}^{N} e^{iQ\cdot r}\, e^{W_j(Q)} \tag{25}$$

where $F_j(Q)$ is the atomic scattering factor for atom j and r_j is the position of atom j in the model structure. The thermal averaging of atomic motion is contained in the atomic Debye-Waller factor exp $(W_j(Q))$. W_j is given by:

$$W_j(Q) = -B_j |Q|^2 \tag{26}$$

where B_j is the atomic temperature factor. The mean square atomic fluctuation $(\Delta r_j)^2$ for atom j is obtained from the refined temperature factors through the relation[42]:

$$\langle (\Delta r_j)^2 \rangle = \frac{3}{8\pi^2} B_j \tag{27}$$

There are therefore four adjustable parameters per atom in the refinement (x_j, y_j, z_j, B_j). In the computer experiments we have carried out to test the assumptions of the nucleic acid refinement model we have generated sets of "observed" structure factors $F_0(Q)$, from the Z-DNA molecular dynamics trajectories. The thermal averaging implicit in Equation 26 is accomplished by averaging the atomic structure factors obtained from coordinate sets sampled along the molecular dynamics trajectories at each temperature:

$$F_0(Q) = \langle F(Q) \rangle = \frac{1}{m} \sum_{k=1}^{M} \sum_{j=1}^{N} F_j(Q)\, e^{iQ\cdot r_j^k} \tag{28}$$

where r_j^k is the position of the *j*th atom in the *k*th coordinate set along the trajectory and M is the number of coordinate sets sampled. In the present study structure factors

corresponding to 3195 reflections between 10Å and 1.7Å were calculated for each of 50 coordinate sets at each temperature. Only the 246 heavy atoms of the hexamer were included in the structure factor calculations; hydrogen atoms were not included in the refinement.

Restrained X-Ray Refinement of the Z-DNA Molecular Dynamics Trajectories

Refinement of the MD average structures against X-ray intensities calculated from the trajectories. Refinement of molecular dynamics average structures against simulated X-ray diffraction intensities was carried out at four temperatures between 165 K

TABLE 3. Parameters for Refinement of Simulated X-Ray Intensities of Z-DNA Hexamer

Parameter	165 K	300 K
Number of bad distances[a]		
Before starting refinement	1	50
After refinement with strong B restraints	0	76
After refinement without B restraints	0	50
R-factor		
Before starting refinement	1.6	22.0
After refinement with strong B restraints	1.6	7.0
After refinement without B restraints	0.9	6.2
Average temperature factor (A^2)		
Before starting refinement	0.5	5.9
After refinement with strong B restraints	0.7	5.1
After refinement without B restraints	0.5	5.5
RMS of delta B[b]		
Before starting refinement	0.3, 0.3, 0.4, 0.3	2.4, 3.5, 8.2, 5.6
After refinement with strong B restraints	.06, .08, .10, .13	0.6, 0.7, 0.2, 0.6
After refinement without B restraints	0.3, 0.3, 0.3, 0.3	3.0, 3.5, 3.8, 4.1

[a]Distances that deviate from ideality by more than two standard deviations.

[b]The four values correspond to the differences in temperature factors for atoms connected by a bond length, for atoms connected by a bond angle, for P—O bond lengths, and for phosphate atoms connected by a bond angle or for atoms involved in hydrogen bonding.

and 300 K with no restraints and with strong restraints on temperature factors. For the initial temperature factors used to start the refinement we tried both individual atomic temperature factors calculated from the dynamics and a single average temperature factor obtained by averaging the molecular dynamics results over all atoms at each temperature. Similar results were obtained for both choices of initial temperature factors. The results reported below correspond to refinements started with individual atomic temperature factors taken from the molecular dynamics results.

The R factors, average temperature factors, rms deviations of the temperature factors, and the number of "bad" distances obtained for the refinement of the molecular dynamics average structures against the simulated X-ray intensities at 165 K and 300 K are listed in TABLE 3. At 165 K the initial R factor and number of bad

distances before refinement are both very small and are not changed significantly after refinement. At 300 K the initial R factor and number of bad distances are 22% and 50, respectively. With tight restraints on temperature factors, the R factor decreased to 7% after three refinement cycles and the number of bad distances increased to 76. With no restraints on B, the R factor decreased to 6.3% after three refinements cycles with 50 bad distances.

At the low temperature (165 K) the effect of refinement with strong temperature factor restraints is to increase the average temperature factor from 0.5 $Å^2$ (exact result) to 0.7 $Å^2$ and to decrease the variances in the temperature factors for the different classes of stereochemical constraints. With strong restraints on B, the refinement resulted in sharp differences between the cytosines and the guanines both for the sugars and bases which were not present in the temperature factors calculated directly from the 165 K trajectories. For example, the ratio of the temperature factors for guanine bases to cytosine bases, which is 1.1 calculated directly from the molecular dynamics simulation, increases to 2.6 after refinement with strong B restraints. In contrast, when refinement is done without temperature factor restraints at 165 K, the average temperature factors and the variances in the temperature factors are very close to the exact molecular dynamics values.

At 300 K the effect of the refinement both with and without strong restraints on temperature factors is to decrease the average thermal factor compared to the exact values. The temperature factor averaged over all atoms calculated directly from the room temperature trajectories is 5.9 $Å^2$ and is reduced by 15% to 5.1 A after refinement with strong B restraints. The average temperature factor (5.5 $Å^2$) obtained from the 300 K refinement without temperature factor restraints is closer to the exact value. The errors in the temperature factors introduced by the refinement at 300 K is largest for the atoms with the largest thermal fluctuations, the phosphates. This result is clearly demonstrated in FIGURE 2, which compares the phosphate temperature factors calculated directly from 300 K trajectories with the results of the two refinements, with and without temperature factor restraints. For example, the two atoms with the largest thermal fluctuations are the C5 and C11 phosphates. The exact B values for these atoms are 15.1 $Å^2$ and 17.0 $Å^2$, respectively; after the refinement without temperature factor restraints the B values are reduced to 10.8 $Å^2$ and 10.9 $Å^2$, and they are reduced even further to 8.8 $Å^2$ and 8.3 $Å^2$, respectively, after the refinement with temperature factor restraints. The errors introduced by the refinement are also apparent in the effect on the temperature factor variances. At room temperature the actual variances in temperature factors computed for each stereochemical class are considerably greater than 1 $Å^2$ (they range from 2.4 $Å^2$ for atoms connected by bonds not involving a phosphorous to 8.2 $Å^2$ for P and O atoms connected by a P—O bond). The refinement of the simulated X-ray intensities at 300 K with temperature factor restraints greatly reduced the variances in the B values for each of the stereochemical classes so that the final variances are less than 1 $Å^2$ (TABLE 3). These results are in accord with a recent analysis of the use of restraints in temperature factor refinements for proteins[8] for which it was shown that the weights used to restrain neighboring atom temperature factors are much more restrictive than the variation in neighboring temperature factor values obtained from protein molecular dynamics simulations. Restraints on the differences in temperature factors between bonded atom pairs have been shown to be uncorrelated with variances in the corresponding bond

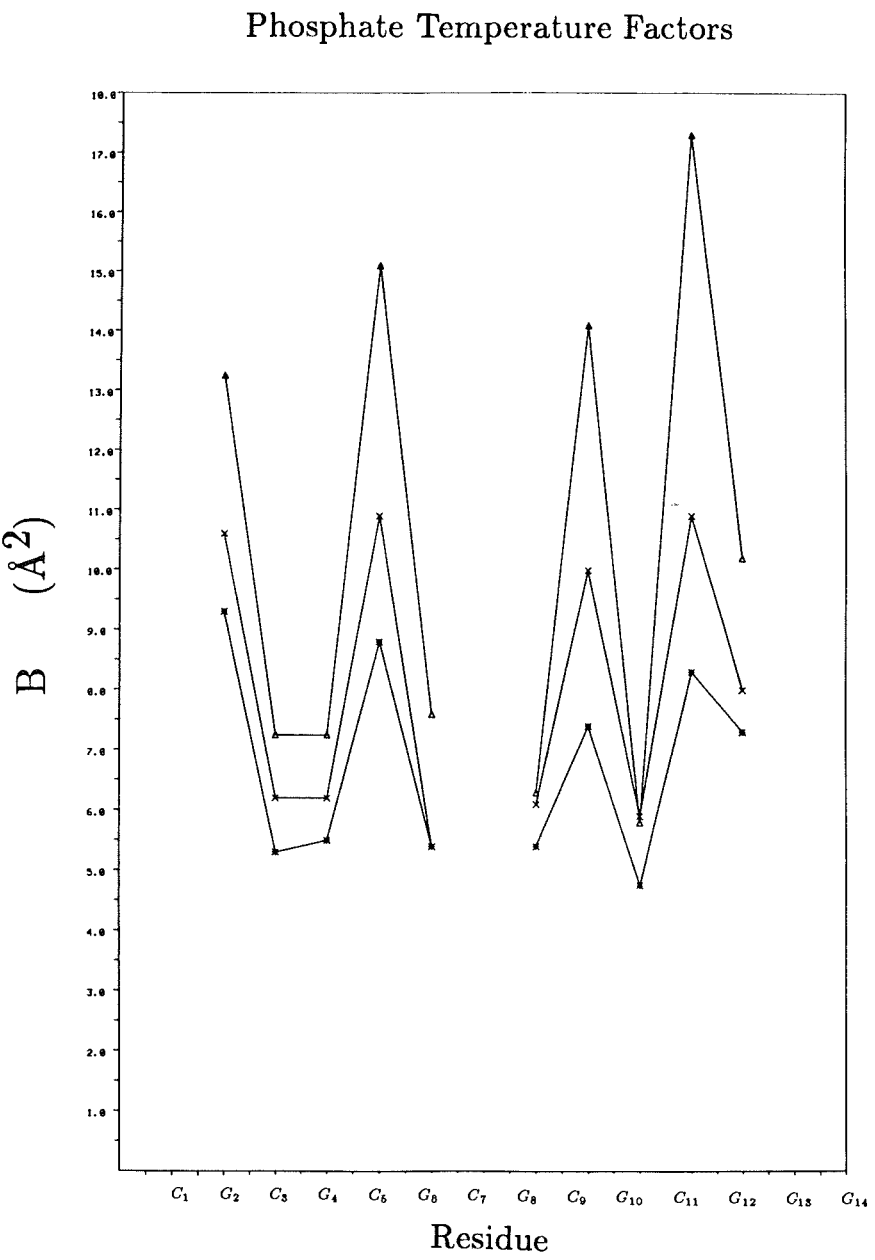

FIGURE 2. Temperature factors for the phosphates of the Z-DNA hexamer at 300 K. Temperature factors were evaluated using the mean square atomic fluctuations calculated directly from the molecular dynamics trajectories; x = temperature factors calculated from the refinement without B-factor restraints; * = temperature factors calculated from the refinement with B-factor restraints.

length distributions.[43] From the present results concerning errors in predicted temperature factor restraints we conclude that the commonly used value of 1 $Å^2$ or less between temperature factors on adjacent atoms[9,44] is too restrictive.

Although, as discussed earlier, there are quantitative errors in temperature factors introduced by the refinement procedure, the temperature-dependence of the atomic mobilities as estimated by the refined temperature factors provides a reasonably accurate description of the true temperature-dependence of the system. In FIGURES 3b and c the mean square atomic fluctuations extracted from the refinement simulations at each temperature and averaged by group are plotted as a function of temperature for comparison with the exact results shown in FIGURE 3a. As to the refinement without temperature factor restraints (Fig. 3b), except for the phosphates at the highest temperature, the extent of anharmonicity (curvature) is in good agreement with the exact result despite the fact that the refinement model assumes isotropic, harmonic motion. The ordering of the atomic fluctuations by groups (bases < deoxyriboses < CpG phosphates < GpC phosphates) at the two higher temperatures (275 K and 300 K) agrees with the exact results, although the agreement is not as good as that at 165 K and 225 K, even though the harmonic model would be expected to be more accurate at low temperature. It is clear from FIGURE 3c that when strong temperature factor restraints are introduced in the refinement, differences in the temperature-dependence of the atomic fluctuations among the groups are suppressed, although the shapes of the curves agree qualitatively with the results calculated directly from the simulations (Fig. 3a). The present results provide a theoretical foundation for the use of Debye-Waller factors obtained from refinements at several temperatures to extract information concerning the anharmonicity of the atomic displacements and underlying potential surface.

SUMMARY

The use of computer simulations to study the internal dynamics of globular proteins and nucleic acids has attracted a great deal of attention in recent years. These simulations constitute the most detailed theoretical approach available for studying internal motions and structural flexibility of biological macromolecules. In this paper we review recent work concerned with the use of computer simulations for the interpretation of, and comparison with, experimental probes of molecular dynamics. New methods for calculating vibrational spectroscopic lineshapes from computer simulations are discussed. A quasiharmonic approximation is described by which classical computer simulations on multidimensional potential surfaces can be used to estimate the effects of anharmonicity on vibrational spectra. A novel aspect of the method is the use of normal-mode eigenvectors as the independent coordinates for Monte Carlo sampling. Results for isolated small molecules and liquid water are reviewed. The construction of vibrational lineshapes from quantum computer simulations using path integral methods with new (quasiharmonic and variable quadratic) reference systems are discussed. Results for small model systems are presented and extension of the methods to large molecules is discussed. The recent use of molecular dynamics simulations to analyze X-ray refinement models for proteins and nucleic acids is also reviewed.

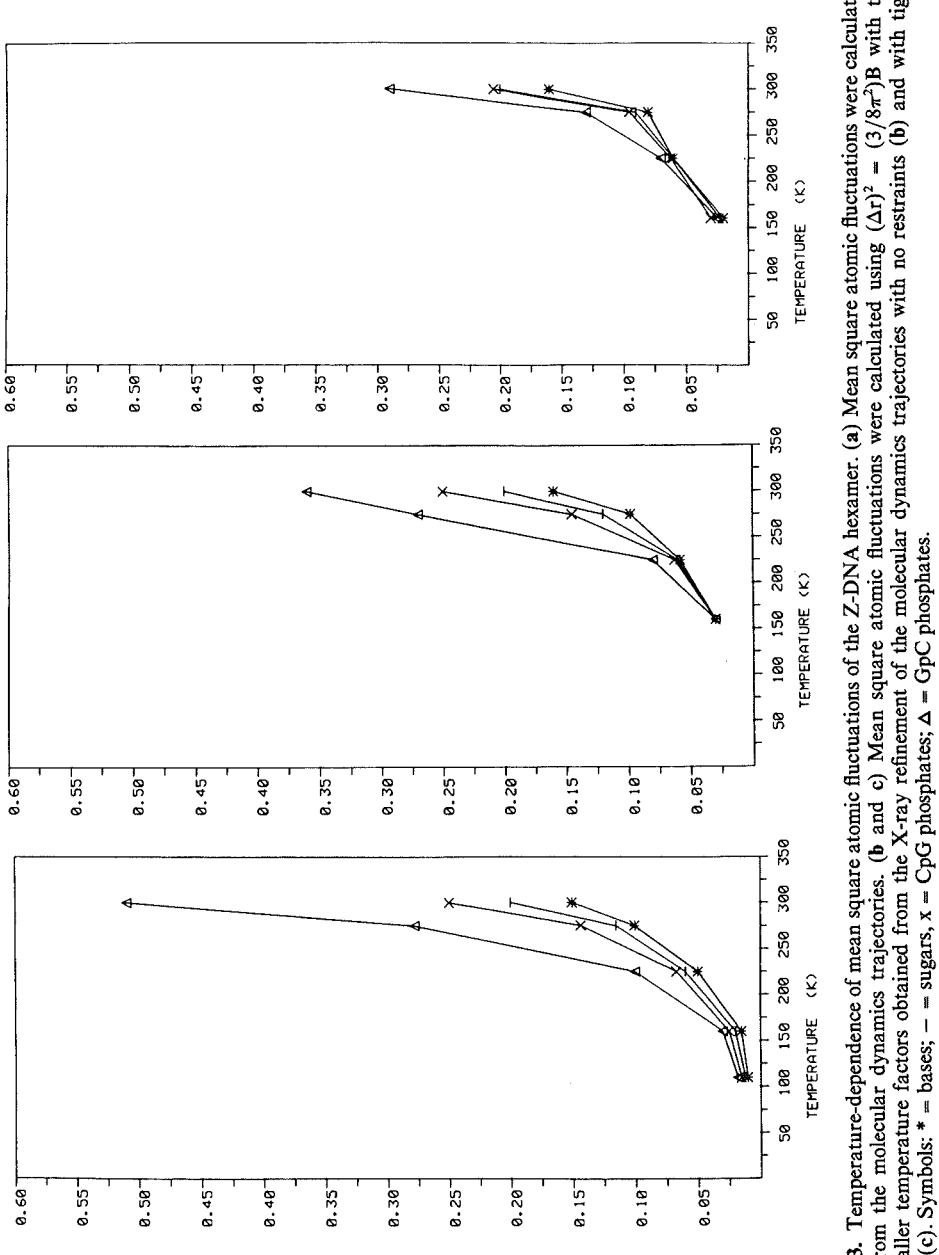

FIGURE 3. Temperature-dependence of mean square atomic fluctuations of the Z-DNA hexamer. (**a**) Mean square atomic fluctuations were calculated directly from the molecular dynamics trajectories. (**b** and **c**) Mean square atomic fluctuations were calculated using $(\Delta r)^2 = (3/8\pi^2)B$ with the Debye-Waller temperature factors obtained from the X-ray refinement of the molecular dynamics trajectories with no restraints (**b**) and with tight restraints (**c**). Symbols: * = bases; – = sugars, x = CpG phosphates; Δ = GpC phosphates.

ACKNOWLEDGMENTS

I am grateful to Richard Friesner and Attila Szabo for collaboration and discussion concerning the quasiharmonic method and to Eric Westhof, John Kuriyan, Martin Karplus, and Greg Petsko for collaboration and discussion concerning X-ray refinement simulations.

REFERENCES

1. KARPLUS, M. & J. A. MCCAMMON. 1981. CRC Crit. Rev. Biochem. **9:** 293.
2. MCCAMMON, J. A. 1984. Rep. Prog. Phys. **47:** 1.
3. ALLISON, S. A. & J. A. MCCAMMON. 1985. J. Phys. Chem. **89:** 1072.
4. BRUNGER, A., C. L. BROOKS & M. KARPLUS. 1985. Proc. Natl. Acad. Sci. USA **82:** 8458.
5. LEVY, R. M. & J. KEEPERS. 1986. Comments on Molecular and Cellular Biophys. **3:** 273.
6. FRIESNER, R. A. & R. M. LEVY. 1984. J. Chem. Phys. **80:** 4488.
7. ZHANG, P., R. M. LEVY & R. A. FRIESNER. Submitted for publication.
8. KURIYAN, J., G. A. PETSKO, R. M. LEVY & M. KARPLUS. 1986. J. Mol. Biol. **90:** 227.
9. WESTHOF, E., S. GALLION, P. WEINER & R. M. LEVY. 1976. 1986. J. Mol. Biol. **91:** in press.
10. WARSHEL, A. J. Chem. Phys. 1971. **55:** 3327.
11. WARSHEL, A. & M. KARPLUS. 1972. Chem. Phys. Lett. **17:** 7.
12. MYERS, A., R. MATHIES, S. TANNOR & E. J. HELLER. 1982. J. Chem. Phys. **77:** 3857.
13. LEVY, R. M., O. ROJAS & R. A. FRIESNER. 1984. J. Phys. Chem. **88:** 4233.
14. NAKAI, S., F. HIRATA & R. M. LEVY. To be published.
15. JORGENSEN, W. 1981. J. Am. Chem. Soc. **103:** 335.
16. BERENDSEN, H., J. P. M. POSTMA, W. F. VAN GUNSTEREN & J. HERMANS. 1981. *In* Intermolecular Forces. B. Pullman, Ed.: 331. Reidel. Dordrecht, Holland.
17. RAHMAN, A. & F. STILLINGER. 1975. J. Chem. Phys. **12:** 5223.
18. REIMERS, J. R. & A. O. WATTS, 1984. Chem. Phys. **91:** 201.
19. CHANDLER, D. & P. G. WOLYNES. 1981. J. Chem. Phys. **74:** 4078.
20. HELLER, E. 1981. Accts. Chem. Res. **14:** 368.
21. THIRUMOLAI, D. & B. BERNE. 1983. J. Chem. Phys. **79:** 5029.
22. GORDON, R. G. 1965. J. Chem. Phys. **43:** 1307.
23. FRAUENFELDER, H., G. PETSKO & D. TSERNOGLU. 1979. Nature **280:** 558.
24. ARTYMIUK, P. J., C. BLAKE, D. GRAU, S. OATLEY, D. C. PHILLIPS & M. STERNBERG. 1979. Nature **280:** 563.
25. WESTHOF, E., D. ALTSCHUH, A. C. BLOOMER, A. MONDRAGON, A. KLUG & M. H. VAN REGENMORTEL. 1984. Nature **311:** 123.
26. TAINER, J., E. D. GETZOFF, H. ALEXANDER, R. A. HOUGHTEN, A. J. OLSON, A. J. LERNER & W. A. HENDRICKSON. 1984. Nature **312:** 127.
27. A. H.-J., G. J. QUIGLEY, F. J. KOLPAK, J. L. CRAWFORD, J. H. VAN BOOM, G. VAN DER MAREL & A. RICH. 1979. Nature **282:** 680.
28. WING, R. M., H. R. DREW, T. TAKANO, C. BROKA, S. TANAKA, K. ITAKURA & R. E. DICKERSON. 1980. Nature **287:** 755.
29. HARTMANN, H., F. PARAK, W. STEIGMAN, G. A. PETSKO, P. D. RINGE & H. FRAUENFELDER. 1982. Proc. Natl. Acad. Sci. USA **79:** 4967.
30. KNAPP, E. W., S. F. FISCHER & F. PARAK. 1982. J. Phys. Chem. **86:** 5042.
31. KARPLUS, M. & J. A. MCCAMMON. 1981. CRC Crit. Rev. Biochem. **9:** 293.
32. LEVY, R. M., R. P. SHERIDAN, J. W. KEEPERS, G. S. DUBEY, S. SWAMINATHAN & M. KARPLUS. Biophys. J. 1985. **48:** 509.
33. WESTHOF, E., T. PRANGE, B. CHEVRIER & D. MORAS. 1985. Biochimie. **67:** 811.
34. WEINER, S., P. KOLLMAN, D. CASE, U. SINGH, G. GHIO, G. ALAGON, S. PROFETA & P. WEINER. 1984. J. Am. Chem. Soc. **106:** 765.
35. PETTIT, J. & M. KARPLUS. 1985. J. Am. Chem. Soc. **107:**1166.
36. LEVITT, M. 1982. Cold Spring Harbor Symp. Quant. Biol. **47:** 251.

37. SINGH, U. C., S. J. WEINER & P. KOLLMAN. 1985. Proc. Natl. Acad. Sci. USA **82:** 755.
38. LEVY, R. M., D. PERAHIA & M. KARPLUS. 1982. Proc. Natl. Acad. Sci. USA **79:** 1346.
39. KEEPERS, J. & R. M. LEVY. To be published.
40. KONNERT, J. H. & W. A. HENDRICKSON. 1980. Acta Crystallog. **A36:** 344.
41. WARSHEL, A. 1977. Ann. Rev. Biophys. Eng., **6:** 273.
42. WILLIS, B. T. & W. PRYOR. 1975. Thermal Vibrations in Crystallography. University Press. London.
43. YU, H., M. KARPLUS & W. A. HENDRICKSON. 1985. Acta. Crystallog. **B41:** 191.
44. WESTHOF, E., P. DUMAS & D. MORAS. 1985. J. Mol. Biol. **184:** 119.