Normal mode analysis using the driven molecular dynamics method. II. An application to biological macromolecules

Martina Kaledin, Alex Brown,[a] Alexey L. Kaledin, and Joel M. Bowman[b]
Cherry L. Emerson Center for Scientific Computation and Department of Chemistry, Emory University, Atlanta, Georgia 30322

(Received 9 April 2004; accepted 8 June 2004)

The driven molecular-dynamics (DMD) method, recently proposed by Bowman, Zhang, and Brown [J. Chem. Phys. **119**, 646 (2003)], has been implemented into the TINKER molecular modeling program package. The DMD method yields frequencies and normal modes without evaluation of the Hessian matrix. It employs an external harmonic driving term that can be used to scan the spectrum and determine resonant absorptions. The molecular motions, induced by driving at resonant frequencies, correspond to the normal-mode vibrations. In the current work we apply the method to a 20-residue protein, Trp-cage, that has been reported by Neidigh, Fesinmeyer, and Andersen [Nature Struct. Biol. **9**, 425 (2002)]. The structural and dynamical properties of this molecule, such as *B*-factors, root-mean square fluctuations, anisotropies, vibrational entropy, and cross-correlations coefficients, are calculated using the DMD method. The results are in very good agreement with ones calculated using standard normal-mode analysis method. Thus, the DMD method provides a viable alternative to the standard Hessian-based method and has considerable potential for the study of large systems, where the Hessian-based method is not feasible. [DOI: 10.1063/1.1777573]

I. INTRODUCTION

Important progress in developing new scientific methods and technologies that integrate physical, chemical, and biological approaches with information science, mathematics, and computational science has been made. In chemistry, computational simulations are a frequent companion to experimental results. Whether used for predicting the structure or the properties of molecular systems, calculations are indispensable for verifying experiments and generating unexpected insight. Nowadays highly sophisticated and accurate computational methods are commonly used to study the structure and reactions of small molecules. The application of these methods to investigate biological macromolecules (e.g., proteins and nucleic acids) faces enormous challenges due to the huge computational demands. Knowledge of the atomic motions and their collective or correlated character in proteins plays an important role in understanding of their biological functions,[1] and thus approximate computational methods are used to simulate the protein dynamics.

One of the principal tools in the theoretical study of biological molecules is the method of molecular dynamics (MD) simulations. This computational method predicts the time dependent behavior of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics, and thermodynamics of biological molecules and their complexes. Limitations of these methods are imposed by the approximate nature of the force fields, the absence of quantum effects, and the limited time scale (order of nanoseconds).[1]

The other major approximate technique is normal-mode analysis (NMA). This method has long been used as a tool for interpreting vibrational spectra of small molecules.[2,3] The frequencies obtained from NMA can be directly related to experimental infrared and/or Raman measurements, and the derived normal modes can be used in characterizing the dynamic behavior of molecules. Advances in computational technology over the last few decades have made normal-mode analysis of proteins and other large molecules practical.[4,5] Although NMA is a quantum method, it is approximate, because only the harmonic motion of the system around a single potential minimum is taken into account. Qualitative and semiquantitative estimates can be made for many properties[6] of macromolecules such as the magnitude of atomic fluctuations, displacement covariance matrix, vibrational entropy, etc.[7]

The standard normal-mode analysis requires the calculation of a mass-weighted second derivative matrix (Hessian) followed by the diagonalization of that matrix. The order of the Hessian matrix is $3N$, where N is the number of atoms in the molecular system. There are a number of bottlenecks associated with its application to large systems, containing thousands of atoms. In particular, the calculation and storage of the Hessian scale quadratically with the size of the system. The diagonalization scales as the cube of the dimension of the Hessian. Thus, for large molecules the storage and diagonalization of the Hessian can quickly exceed the memory of a typical workstation.

[a]Permanent address: Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2, Canada.
[b]Author to whom correspondence should be addressed. Electronic mail: bowman@euch4e.chem.emory.edu

Awareness of the limitations of the Hessian-based normal-mode analysis for large molecular systems resulted in the development of several approaches designed to overcome the difficulties with the storage and diagonalization of the Hessian matrix. In proteins, the low-frequency normal modes that correspond to the large-scale conformational change are often of central interest. Previous studies reveal the success of approximate coarse-grained protein models[8,9] that are useful in characterizing large-scale cooperative motions in systems composed of more than several thousand residues. This approach is based on a block normal-mode algorithm that projects the Hessian matrix into local translation/rotation basis vectors and, therefore, reduces the size of the matrix involved in diagonalization.

In our previous paper,[10] we proposed a driven molecular dynamics (DMD) approach that can be used to obtain normal modes without evaluation of the Hessian matrix. We verified the method on HOD and $H_5O_2^+$, and discussed its computational aspects and its potential application for large molecular systems.[10] In this work, we apply and test the DMD method to a small protein Trp-cage.[11] We calculated the structural and dynamical properties of this molecule, such as *B*-factors, root-mean square fluctuations, vibrational entropy, anisotropies, and cross-correlation coefficients using the normal modes and frequencies obtained from the DMD simulation. The results are compared to those calculated using standard normal-mode analysis method.

In the following sections, we briefly review the algorithm of the DMD approach and describe our implementation in the TINKER program.[12] Then we discuss the parameters used for the DMD simulations, their meaning, and impact on protein properties, as well as performance and accuracy of the DMD method.

II. METHOD

As described in a previous work,[10] the principle of the DMD method is to employ an external harmonic driving term to scan the frequency spectrum and determine resonant absorptions, which under mild driving equal to the normal-mode frequencies. The molecular motions, induced by driving at resonant frequencies, correspond to the normal-mode vibrations. The Hamiltonian of a molecular system consisting of N atoms used in DMD simulation is given by

$$H(\mathbf{p},\mathbf{q},t)=H_0(\mathbf{p},\mathbf{q})+U(t), \tag{1}$$

where $\mathbf{q}$ and $\mathbf{p}$ represent the $3N$ atomic Cartesian coordinates and momenta, respectively. The molecular Hamiltonian H_0 in this equation is given by

$$H_0(\mathbf{p},\mathbf{q})=\sum_{i}^{3N}\frac{p_i^2}{2m_i}+V(\mathbf{q}). \tag{2}$$

The driving term $U(t)$ depends only on internuclear distances r_{ij} and is given by

$$U(t)=\sum_{i,j}\lambda_{ij}r_{ij}\sin(\omega_n t), \tag{3}$$

where λ_{ij} are the coupling constants. Hamilton's equations of motion are as follows:

$$\frac{\partial H}{\partial p_{\alpha,i}}=\dot{q}_{\alpha,i}=\frac{p_{\alpha,i}}{m_i},$$
$$-\frac{\partial H}{\partial q_{\alpha,i}}=-\dot{p}_{\alpha,i}=-\frac{\partial V}{\partial q_{\alpha,i}}-\sum_j\lambda_{ij}\frac{\alpha_i-\alpha_j}{r_{ij}}\sin(\omega_n t), \tag{4}$$
$$i=1,...,N,\quad \alpha=x,y,z.$$

The absorption measure that we adopt is the average total internal energy of the molecule after a finite time of driving, i.e.,

$$\langle E\rangle=\frac{1}{t}\int_0^t H_0(\tau)d\tau. \tag{5}$$

At nonresonant frequencies, the absorbed energy is small and oscillatory with time, while on resonance it increases rapidly with time. The absorption energy is also a good measure to estimate the stability of the calculation with respect to the driving strength. This issue will be discussed in further detail in the following section.

The computational implementation of this method is straightforward. The driving term and its derivative, which are easy to evaluate, can be readily incorporated into any molecular dynamics simulation program. We have done this with the program TINKER.[12] This program provides an excellent basis for molecular mechanics and dynamics with some special features to simulate biopolymers. It has the ability to use several common force field parameter sets and it includes a variety of novel algorithms for geometry optimization, potential surface scanning, normal-mode analysis, solvation effects, etc. For simulation involving biopolymers, such as proteins and nucleic acids, the initial positions for the atoms of the biopolymer are usually obtained from a known x-ray structure or NMR spectrum. If the molecular structure is obtained, for example, in the PDB format (Protein Data Bank),[13] this can be easily transformed into the Cartesian coordinates using the pdbxyz program of TINKER. Then the structure is usually refined using an energy minimization algorithm with the particular potential function (force field) that is employed.

We implemented the driving term into the velocity Verlet integration program of TINKER. The scanning procedure to obtain the normal-mode frequencies and corresponding normal-mode vibrations we adopted is straightforward. The system is initially at rest at a stationary point, and the driving force at frequency ω_n is weak enough to be harmonic. During the simulation the absorption energy [Eq. (5)] is monitored to identify a resonant frequency. For each frequency a single trajectory is propagated for picoseconds. To date, the available simulations of proteins and nucleic acid systems range in length from picoseconds to nanoseconds.[1] If the frequencies alone are known, e.g., from a spectral analysis of the velocity autocorrelation function, these can be used directly with DMD to obtain the normal modes by driving exactly at the known frequencies. Otherwise, the DMD normal modes for the resonant frequencies are calculated from the mass-weighted coordinates obtained any time in the trajectory, after significant absorption of energy has occurred.

FIG. 1. Trp-cage protein (PDB entry 1L2Y). In the structure the backbone trace is shown as a tube.

III. RESULTS

To test the efficiency and accuracy of the DMD approach, we performed calculations of structural and dynamical properties for a 20-residue protein (304 atoms) Trp-cage[11] (PDB code: 1L2Y), depicted in Fig. 1. The DMD results are compared to those obtained from the standard NMA. All calculations presented in this work were carried out *in vacuo*. In both NMA and DMD simulations, the Trp-cage protein was described with the AMBER force field *ff*98 for nucleic acids.[14] The structure was first energy minimized by quasi-Newton nonlinear optimization[15] until the rms gradient was less than 10^{-6} kcal/(mole Å).

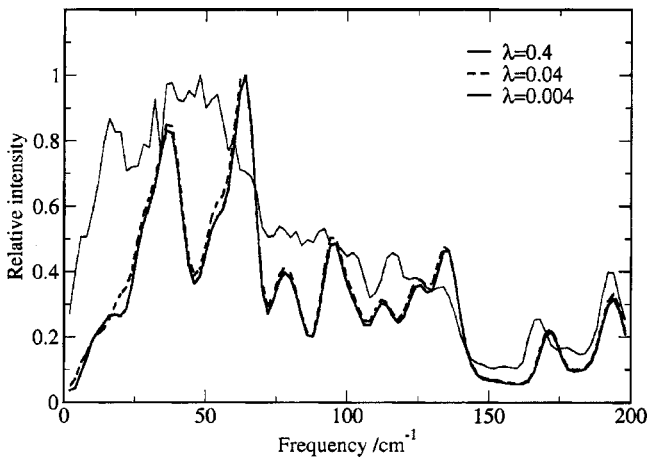

FIG. 2. Average absorption energy as a function of frequency calculated for various driving parameters λ in cm^{-1} $Å^{-1}$ units.

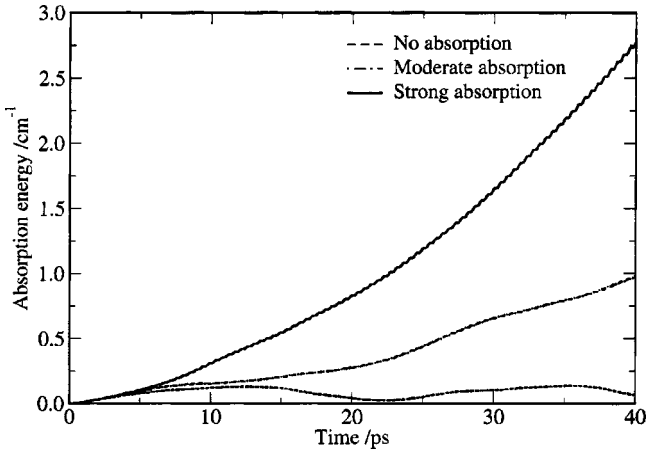

FIG. 3. Typical energy profiles for nonabsorbing low frequency mode ω=33.5 cm^{-1}, moderately absorbing frequency ω=35.0 cm^{-1}, and strongly absorbing frequency ω=36.5 cm^{-1}.

We examine properties calculated from frequencies and normal modes for the two methods, such as root-mean square atomic fluctuations, temperature-dependent *B*-factors, the vibrational entropy, anisotropies, and cross-correlation coefficients. In most of the DMD calculations presented here, the Trp-cage frequency spectrum is scanned with a frequency step size equal to 1 cm^{-1};[16] each trajectory is propagated up to 5 ps with 10 000 integration steps, unless stated otherwise. The initial velocities of the molecular system are set to zero.

A. The driving parameter λ

First we discuss the issue of the driving strength. For simplicity, the driving parameter [λ parameter in Eq. (3)] was chosen to be the same for all bonds. The driving applied to all bonds ensures that all normal modes will be excited. Note that the choice of equal driving parameters is not necessary and was made for convenience. For example, if a molecule has symmetry (which is not the case for a protein), normal-mode frequencies can be driven separately with driving parameter matrix that transforms according to the appropriate irreducible representation of a point group.[10] It should

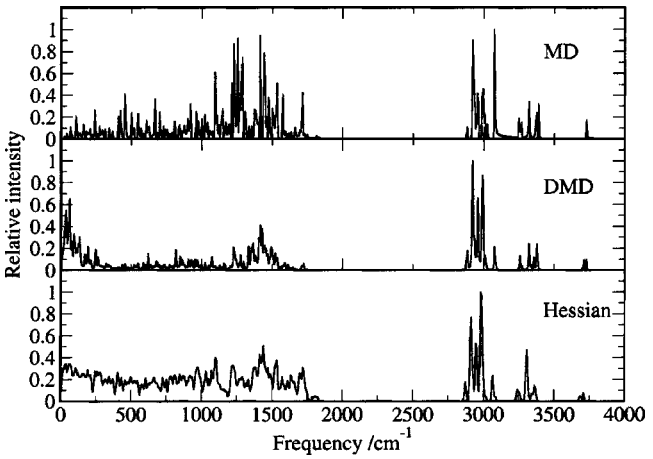

FIG. 4. Distributions of frequencies calculated by standard NMA (Hessian), MD simulations with and without the driving terms (DMD and MD, respectively).

TABLE I. Statistics on the threshold to select absorbing frequencies.

Threshold	N_m/N_f [a]	ΔRMS[b]	S(300 K)[b]
5.0×10^{-4}	0.64	+64%	+32%
2.5×10^{-4}	1.09	+26%	+11%
1.0×10^{-4}	1.93	−6%	−2%
7.5×10^{-5}	2.09	−10%	−5%
5.0×10^{-5}	2.23	−13%	−7%

[a]N_m is the number of absorbing frequencies in the molecule and N_f is the number of normal mode frequencies in molecule.
[b]The difference between the DMD and Hessian root-mean square fluctuations and entropies.

be noted, however, the modes are not driven ''equally'' for given driving time, i.e. some modes absorb more energy than others. The driving parameter has to be selected carefully, because it determines the amplitudes of the atomic motion. In the case of a resonant frequency, as the trajectory propagates, the molecule continues to absorb energy, and molecular motions can exceed the small amplitude, harmonic limit. Driving beyond this limit is another feature of the DMD method, since it offers the study of coupled anharmonic motion, with, however, the attendant complications of possibly ''chaotic motion.'' However, the focus of the present study is to compare the DMD results to those obtained from the NMA, so it is desirable to keep molecular vibration in the small amplitude region. This was done by stopping the driving when the absorbed energy is roughly 10% of the harmonic frequency, except for the lowest frequency modes, where the cutoff is 20 or 30%. This threshold is also a useful one for automated application of the method.

Certainly, the longer the propagation time is, the smaller the driving parameters should be to keep the atomic motions within the harmonic limit. We tested several driving parameters by analyzing the average absorption energy as a function of frequency as shown in Fig. 2. We gradually decreased the driving parameter until the shape of the function did not change. This convergence test is important to identify spurious absorption peaks at very low frequencies. Finally we chose the driving parameter $\lambda=0.04$ $\mathrm{cm}^{-1}\,\mathrm{Å}^{-1}$ for the present calculations of properties of proteins derived from the normal modes and frequencies.

Typical energy profiles for the nonabsorbing, moderately absorbing, and resonant frequencies are shown in Fig. 3. The average absorption energy [Eq. (5)] is small and oscillatory for nonabsorbing frequencies, while it rises rapidly at resonant frequencies.

B. Spectrum

The ''absorption'' spectrum, i.e., the energy absorbed from DMD calculations, along with the power spectrum obtained by the Fourier transform of the velocity autocorrelation function generated by a standard MD simulation, also for 5 ps, and the results from the standard NMA are shown in Fig. 4. Unlike the discrete frequencies in the normal-mode analysis, the frequency distributions in the MD and DMD simulations are quasicontinuous. In order to better compare the discrete NM spectrum with the MD and DMD distributions, the NM distribution has been represented as a sum of Gaussians with a resolution corresponding to the Fourier transform limit of a 5 ps trajectory. These three spectra are not strictly equivalent; however, they should agree in the positions of the peaks, and on the scale of resolution of this figure, there is good agreement. All spectra have been normalized so that the largest peak value is equal to 1.

C. Atomic fluctuations

The DMD normal modes for the resonant frequencies are calculated from the Cartesian coordinates obtained any time in the trajectory, after significant absorption of energy has occurred. Next we consider several properties computed from the normal modes or the frequencies. One physical quantity of particular interest is the thermal variance in the position of atoms at equilibrium, $\langle(\Delta x_j)^2\rangle$, $\langle(\Delta y_i)^2\rangle$ and $\langle(\Delta z_i)^2\rangle$. For each variance the widely used classical expression for the atomic fluctuation of atom i is, e.g., for $\langle(\Delta x_i)^2\rangle$,

$$\langle(\Delta x_i)^2\rangle=\frac{k_BT}{m_i}\sum_{j=1}^{3N-6}\frac{u_{ij}^2}{\omega_j^2}, \tag{6}$$

where ω_j is the frequency of mode j and u_{ij} is the corresponding projection of normal mode j on the Cartesian coordinates of atom i. Since the present implementation of DMD method does not resolve exact NM frequencies but resonant frequencies extracted from the uniform scan of the spectrum, the sum in the Eq. (6) can be formally replaced by an integral:

TABLE II. Root-mean square fluctuations $[\langle\Delta r_i^2\rangle]^{1/2}$ in Å.

	DMD			Hessian		
Method	Mean[a]	Minimum	Maximum	Mean[a]	Minimum	Maximum
All atoms	0.48 (0.19)	0.21	1.47	0.51 (0.21)	0.24	1.59
Backbone	0.36 (0.11)	0.21	0.61	0.36 (0.10)	0.24	0.58
Side chains	0.51 (0.21)	0.23	1.47	0.55 (0.23)	0.28	1.59
N	0.40 (0.14)	0.21	1.20	0.42 (0.15)	0.25	1.34
C	0.44 (0.10)	0.21	1.10	0.44 (0.10)	0.24	1.22
O	0.45 (0.09)	0.29	0.75	0.49 (0.12)	0.29	0.89
H	0.53 (0.15)	0.23	1.47	0.58 (0.16)	0.30	1.59
C_α	0.37 (0.11)	0.21	0.61	0.37 (0.11)	0.26	0.58

[a]The numbers are averages over all Trp-cage atoms for a particular class. Numbers in parentheses are standard deviations.

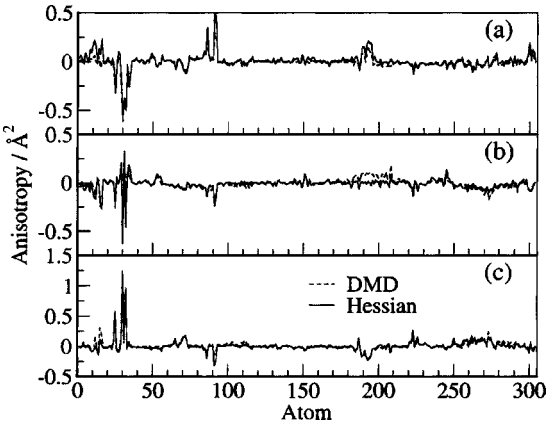

FIG. 5. Anisotropy of fluctuations in Å^2 of each atom measured by the deviation of (a) $\langle(\Delta x_i)^2\rangle$, (b) $\langle(\Delta y_i)^2\rangle$, and (c) $\langle(\Delta z_i)^2\rangle$ from the isotropic value $\langle(\Delta r_i)^2\rangle/3$ calculated by the DMD and standard NMA (Hessian).

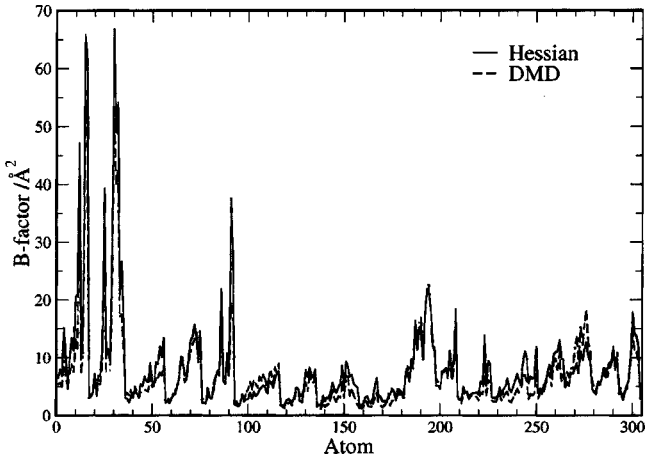

FIG. 6. A comparison of *B*-factors calculated by NMA (Hessian) and low resolution DMD simulation at $T=300$ K.

$$\langle(\Delta x_i)^2\rangle = \frac{k_B T}{m_i} \frac{\int_0^{\omega_{\max}} \frac{1}{\omega^2} W(\omega)[u_i(\omega)]^2 d\omega}{\int_0^{\omega_{\max}} \frac{W(\omega)}{N_f} d\omega}, \tag{7}$$

where $W(\omega)$ is a weighting function (see below), the integral in the denominator of Eq. (7) is a normalization factor, and N_f equals the number of normal modes ($3N-6$ for N atoms). For the purpose of evaluation of the integrals the following simple discretization procedure is used. The weight factors $W(\omega)$ are 0 for nonabsorbing frequencies and 1 for absorbing ones, and in the present case $\omega_{\max}=4000\,\mathrm{cm}^{-1}$.

At nonresonant frequencies the absorbed energy of the molecule is small and oscillatory with time. In our calculations, the oscillatory limit was about 10^{-4} (of a normalized integrated spectrum), and all driving frequencies exceeding this threshold value are considered as absorbing frequencies. With this condition Eq. (7) becomes

$$\langle(\Delta x_i)^2\rangle = \frac{N_f}{N_m} \frac{k_B T}{m_i} \sum_{j=1}^{N_m} \frac{u_{ij}^2}{\omega_j^2}, \tag{8}$$

where N_m is the number of absorbing frequencies. We tested the DMD results with respect to the threshold value and found stable and accurate results relative to the exact ones provided the threshold is low enough so that N_m is twice larger than N_f (see Table I). With this inequality very few true normal modes are missed in the above equation. There will be approximate copies of normal modes appearing in the sum; however, the normalization in this sum approximately (and evidently accurately) accounts for this.

Statistics on properties related to atomic fluctuations of Trp-cage protein are collected in Table II and compared to the exact Hessian-based results. For most properties the error is roughly 10% or less. (If exact normal-mode frequencies are used, the agreement of the DMD and Hessian results is nearly perfect; the error is 2%.) The root-mean square fluctuations averaged over all protein heavy atoms (C,N,O) are in the ranges an 0.36–0.58 Å; backbone atoms tend to have smaller fluctuations, and side chain atoms tend to have larger fluctuations. There is an increase in the magnitude of the fluctuations as one goes from the center of the protein out toward the terminal groups.

D. Anisotropy of motion

In the particular case of isotropic motion the fluctuations in all directions are equal and given by

$$\langle(\Delta x_i)^2\rangle = \langle(\Delta y_i)^2\rangle = \langle(\Delta z_i)^2\rangle = \frac{\langle(\Delta r_i)^2\rangle}{3} \tag{9}$$

for each residue (atom) i. The deviation of $\langle(\Delta x_i)^2\rangle$, $\langle(\Delta y_i)^2\rangle$, $\langle(\Delta z_i)^2\rangle$ from the isotropic value $\langle(\Delta r_i)^2\rangle/3$ provides a measure of the anisotropy of fluctuations. The quality of the DMD results is verified against the exact Hessian-based calculations shown in Fig. 5. The largest differences between the DMD and Hessian calculations are for the atoms executing the largest fluctuations, e.g., atoms in the terminal amino-acid groups and hydrogen atoms.

E. *B* factors

The atomic fluctuations are related to temperature-dependent crystallographic factors (*B*-factors) according to the well-known expression

$$B_i = 8\pi^2 \frac{\langle(\Delta r_i)^2\rangle}{3}. \tag{10}$$

Figure 6 shows the *B*-factors at $T=300$ K for all atoms of the Trp-cage protein calculated from the Hessian normal modes [Eq. (6)] and the DMD normal modes [Eq. (8)]. The agreement between the standard Hessian and DMD simulations is very good, lending support to the use of DMD method for further investigation of protein dynamics.

F. Entropy

Dynamical techniques are useful for understanding the internal motions of complex systems as well as for evaluat-

ing thermodynamic properties. As far as proteins are concerned, the computational estimate of entropy is of basic interest in protein folding and ligand binding. In this work, we test the calculation of the absolute entropy of Trp-cage using the driven MD approach. The resonant frequencies obtained from the same 1 cm^{-1} scan procedure are used in the standard expression for entropy,[9] where again the sum is replaced by an integral,

$$S_v = \frac{\int_0^{\omega_{\max}} \left[\frac{\hbar\omega}{T(e^{\hbar\omega/k_BT}-1)} - k_B \ln(1-e^{(-\hbar\omega/k_BT)}) \right] W(\omega) d\omega}{\int_0^{\omega_{\max}} \frac{W(\omega)}{N_f} d\omega}, \tag{11}$$

and then evaluated as described in Sec. III B. The absolute values of entropy as a function of temperature are given in Table III along with exact NMA results. As seen, the DMD entropies are in very good agreement with the exact NMA results.

G. Cross-correlation coefficients

Finally, we consider cross-correlation coefficients; these are a general indication of the degree of the collective motion in protein. The cross-correlation coefficient between atoms i and j is defined as[17]

$$C_{ij} = \frac{\langle \Delta r_i \Delta r_j \rangle}{(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle)^{1/2}}. \tag{12}$$

These coefficients range from a value of -1 (completely anticorrelated motions) to a value of $+1$ (completely correlated motions). They reflect correlation of displacements along a straight line. In other words, two atoms moving exactly in phase and with the same period, but along perpendicular lines, will have a cross correlation of zero. Initially the calculation was done using the 1 cm^{-1} low resolution DMD data set of frequencies and modes; however, the results were not in satisfactory agreement with an exact normal-mode calculation. The reason for this is that cross-correlation coefficients are sensitive to the orientation of modes, and thus accuracy of these modes is essential. For most proteins, accurate values of $\langle (\Delta x_i)^2 \rangle$ can be obtained with as few as 30 lowest-frequency normal modes. Indeed, it has been shown that low-frequency normal modes of proteins, with frequencies under 30 cm^{-1}, are responsible for most of their atomic displacements.[18] In new DMD calculations we carried out the driving at exact Hessian normal-mode frequencies up to 200 cm^{-1}. Using these, we calculated the cross correlations of the fluctuations of all backbone C_α's and did obtain accurate cross-correlation coefficients, plotted in Fig. 7 next to the exact normal-mode results. Figure 7(c) is a result of the high resolution (40 ps simulation) DMD scan performed up to 100 cm^{-1} with 0.25 cm^{-1} frequency step.

TABLE III. Trp-cage absolute entropies in kcal/mole K.

Temperature/K	Hessian	DMD
300	0.820	0.807
400	1.058	1.057
500	1.286	1.293
1000	2.159	2.154
1500	2.685	2.662
2000	3.030	2.997

H. Resolution

We expect that a large molecule, such as a protein, has many regions of closely spaced frequencies. This presents a challenge to the DMD approach to resolve closely spaced frequencies. However, it is also important to determine how accurately we need to resolve frequencies in order to determine properties of proteins accurately. The resolution power of the DMD method is demonstrated in Fig. 8. We scanned the spectrum in a small frequency range with a step of 0.25 cm^{-1}. Each trajectory was run for up to 40 ps with driving parameter $\lambda = 0.0004$ cm^{-1} $Å^{-1}$ and as the trajectory was propagated, we monitored the average absorption energy. The resolving power of the DMD approach increases significantly in going from 10 to 40 ps. After 30 ps the two peaks, separated by about 2 cm^{-1}, are clearly resolved. The overlaps of the corresponding DMD normal modes with the exact NMA are 0.90 and 0.99, respectively. For regions of higher density we need to propagate the trajectory longer and scan with a very small frequency step. Note that for longer propagation time smaller driving parameters should be used to preserve small amplitude behavior, as was mentioned before. Larger time propagation comes with increase in computational time, so it is important to determine whether typical properties from a NMA can be accurately obtained with a ''low resolution'' scan of the frequencies. In this section we have demonstrated that average molecular properties of proteins, such as root-mean square fluctuations and *B*-factors, can be obtained accurately even from a very low resolution DMD calculation.

IV. SUMMARY AND CONCLUSIONS

The DMD method, applied in the simplest possible fashion, i.e., by a uniform scan in frequencies, does provide an accurate description of averaged quantities such average atomic fluctuations and entropy for the Trp-cage protein, in comparison to Hessian normal-mode results. For more de-

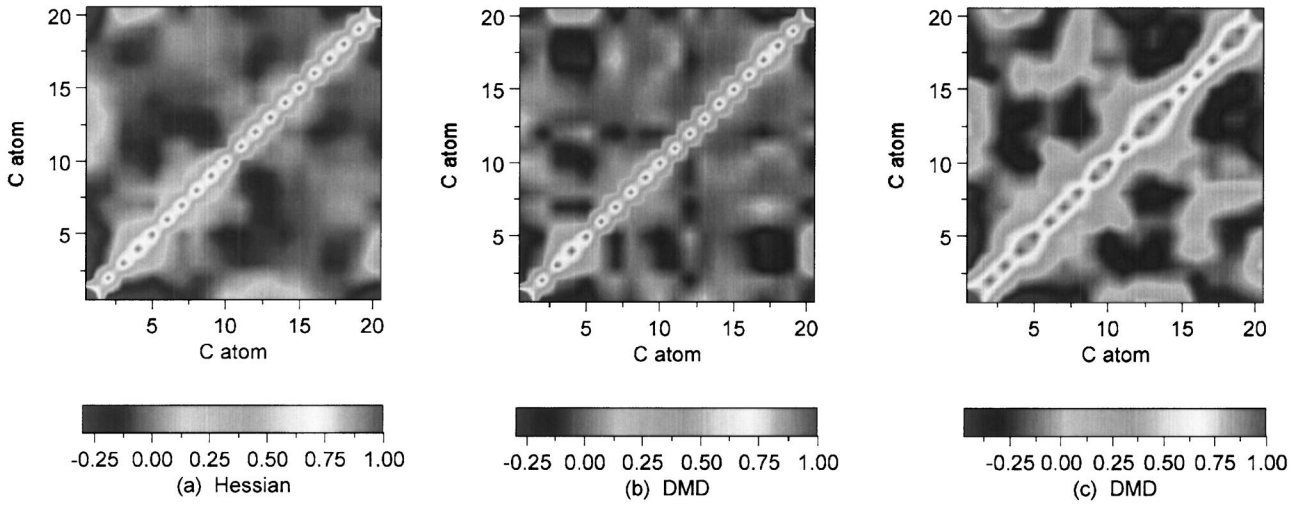

FIG. 7. Cross-correlation maps for C_α atoms calculated from (a) Hessian normal modes and frequencies, (b) DMD normal modes at exact Hessian frequencies, and (c) DMD normal modes at frequencies obtained from the high resolution DMD calculation.

tailed properties such as cross-correlation coefficients, accurate results require driving at exact normal-mode frequencies. (Of course all DMD calculations would benefit from knowledge of the exact normal-mode frequencies.) Obviously, to obtain these frequencies in a straightforward scan would be computationally very time consuming, but favorable for parallel computation. A variety of spectral deconvolution methods could also be applied to the spectrum. Another approach is to use the velocity autocorrelation function from a standard MD calculation but to apply new and promising signal processing methods to obtain accurate frequencies. For example, the filter diagonalization technique developed by Neuhauser and co-workers[19,20] and related but newer techniques by Taylor and co-workers[21,22] appear very promising for this purpose.

Finally, it is important to note that normal-mode analysis has well-known limitations, i.e., mode-coupling is ignored. Thus, the approach cannot be applied to the exciting, new field of two-dimensional infrared spectroscopy, which can probe correlated, time-dependent motion of proteins and peptides[23–28] and which also probes modes throughout the spectrum. The DMD method can be extended to this region; however, since it is based on (classical) molecular dynamics, it will have to be tested first against quantum calculations on small benchmark systems.

In summary, the agreement of the properties calculated using the DMD approach and the standard NMA is very good. Thus, we conclude that the DMD method is a viable alternative to the standard Hessian-based normal mode analysis. Since the memory requirements of DMD are very modest compared to the Hessian-based approach, we see no obstacle to implementing it to a large bimolecule, where the Hessian-based approach could not be applied (on a common cluster of workstations). We are planning such an implementation in the near future.

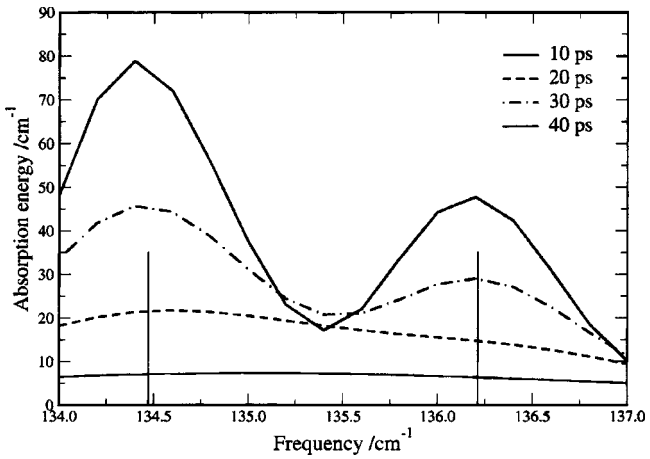

FIG. 8. Resolution power of the DMD method. (a) DMD calculations are carried out for 40 ps with 80 000 integration steps and driving parameter $\lambda=0.0004$ cm^{-1} Å^{-1}. The Hessian normal modes are shown as impulses at 134.5 and 136.2 cm^{-1}.

ACKNOWLEDGMENTS

We thank Professor A. Roitberg (University of Florida) for discussions and for suggesting the test with Trp-cage protein. J.M.B. acknowledges the National Science Foundation (Grant No. NSF CHE-0131482) for financial support.

[1] C. L. Brooks III, M. Karplus, and B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, Advances in Chemical Physics, Vol. 71 (Wiley and Sons, New York, 1988).
[2] G. Herzberg, *Molecular Spectra and Molecular Structure. II. Infrared and Raman Spectra of Polyatomic Molecules* (Van Nostrand, New York, 1945).
[3] E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955).
[4] M. Levitt, C. Sander, and P. S. Stern, J. Mol. Biol. **181**, 423 (1985).
[5] T. Nishikawa and N. Go, Proteins **2**, 308 (1987).
[6] B. R. Brooks, D. Janezic, and M. Karplus, J. Comput. Chem. **16**, 1522 (1995).
[7] J. Ma and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **95**, 8502 (1998).
[8] F. Tama, F. X. Gadea, O. Marques, and Y. Sanejouand, Proteins: Struct., Funct., Genet. **41**, 1 (2000).
[9] G. Li and Q. Cui, Biophys. J. **83**, 2457 (2002).
[10] J. M. Bowman, X. Zhang, and A. Brown, J. Chem. Phys. **119**, 646 (2003).
[11] J. W. Neidigh, R. M. Fesinmeyer, and N. H. Andersen, Nat. Struct. Biol. **9**, 425 (2002).

[12] TINKER Software Tools for Molecular Design, Version 4.1, Washington University School of Medicine, June 2003, available from http://dasher.wustl.edu/tinker
[13] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000).
[14] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[15] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, New York, 1999), Sec. 9.1.
[16] We chose the frequency step equal to 1 cm^{-1} for a 5 ps simulation. The choice of a frequency step size is essentially dictated by the frequency resolution.
[17] T. Ichiye and M. Karplus, Proteins **11**, 205 (1991).
[18] B. R. Brooks and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **82**, 4995 (1985).
[19] M. R. Wall and D. Neuhauser, J. Chem. Phys. **102**, 8011 (1995).
[20] A. J. R. da Silva, J. W. Pang, E. A. Carter, and D. Neuhauser, J. Phys. Chem. A **102**, 881 (1998).
[21] J. Main, P. A. Dando, D. Belkic, and H. S. Taylor, J. Phys. A **33**, 1247 (2000).
[22] S. D. Kunikeev and H. S. Taylor, J. Phys. Chem. A **108**, 743 (2004).
[23] P. Hamm, M. Lim, and R. M. Hochstrasser, J. Phys. Chem. B **102**, 6123 (1998).
[24] P. Hamm, M. Lim, W. F. DeGrado, and R. M. Hochstrasser, Proc. Natl. Acad. Sci. U.S.A. **96**, 2036 (1999).
[25] P. Hamm, M. Lim, W. F. DeGrado, and R. M. Hochstrasser, J. Chem. Phys. **112**, 1907 (2000).
[26] S. Woutersen and P. Hamm, J. Phys. Chem. B **104**, 11316 (2000).
[27] S. Woutersen and P. Hamm, J. Chem. Phys. **114**, 2727 (2001).
[28] C. Sheurer, A. Piryatinski, and S. Mukamel, J. Am. Chem. Soc. **123**, 3114 (2001).