

Atomistic model of DNA: Phonons and base-pair openingF. Merzel,¹ F. Fontaine-Vive,^{2,3} M. R. Johnson,² and G. J. Kearley⁴¹*National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia*²*Institut Laue Langevin, Boîte Postale 156, 38042 Grenoble Cedex 9, France*³*Radiation, Reactors and Radionuclides Department, Faculty of Applied Sciences, Delft University of Technology, Mekelweg 15, 2629 JB Delft, The Netherlands*⁴*Bragg Institute, Building 87, Australian Nuclear Science and Technology Organisation, PMB 1 Menai, NSW 2234, Australia*

(Received 14 February 2007; revised manuscript received 28 May 2007; published 14 September 2007)

A fully atomistic model of B-DNA using the CHARMM (chemistry at Harvard molecular mechanics) force field is presented. Molecular dynamics simulations were used to prepare an equilibrium structure. The Hessian of interatomic forces obtained from CHARMM for the equilibrium structure was used as input to a large scale phonon calculation. The calculated dispersion relations at low frequency are compared with recently published experimental data, which shows the model to have good accuracy for the low frequency, vibrational modes of DNA. These are discussed in the context of base-pair opening. In addition to the widely reported modes at, or below, ~ 12.5 meV, a continuous band of modes with strong base-pair opening character is found up to 40 meV, which coincides with the typical denaturation temperature of DNA.

DOI: [10.1103/PhysRevE.76.031917](https://doi.org/10.1103/PhysRevE.76.031917)

PACS number(s): 87.14.Gg, 31.70.Ks, 33.20.Tp

I. INTRODUCTION

Understanding the function of biological molecules has evolved from being structure-based to including a knowledge of molecular dynamics [1]. In the case of deoxyribonucleic acid (DNA), a key point of interest is base-pair opening as such dynamics play a key role in replication, transcription, and denaturation [2]. These processes all involve the splitting of the double helix into single strands, which is initiated locally by the breaking of interbase hydrogen bonds and the formation of bubbles, which span several base pairs. Although these processes involve proteins interacting with DNA, they are thought to be driven by the dynamics of DNA itself.

Experimentally, the dynamics of DNA has attracted considerable attention. Raman spectroscopy has been used to study low frequency dynamics of A- and B-DNA [3,4] in solution and in liquid crystalline forms as a function of temperature. Peaks and broad features can only be described as corresponding to localized or delocalized modes, the latter being intra- or inter-helical. Optical spectroscopy is limited to $k \sim 0$, but dispersion associated with delocalized modes is typically measured by coherent, inelastic neutron scattering (INS). For large systems, diffusion at small wave vectors should be measured but kinematic restrictions make this impossible for INS and a strongly dispersive mode was measured in the vicinity (~ 20 nm⁻¹) of the base pair Bragg peak [5]. The kinematic problem does not apply to inelastic x-ray scattering (IXS) and the same dispersive mode has recently been measured close to the gamma point in a range of DNA samples [6–8]. The combined result of these INS and IXS measurements is of an acoustic-type phonon that has a maximum frequency of ~ 12.5 meV at 10 nm⁻¹, a minimum of 2–4 meV at ~ 20 nm⁻¹, and at higher Q the phonon is strongly damped and can no longer be measured.

While many experiments have been designed and performed to study DNA and its key processes, models are essential for a detailed understanding. Biological activity oc-

curs close to room temperature and it therefore tends to involve low frequency, collective modes. Given past limitations in computing power and, otherwise, the need to access long times in dynamics simulations in order to obtain statistically meaningful results, a number of mesoscale models have been developed. Such models reduce the large number of atomistic degrees of freedom to a limited set so that DNA is typically represented by a one-dimensional chain of effective atoms with a Morse potential describing the hydrogen-bond potential between the base-pair molecules [9,10]. The Peyrard-Bishop-Dauxois model [11] has been parametrized to reproduce denaturation curves and describes base-pair opening frequencies at 5 and 10 meV for adenine-thymine (A-T) and guanine-cytosine (G-C) pairs, respectively.

At the other end of the scale of precision of total energy calculations, density functional theory based methods are commonly used to investigate atomistic models of a few base-pairs and, exceptionally, to calculate the electronic structure of a full periodic model [12,13]. However, these methods are still too time-consuming to be used to investigate the dynamics of a periodic model.

While atomistic models have, in principle, too many degrees of freedom, the parametrization of interactions is simpler and including all degrees of freedom avoids any artifacts associated with the definition of bead particles. This type of approach was originally developed by Prohofsky, based on a Urey-Bradley force field [14], although approximations were introduced, for example, treating the presence of the second strand in the double helix as a perturbation of the first [15] and modeling changes in hydration by an electrostatic perturbation [16]. Models of this kind predict base-pair opening frequencies between 7.3 and 10.4 meV for A-T and between 11.8 and 15.4 meV for G-C [17].

Nowadays, force fields like CHARMM (chemistry at Harvard molecular mechanics), which have evolved over several decades, enable a wide range of biomolecular structure and dynamics to be simulated with good accuracy, including DNA [18]. Thus it is possible to perform simulations without refining the parameters in the force field to reproduce experi-

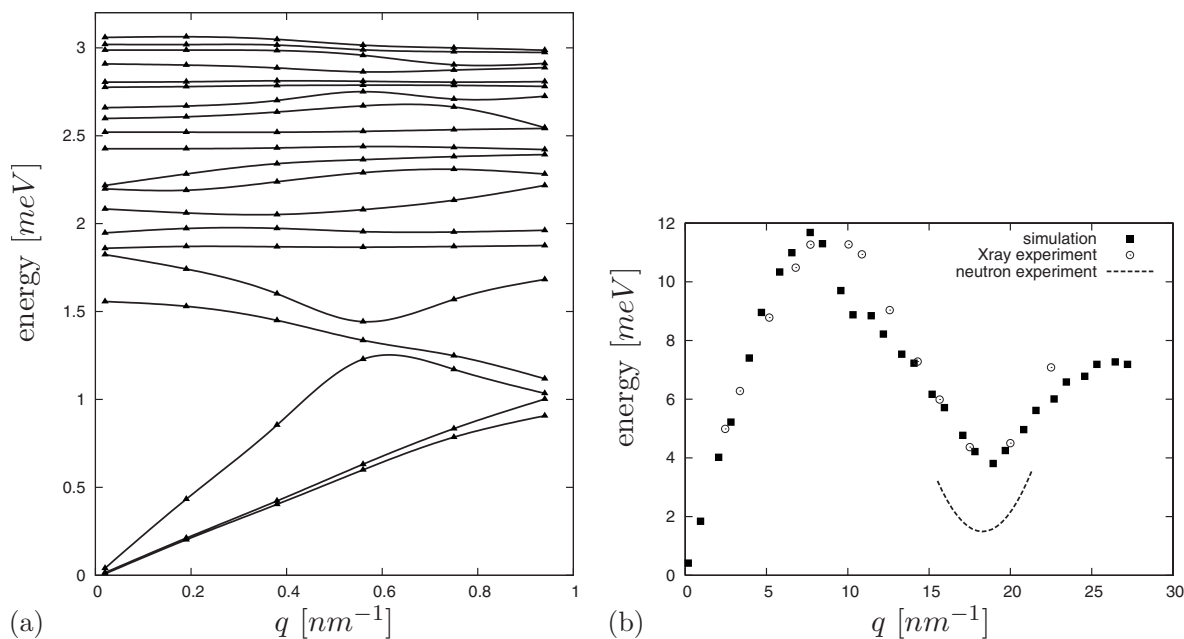


FIG. 1. Calculated dispersion curves of the 20 lowest modes (a) and the apparent dispersion (b) obtained by fitting the spectral intensity. X-ray data are taken from [8].

mental observables and the numerical model can be used to explore in detail the behavior of a complex system. In this paper, we therefore use the CHARMM force field to investigate the normal modes of a fully atomistic, hydrated model of DNA, comparing simulation results with experimental data wherever possible, in particular concerning low frequency phonons. Normal modes are investigated in terms of bead particles. In addition to the simplified description and parametrization of interactions, another advantage of the fully atomistic simulation is that the normal modes can be analyzed for different bead descriptions, ranging from atoms, through base, sugar, and phosphate molecules to nucleotides.

II. COMPUTATIONAL MODEL

The CHARMM force field and molecular dynamics (MD) code has been shown to reproduce average, structural properties of different forms of DNA [18], but it has not been applied to study phonons and molecular vibrations. We have done precisely this to obtain an atomistic picture of the vibrational modes of DNA, with particular interest in the low frequency modes. In order to compare our B-DNA model with the INS and IXS phonon data, we have set up a phonon calculation for arbitrarily large systems, using the direct method as embodied in the PHONON code [19].

An initial B-DNA structure was obtained from the Protein Data Bank. This structure was adapted slightly to give a rectangular, periodic model with dimensions $a=b=3.23$ nm and $c=3.35$ nm where c is parallel with the helix axis. The ten base-pair sequence is CTCTGCTACT (for one strand) and the simulation box was filled with 20 lithium counterions and almost 1000 water molecules giving a density of 1.14 g/cm^3 . MD (NVE) simulations at 300 K, followed by a quench were used to generate the equilibrium structure.

While the water molecules and counterions are not of direct or immediate interest in the low frequency dynamics of DNA, their presence in the MD simulation is essential for the stability and good structural properties of the double helix.

For such large systems, the Hessian has to be determined by finite displacements and this matrix was given as input to our phonon code. To diagonalize the complex, dynamical matrix, which contains $\sim 5 \times 10^8$ matrix elements, we used the LAPACK routine “zcheev.” Wave vectors along the z (helical) axis were chosen to investigate the measured dispersion curves (see Fig. 1). Spectral intensity at a given frequency, for comparison with INS and IXS experimental data, is determined from the eigenvectors according to the equation [20]

$$S(\vec{q}, \omega)_{coh} = \sum_{\vec{G}} \sum_{\vec{k}, j} \frac{\hbar}{2\omega_{\vec{k}j}} \times \left| \sum_{\mu} \sigma_{\mu}^{coh} \frac{\vec{q} \cdot \vec{e}_{\mu\vec{k}j}}{\sqrt{M_{\mu}}} \exp[-W_{\mu}(\vec{q}) + i\vec{q}\vec{r}_{\mu}] \right|^2 \times [n(\omega_{\vec{k}j}) + 1] \delta(\omega - \omega_{\vec{k}j}) \delta(\vec{q} + \vec{k} - \vec{G}), \quad (1)$$

where \vec{q} is the scattering vector, σ_{μ} is the corresponding atomic scattering length, $[n(\omega) + 1]$ refers to the phonon creation process (absorption spectrum), and $n(\omega)$ is the mean number of phonons of frequency ω at temperature T according to the Bose-Einstein statistics

$$[n(\omega) + 1] = \frac{\exp(\hbar\omega/k_B T)}{\exp(\hbar\omega/k_B T) - 1}. \quad (2)$$

The factor $\exp[-2W_{\mu}(\vec{q})]$ is called the Debye-Waller factor:

$$W_{\mu}(\vec{q}) = \frac{\vec{q} \cdot \mathbf{B}(\mu) \cdot \vec{q}}{2}, \quad B_{\alpha\beta}(\mu) = \langle u_{\mu\alpha} u_{\mu\beta} \rangle_T, \quad (3)$$

where $\mathbf{B}(\mu)$ is a 3×3 symmetric tensor representing the thermodynamic mean square displacement of an atom μ , which can be expressed by the partial atomic phonon density of states $g_{\alpha\beta,\mu}(\omega)$:

$$B_{\alpha\beta}(\mu) = \frac{3N\hbar}{2M_{\mu}} \int_0^{\infty} \frac{d\omega}{\omega} g_{\alpha\beta,\mu}(\omega) \coth\left(\frac{\hbar\omega}{2k_B T}\right). \quad (4)$$

With 3551 atoms in the model, a complete mode assignment is not possible or meaningful and a number of projections operators, in addition to simple projections on to axial and radial coordinates with respect to the helix, are defined to extract the characteristic dynamics of the double helix.

Breathing:

$$B = \frac{1}{N_p} \sum_p \frac{\left| \frac{\vec{r}}{|\vec{r}|} (\vec{d}_{p1} - \vec{d}_{p2}) \right|}{2 \max(|\vec{d}_{p1}|, |\vec{d}_{p2}|)}. \quad (5)$$

Swing:

$$S = \frac{1}{N_p} \sum_p \frac{\left| \frac{\vec{z}}{|\vec{z}|} (\vec{d}_{p1} - \vec{d}_{p2}) \right|}{2 \max(|\vec{d}_{p1}|, |\vec{d}_{p2}|)}. \quad (6)$$

Rotation:

$$R = \frac{1}{N_p} \sum_p \frac{\left| \left(\frac{\vec{r}}{|\vec{r}|} \times \frac{\vec{z}}{|\vec{z}|} \right) (\vec{d}_{p1} - \vec{d}_{p2}) \right|}{2 \max(|\vec{d}_{p1}|, |\vec{d}_{p2}|)}. \quad (7)$$

d_{p1} and d_{p2} are the net displacement vectors of beads, defined, for example, as nucleotides, and r is the interbead vector. Replacing $-$ in Eq. (5) by $+$ gives a radial translation projector (RT) and a similar transformation in Eq. (6) gives an axial translation operator (AT). Note that these definitions do not allow for correlated movement between base pairs to be detected. The extent of delocalization of a vibrational mode is given by the participation ratio (PR),

$$\mathcal{R}_P = \frac{1}{N} \frac{\left[\sum_{i=1}^N |\vec{e}_i(\omega)|^2 \right]^2}{\sum_{i=1}^N |\vec{e}_i(\omega)|^4}, \quad (8)$$

where e_i and ω are the eigenvectors and corresponding eigenfrequencies. The PR is unity for acoustic phonons and is close to zero for a strongly localized mode.

III. RESULTS

Figure 1 shows the dispersion relations for the first 20 modes in the direction of the helix axis. We note that the acoustic phonons have a maximum frequency of 1 meV at the Brillouin zone boundary. At higher frequencies, we find

optic modes with limited dispersion. At the gamma point the optic modes start at 1.7 meV and there are ~ 1000 modes in the first 12.5 meV. This result is consistent with the observation by Raman spectroscopy of an optic mode at 2 meV for B-DNA. Since we do not calculate Raman intensity, we cannot exclude the possibility of the spectral profile having a maximum at higher frequency, but the dispersion curves show that the optic mode cannot occur at lower frequency.

Calculating the spectral intensities for all modes for coherent IXS and INS for wave vectors up to 30 nm^{-1} gives total spectral profiles. These are well-fitted with a simple Gaussian, which enables the characteristic frequency and width of the spectra to be determined for each wave vector. The dispersion curve obtained in this way [Fig. 1(b)] agrees well with the experimental data in terms of frequency and width, the width being comparable to the frequency. The experimental signal is not a direct measurement of the acoustic phonon but it is the projection of spectral intensity over a large number of mainly optic modes. The sound velocity derived from Fig. 1 at low q and over an extended q range ($< 5 \text{ nm}^{-1}$) is 2900 m/s.

The comparison of experimental and numerical results indicates that the interactions in the CHARMM force field are reasonably well-described. With 3551 atoms in our model, a complete mode assignment is not possible or meaningful. Furthermore, we have considered a realistic, random sequence of base pairs which, with the water, tends to smear out any particular vibrational character. Figure 2 shows the frequency dependence of the various projections, for beads defined as nucleotides, and the participation ratio for the gamma point vibrations. Modes tend to be delocalized over the first 5 meV and at higher frequency (up to 35 meV) there is a significant decay in the PR. Modes with breathing character (out-of-phase radial motion of the nucleotides) peak at 14.2 meV. We have checked the contributions of A-T and G-C pairs in the breathing projection and find only a very weak tendency for the A-T contribution to occur at a lower frequency than that of G-C. At ~ 12.5 meV vibrational modes are already significantly localized, which means there is no coherent base pair opening over a significant length of the helix. The in-phase, radial motion of the nucleotides is strongest at low frequency (< 12.5 meV) since these modes do not involve stretching of the inter-base hydrogen bonds. In accordance with these ‘‘breathing’’ projections, the simple radial projection shows maxima at ($>$) 15 and 6 meV. The axial projection is almost a direct inversion of the radial projection, although the modulation of these projections as a function of frequency is weak. The out-of-plane rocking of the nucleotides is pronounced at ~ 3 meV. The torsional character of modes is only marginally pronounced at 6 meV.

The extent of delocalization and polarization is illustrated by displacement vector plots for particular modes. Figure 3 shows three modes with pronounced radial character (top left panel), with axial character due to in-phase motion of nucleotide pairs (top right panel) and with radial breathing character (bottom panel). The third mode has the strongest ‘‘breathing’’ polarization of all modes as defined by Eq. (1) and indicated by the ring in Fig. 2. The PR for this mode is about half of the PR for the lower frequency modes shown in Fig. 3 and, accordingly, only three or four nucleotide pairs

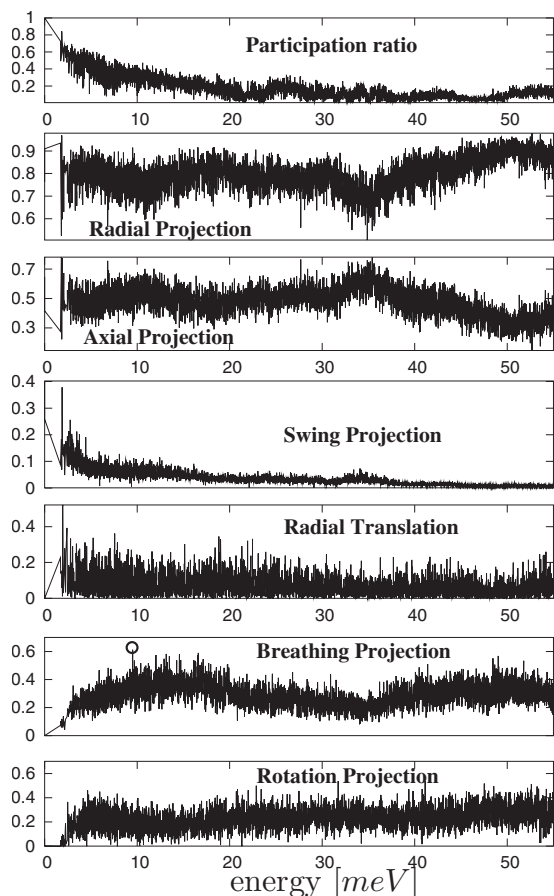


FIG. 2. Projection analysis of eigenvectors as a function of frequency.

have significant displacements. This degree of localization is consistent with the fact that we have not found a significant separation in frequency of the A-T and G-C contributions in the base-pair opening projection. In our DNA model, there are no extended sequences of a particular kind of base pair, compared to the extent of localization, so all normal modes with base-pair opening character involve a mixture of base pairs.

While the foregoing assignments are broadly consistent with previous work, we draw attention to the marked rise in the breathing projection just above 35 meV in Fig. 2. This frequency range coincides with the temperature of DNA denaturation ($\sim 100^\circ\text{C}$) and these modes could therefore contribute to the related, biologically important processes, augmenting the vibrational amplitude of the low frequency modes for base-pair opening. The PR at >35 meV is about half the value at 12.5 meV indicating modes localized over a couple of base pairs.

By redefining the beads, we can check that the average motion of nucleotide beads does actually correspond to a stretching of the interbase hydrogen bonds. Using base molecules as beads (see Fig. 4) shows that there is an almost continuous range of modes with base-pair opening character up to 40 meV. The difference between the two types of projections indicate that intranucleotide distortions, bending degrees of freedom between the component molecules, reduce

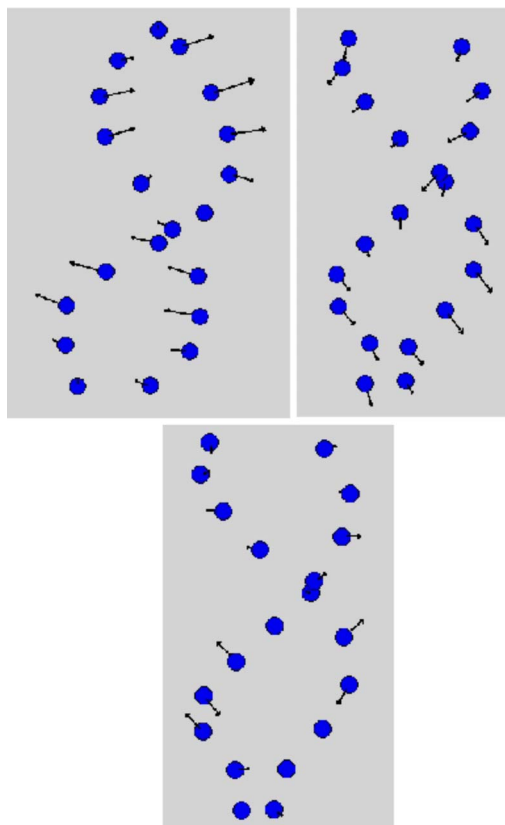


FIG. 3. (Color online) Selected eigenvectors, in a bead representation, showing strong radial (top left), axial (top right), and breathing (bottom) character. The mode frequencies are, respectively, 1.92, 1.71, and 9.51 meV.

the base-pair opening characteristics in the frequency range around 20 meV when nucleotides are used as beads.

IV. CONCLUSION

The fully atomistic model presented here complements the bead models that have been used to date to investigate

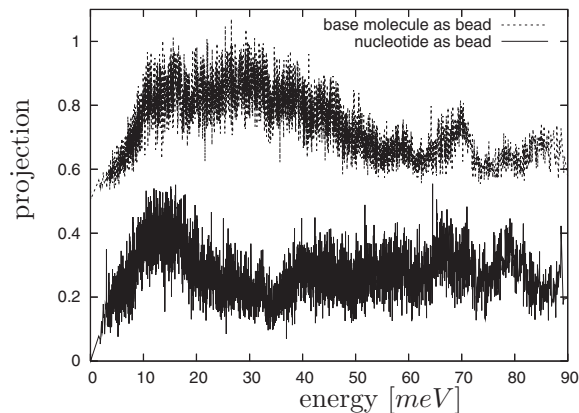


FIG. 4. Breathinglike character of modes as a function of energy using a base molecule as a bead and a whole nucleotide as a bead, respectively.

the dynamics of DNA. In particular the problem of parametrizing the interactions in a bead model is avoided by using the CHARMM force field in which force constants have been developed based on a wealth of experimental data and quantum chemistry calculations. The inconvenience of large systems is the scale of the phonon calculations and the number and complexity of the eigenvectors. These problems have been handled using large scale matrix diagonalization, projection operators, and different bead representations for the displacement vectors. Our calculations reproduce well the IXS measurements of the apparent dispersion of a low frequency excitation. The calculations show clearly that acoustic phonons have a maximum frequency of ~ 1 meV and the majority of the observed dispersion is due to the wave-vector dependence of the intensity of optic modes. The lowest lying optic modes are also consistent with Raman measurements. In terms of specific modes, we have focused on the low frequency breathing modes of nucleotide pairs. The corresponding projection operator reveals a broad band of modes with a maximum at 14 meV, which is consistent with previous estimations for the breathing mode. The PR shows that these modes are typically localized on ~ 4 nucleotide pairs. By splitting the nucleotide beads into smaller base, sugar, and phosphate molecule beads, a continuum of progressively more localized, higher frequency modes up to 40 meV is revealed with pronounced breathing character and these modes enhance the relevant vibrational amplitudes at the temperature of DNA denaturation.

The fully atomistic model offers two important advantages in establishing these findings. First, all degrees of free-

dom are retained and the vibrational character of all possible modes is investigated. A model based on nucleotide beads would not show the higher frequency base-pair opening modes. Second, it is essential to be able to reduce the number of degrees of freedom in order to obtain a clear understanding of the vibrations and this can be done continuously from beads as atoms to beads as nucleotides in the analysis of the displacement vectors.

As a result of our findings, a number of extensions to this work are envisaged. In the near future it will be possible to perform electronic structure calculations on DNA sufficiently quick to allow the dynamics of DNA to be investigated with an unprecedented level of precision. The work presented here shows that the essential degrees of freedom to be retained are those of the bases, sugars, and phosphates (six each) so that a dynamical matrix could be constructed for 360 translational and rotational degrees of freedom for B-DNA. This type of model would still be based on the zero Kelvin equilibrium structure of DNA, which is relevant to the extent that it has allowed us to reproduce well the apparent dispersion of low frequency modes, which were measured at room temperature [8]. Between room temperature and the melting temperature, it will clearly be of interest to investigate the molecular dynamics simulations that we have actually used to generate the initial structure of DNA. While such models can be used to explore bubble formation directly, the way in which vibrational modes change with temperature could be investigated with principal component analysis [21].

-
- [1] M. Peyrard, *Nat. Phys.* **2**, 13 (2006).
 [2] T. S. van Erp, S. Cuesta-Lopez, J. G. Hagmann, and M. Peyrard, *Phys. Rev. Lett.* **95**, 218104 (2005).
 [3] H. Urabe *et al.*, *J. Chem. Phys.* **82**, 531 (1985).
 [4] Y. Tominaga *et al.*, *J. Chem. Phys.* **83**, 5972 (1985).
 [5] H. Grimm *et al.*, *Phys. Rev. Lett.* **59**, 1780 (1987).
 [6] Y. Liu *et al.*, *Phys. Chem. Chem. Phys.* **6**, 1499 (2004).
 [7] Y. Liu *et al.*, *J. Chem. Phys.* **123**, 214909 (2005).
 [8] M. Krisch, A. Mermet, H. Grimm, V. T. Forsyth, and A. Rupprecht, *Phys. Rev. E* **73**, 061909 (2006).
 [9] M. Peyrard and A. R. Bishop, *Phys. Rev. Lett.* **62**, 2755 (1989).
 [10] S. Cocco and R. Monasson, *J. Chem. Phys.* **112**, 10017 (2000).
 [11] N. K. Voulgarakis *et al.*, *Nano Lett.* **4**, 629 (2004).
 [12] P. J. de Pablo, F. Moreno-Herrero, J. Colchero, J. Gomez Herrero, P. Herrero, A. M. Baro, P. Ordejón, J. M. Soler, and E. Artacho, *Phys. Rev. Lett.* **85**, 4992 (2000).
 [13] F. L. Gervasio, P. Carloni, and M. Parrinello, *Phys. Rev. Lett.* **89**, 108102 (2002).
 [14] J. M. Eyster *et al.*, *Biopolymers* **13**, 2505 (1974).
 [15] J. M. Eyster *et al.*, *Biopolymers* **13**, 2527 (1974).
 [16] J. M. Eyster *et al.*, *Phys. Rev. Lett.* **38**, 371 (1977).
 [17] R. Beger *et al.*, *Biophys. J.* **58**, 437 (1990).
 [18] M. Feig and B. M. Pettitt, *Biophys. J.* **75**, 134 (1998).
 [19] K. Parlinski, in *Neutrons and Numerical Methods N₂M*, edited by M. R. Johnson, G. G. Kearley, and H. G. Buttner, AIP Conf. Proc. No. 476 (AIP, New York, 1999), p. 121.
 [20] S. Lovesey, *Theory of Neutron Scattering from Condensed Matter*, International Series of Monographs on Physics 72 (Oxford Science, Oxford, 1984).
 [21] B. R. Brooks *et al.*, *J. Comput. Chem.* **16**, 1522 (1995).