

Atomistic modeling of the low-frequency mechanical modes and Raman spectra of icosahedral virus capsids

Eric C. Dykeman and Otto F. Sankey

Department of Physics, Center for Biological Physics, Arizona State University, Tempe, Arizona 85287-1504, USA

(Received 30 September 2009; revised manuscript received 7 January 2010; published 19 February 2010)

We describe a technique for calculating the low-frequency mechanical modes and frequencies of a large symmetric biological molecule where the eigenvectors of the Hessian matrix are determined with full atomic detail. The method, which follows order N methods used in electronic structure theory, determines the subset of lowest-frequency modes while using group theory to reduce the complexity of the problem. We apply the method to three icosahedral viruses of various T numbers and sizes; the human viruses polio and hepatitis B, and the cowpea chlorotic mottle virus, a plant virus. From the normal-mode eigenvectors, we use a bond polarizability model to predict a low-frequency Raman scattering profile for the viruses. The full atomic detail in the displacement patterns combined with an empirical potential-energy model allows a comparison of the fully atomic normal modes with elastic network models and normal-mode analysis with only dihedral degrees of freedom. We find that coarse-graining normal-mode analysis (particularly the elastic network model) can predict the displacement patterns for the first few (~ 10) low-frequency modes that are global and cooperative.

DOI: [10.1103/PhysRevE.81.021918](https://doi.org/10.1103/PhysRevE.81.021918)

PACS number(s): 87.64.kp, 62.25.De, 87.15.A-

I. INTRODUCTION

Nearly every living creature on earth is affected by viruses. Viruses often are pathogenic and infect plant, animal, and bacteria cells. Viruses cause damage to cultivated plants and livestock and produce a variety of diseases in humans, which can lead to death. Viruses are parasites and once the host cell is infected, its metabolism is commandeered by the virus and used to replicate more viral particles.

The assembly, thermodynamics, and dynamics of virus capsids are an important research area as these pathogens strongly affect human health. Modeling using atomistic approaches is difficult due to the large size of these gene-carrying creatures. The most common assembly allotrope is the icosahedral form which contains a minimum of 60 interacting proteins. A full description of the cooperative dynamics in terms of a complete set of normal modes is still not possible. Yet for many applications, one is mainly interested in the global motion distortions of the capsid which generally are at low frequency (~ 30 GHz). These global motions are relevant for the maturation of the virion, attachment to cells or other surfaces, or the response to external mechanical probes such as atomic force microscope tips. In these applications, where the low-frequency modes are desired, an exact solution is difficult. It is the intention of this work to produce an exact solution of the normal-mode problem for large icosahedral viruses and to compare with approximate methods. Additionally we determine the Raman scattering intensity profile at low frequency.

Another motivation for this work are recent experiments with short pulses of light that inactivate viruses. One possible alternative to drug therapies is to use a physical method to inactivate the virus based on its mechanical properties. A particularly promising approach relies on stimulating (through some means) resonant excitations of a virus capsid's vibrational modes; large amplitude excitations produce instabilities in the capsid causing it to break apart. Such a hypothesis has been discussed in the past [1] and very early

work provides evidence of ultrasonic energy absorption in viruses [2]. Recent experiments have examined the feasibility of such treatments using ultrashort laser pulses. Tsen and colleagues [3–7] performed several such experiments with M13 bacteriophages and other viral particles in solution. Using conditions suggesting impulsive stimulated Raman scattering (ISRS) with near infrared and visible light, they have been able to inactivate M13 phages. However, it is still unclear if ISRS is the (sole) mechanism involved in the inactivation of the M13 phages or if other process help (or dominate) to enable the inactivation.

Normal modes are also useful for understanding the cooperative motion of flexible molecules and their change in conformation. For example enzymes undergo allosteric transitions upon binding with other molecules or substrates. The displacement patterns in a normal-mode analysis (NMA) give guidelines to the mechanical operation of these nanomachines [8]. Often simplified mode patterns are used [9], and a few normal modes are not always enough [10,11].

The “general” method for determining the normal modes of vibration of a molecule requires one to solve the $N_f \times N_f$ matrix equation,

$$\vec{D}\vec{e} = \lambda\vec{e}, \quad (1)$$

where \vec{D} is the dynamical matrix with mass-weighted elements $D_{ij} = \Phi_{ij} / \sqrt{(m_i m_j)}$, where $\vec{\Phi}$ is the force constant matrix (Hessian) for the molecule given by

$$\Phi_{ij} = \frac{\partial^2 V}{\partial x_i \partial x_j}. \quad (2)$$

The eigenvalues of Eq. (1) λ , are the squared frequencies ω^2 while the eigenvectors \vec{e} are the mass-weighted displacement patterns of the molecular motions; $\vec{e} = \vec{M}^{1/2} \vec{\eta}$ with $M_{ij} = m_i \delta_{ij}$. In a classical atomistic model, there are typically five potential-energy terms which contribute to the total potential energy (V) of a molecule corresponding to bond, angle, di-

hedral, Coulomb, and van der Waals interactions.

Diagonalizing the matrix equation [Eq. (1)] provides a complete set of atomic modes for the molecule. However, as the number of atoms in the molecule increases, the amount of computer memory required to store the dynamical matrix elements and the time required to diagonalize it (scaling as order N^3) quickly become unfeasible. At the present time, the number of atoms that can be treated by a direct diagonalization of the dynamical matrix is roughly on the order of a few thousands, which leaves large proteins and protein complexes like viruses with hundreds of thousands or millions of atoms out of reach.

Past normal-mode analysis of large proteins or macromolecular structures has relied on methods such as continuum elastic theory [12], elastic network models (ENMs) [13,14], or the rotation translation block method [15]. These methods coarse grain the molecule (i.e., use a reduced basis set to construct the vibrational modes), so that a diagonalization of Eq. (1) is possible. Recent work has extended continuum elasticity theory to icosahedral symmetry [16].

In this paper, we use a method based on order N techniques from electronic structure theory [17] to determine the low-frequency normal modes and frequencies of three large icosahedral viruses (with 10^5 – 10^6 atoms), the polio virus (PV), the cowpea chlorotic mottle virus (CCMV), and the hepatitis B virus (HBV) using a fully atomistic classical force-field model and a *full* basis set. A full basis set means that unlike coarse-graining procedures, the vibrational modes are constructed from the individual atomic displacements (three for each atom in the system) and hence is complete. The work presented here provides (1) a more detailed description of the methods used to determine the atomistic normal modes of icosahedral viruses [18], particularly the group theory aspects, (2) a comparison of atomistic calculations with other coarse-graining normal-mode analysis performed on the same viruses [15,19], and (3) a Raman spectra profile of each virus based on a bond polarizability model [20–23].

We note that all of the normal modes calculated in this work are for the capsid without genomic material inside. While the genome will likely affect the vibrational modes of the capsid, it is often just as important to calculate the normal modes of the empty capsid. For example, many empty capsids undergo either a reversible or irreversible swelling under the action of a protease or a simple change in pH . One such example is the cowpea chlorotic mottle virus (studied in this work) which undergoes a reversible swelling when the pH is increased. Similarly, the Hong Kong 97 (HK97) virus, undergoes a maturation event (an irreversible swelling of the empty procapsid) *in vitro* when the delta domain of the protein consisting of the 102 N -terminal amino acids is cleaved and the pH is subsequently changed [24]. In the case of HK97, this maturation event is likely to be an important step in virus assembly as the ability of the empty HK97 procapsid to swell is a necessary requirement to accommodate the genome during packaging by its terminase.

First, we briefly review the phonon functional method. Next, we provide a brief review of group theory followed by a description of how the operation of the smaller symmetry related dynamical matrices for an irreducible representation (irrep.) of the group can be performed without storage of

either the full dynamical matrix or the smaller symmetry related dynamical matrices. The following section applies the methods discussed to the polio virus, cowpea chlorotic mottle virus, and hepatitis B virus and compares the low-frequency fully atomistic modes with the low-frequency modes calculated in other coarse-grained calculations providing information about the reliability of using coarse-graining methods in the predictions of normal modes. In addition, Raman spectra are calculated for each virus using a bond polarizability model. Finally, the work is summarized and concluded.

II. ORDER N METHOD FOR ELECTRONIC STRUCTURE CALCULATIONS

Often in complexes such as viruses, the low-frequency vibrations (e.g., <25 cm^{-1}) tend to be of the most interest in a NMA since these modes illustrate the large global motions of the molecule. These large global motions are also the ones that are most likely to break apart the virus during resonant excitations with external probes. For a large molecule there are many high-frequency states and only a few low-frequency states that are of particular interest.

Our strategy is to use a phonon energy functional [18] that, when minimized, contains the subset of M lowest-frequency eigenvectors (and eigenvalues) of the dynamical matrix. The phonon functional G_p is deceptively simple,

$$G_p = \text{Min}\{\text{Tr}_M[H + H(1 - S)]\}. \quad (3)$$

The matrices H “Hamiltonian” and S “overlap” are *small* $M \times M$ matrices with elements defined in terms of the dynamical matrix \vec{D} and the vectors $|u\rangle$,

$$H_{ij} = \langle u_i | \hat{D}_s | u_j \rangle, \\ S_{ij} = \langle u_i | u_j \rangle. \quad (4)$$

The value of M is chosen by the user and specifies the number of lowest-frequency states that are to be obtained from the full $3N$ spectrum of states. The operator \hat{D}_s is the shifted dynamical matrix operator $\hat{D}_s = \hat{D} - \lambda_L \hat{I}$, where λ_L is the largest eigenvalue of \hat{D} . The purpose of the shift is to produce a negative-definite spectrum which guarantees to produce an energy minimum in the phonon energy functional [17]. In Fig. 1 we plot a contour plot of the energy functional for a two-atom dimer in one dimension. We write the displacement pattern as a two-component vector (C_1, C_2) for the displacements of atoms 1 and 2, respectively. There are two modes—the translational mode of zero frequency and the stretching mode. In this trivial example, the lowest-frequency mode is the translation, and for $M=1$ in the energy functional, this is the state that will be produced, $|u\rangle = \pm(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. As can be seen from Fig. 1, a minimum is produced when the eigenvalue is negative (frequencies shifted downward) and a maximum is produced when the frequencies are unshifted and allowed to be positive. To ensure that the phonon energy functional produces an energy minimum, the spectrum of \hat{D} is shifted by the largest eigenvalue pro-

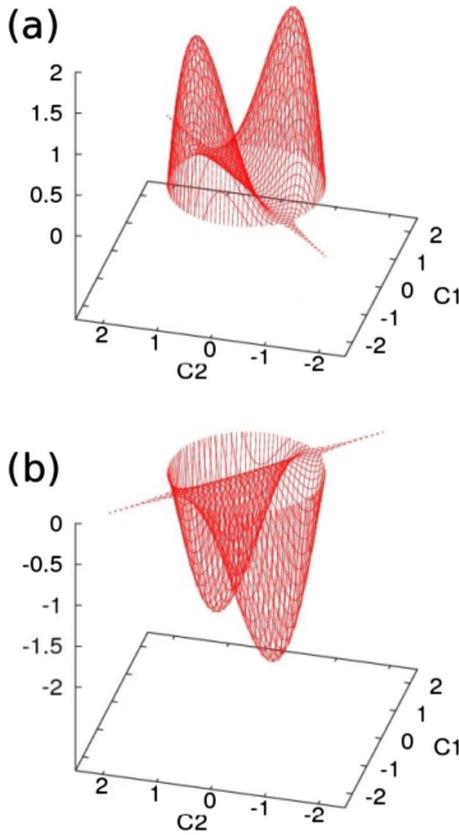


FIG. 1. (Color online) The energy landscape of the phonon functional for a dimer molecule in one dimension. The values C_1 and C_2 label the displacement pattern of the dimer molecule as a two-component vector (C_1, C_2) . The lowest-frequency mode is the translational mode at zero frequency. The landscape for (a) the unshifted dynamical matrix and (b) the shifted dynamical matrix with all negative eigenvalues. Note that (a) forms a maximum at the high-frequency eigenvector $(C_1, C_2) = \pm(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, while (b) forms a minimum at the correct low-frequency mode $(C_1, C_2) = \pm(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

ducing an entirely negative spectrum. The shift leaves the eigenvectors unaffected. The functional is very similar to that used in electronic structure theory [17], except that there are no “occupied” and “unoccupied” subspaces. We simply choose M desired states that are the lowest in frequency.

The energy functional, when minimized, produces the exact eigenvalues and eigenvectors of the lowest M modes of the dynamical matrix. Figure 2 shows a block diagram of the process to achieve this which involves seven (7) major steps. We now describe the steps in some details.

The user first decides on the number (M) of eigenstates desired (step 1). Typically $M=100$ is sufficient to reach frequencies around 10–25 cm^{-1} (depending on the size of the virus) and this is typically the number used in the present work. The search for the M lowest phonon states of \hat{D} begins by choosing a random set of M vectors $|u_i\rangle$ (step 2) that need not be orthogonal or even normalized as the phonon functional will enforce orthonormality. For each vector, the shifted dynamical matrix is applied (step 3), and the matrices H_{ij} and S_{ij} are formed. The operator \hat{D}_s acting on a vector is accomplished by treating each component separately—much

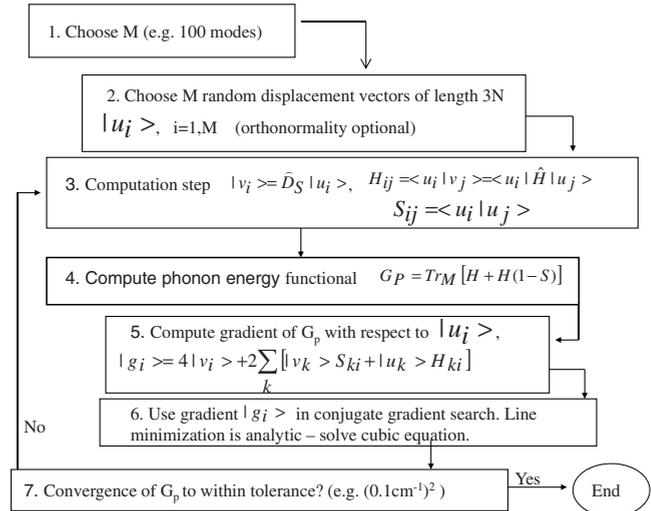


FIG. 2. The flow chart block diagram of the energy functional.

like determining the force on an atom in a molecular dynamics (MD) step. An MD step is tantamount to computing, atom by atom, the change in the energy to a first-order virtual displacement. The operation of \hat{D}_s on a vector component is the second-order displacement. Computationally, the time necessary for the second derivative is marginally more than the first derivative, so that applying \hat{D}_s to a vector is roughly equivalent (in CPU time) to an MD step. For M states, this is equivalent to M MD steps. Thus, the full dynamical matrix is never stored or computed, but rather only the small $M \times M$ matrices H_{ij} and S_{ij} and M vectors $|u_i\rangle$.

The phonon energy functional [Eq. (3)] is then computed and minimized (steps 4–6). Although the phonon functional is a multivariate function of $3MN$ variables (M vectors times $3N$ components) the minimization procedure proceeds quickly since *any* set of orthonormal vectors that span the subspace of the M lowest states \hat{D} is a solution of the minimization. The minimization of the energy functional is performed using an iterative conjugate gradient method (steps 5 and 6). Generally the conjugate gradient method requires a one-dimensional line minimization along a search direction. This minimization normally is tedious and a relatively expensive computation. Here, fortunately, the line minimization can be accomplished analytically. The condition for a minimum generates a cubic equation which can be easily solved to determine the step length needed to find the minimum. Complete details of this step have been given in Ref. [32]. The minimization procedure will require only M dynamical matrix operations (on each of the search directions) for every minimization step. The total number of full vectors of length $3N$ that must be stored is $5M$.

The convergence of the functional G_P can be monitored during the iteration process. Figure 3, which will be described fully in the results section, shows convergence for the polio virus as a function of iteration step. A few thousand steps are sufficient to converge the frequencies to a small fraction of a wave number. In general, about 3000 minimization steps are sufficient to converge the phonon functional

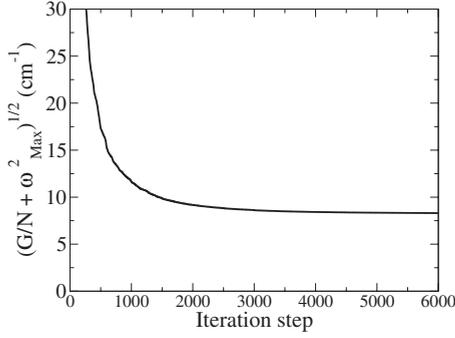


FIG. 3. Convergence of the phonon functional G for the A symmetry modes of the polio virus. The function on the ordinate is tantamount to convergence of the root-mean-square frequency of the lowest 50 modes.

when $M=100$. Since each minimization step requires M operations of the dynamical matrix on a vector, the rough equivalent of M molecular-dynamics steps, the computational effort to compute the M lowest-frequency modes can be thought of as 3×10^5 molecular-dynamics steps (or as a 0.3 ns molecular-dynamics simulation). For a virus where symmetry can be used to reduce the size of the problem, the computational time can be thought of as 0.3 ns molecular-dynamics simulation on a single unit cell (1 of 60).

After the minimization, the vectors $|u_i\rangle$ are an orthonormal set of vectors that span the space of the lowest M states of \hat{D} . The true eigenvectors of $\hat{D}(|e_i\rangle)$ are obtained from a diagonalization of the small $M \times M$ matrix equation,

$$HC(i) = \lambda' SC(i). \quad (5)$$

The eigenvalues λ' are shifted and are related to the true eigenvalues of \hat{D} via $\lambda' = \lambda - \lambda_L$. The vector $C(i)$ of length M gives the linear combination of basis vectors $|u\rangle$ that produce the eigenvector $|e_i\rangle$, i.e.,

$$|e_i\rangle = \sum_{j=1}^M C_j(i) |u_j\rangle. \quad (6)$$

III. SYMMETRY

Although the phonon functional method is capable of determining the lowest-frequency modes of large ($>10^6$) matrices that result for icosahedral viruses, the use of group theory can (i) reduce the computational expense of minimizing the phonon functional and (ii) group the eigenvectors in categories which describe their irreps. The last point is particularly important when searching for modes that may be sensitive to stimulation by external probes, such as Raman light scattering, which have strict symmetry selection rules (A and H irreps. are “allowed” in icosahedral symmetry).

To use group theory in the phonon functional method, a method to calculate the operation of the smaller symmetry related dynamical matrices on a vector is required. This is further complicated by the requirement that the matrix not be stored in computer memory as it is typically too large. First we briefly review some background from group theory then

TABLE I. The character table of the icosahedral group I . The quantity τ is the golden mean, $\frac{1+\sqrt{5}}{2}$.

	$1E$	$12C_5$	$12C_5^2$	$20C_3$	$15C_2$
A	1	1	1	1	1
T_1	3	τ	$1-\tau$	0	-1
T_2	3	$1-\tau$	τ	0	-1
G	4	-1	-1	1	0
H	5	0	0	-1	1

follow with a description of a method that determines the symmetry related dynamical matrices times a vector without requiring storage of the matrix.

A. Background and review

The icosahedral point group I , is one of very high symmetry. Group theory is used to classify the modes according to their irreducible representations, but can also be used to simplify the computation of the eigenvectors and frequencies. The character table of group I is given in Table I. The symmetry operators for an element T of the group are denoted by $\hat{R}(T)$ and fall into one of five classes; the identity E , 12 C_5 operations, 12 C_5^2 operations, 20 operations around C_3 axes, and 15 operations around C_2 axes. There are five irreps. associated with the group I : A , T_1 , T_2 , G , and H with degeneracies of 1, 3, 3, 4, and 5, respectively.

The total number of symmetry operations is n_G , where n_G is 60 for the regular icosahedral group. These operations relate the proteins on the coat by symmetry. Capsids with triangulation number $T=n$ have $60n$ proteins on the coat. We thus consider the coat to contain 60 cells (which we index as $J=[1, 60]$). Each cell contains n proteins and each protein is copied 60 times by the application of the 60 symmetry operators $\hat{R}(T)$. The total number of degrees of freedom N_f of the complete viral capsid is thus $N_f=60 \times 3N_b$, where N_b is the number of atoms (“basis” atoms) in the n proteins in a single cell.

Each of the N_f eigenvectors of the $N_f \times N_f$ Cartesian dynamical matrix \hat{D} for the complete virus can be separated and labeled by three indices; the irrep. p , row m , and appearance number α . The index p ranges from $[1, 5]$ and denotes one of the five irreps. (A , T_1 , etc.) of group I . The index m ranges from $[1, d_p]$ and labels the eigenvector by the d_p orthogonal degenerate states. Finally, α ranges from $[1, n_p]$, where n_p is the number of times that irrep. p “appears” in the Cartesian dynamical matrix \hat{D} . One can easily verify that $n_p=3N_b d_p$ and, using the property $n_g=\sum_p d_p^2$, that the total number of eigenvectors is equal to $\sum_p n_p d_p=N_f$.

The eigenvectors of the full virus capsid $\hat{e}_{m\alpha}^p$ (of length N_f) of row m for irrep. p transform according to the irreducible representation Γ matrices $\Gamma_{mn}^p(T)$,

$$\hat{R}(T)\hat{e}_{n\alpha}^p = \sum_{m=1}^{d_p} \Gamma_{mn}^p(T)\hat{e}_{m\alpha}^p. \quad (7)$$

Similarly, any linear combination of eigenvectors from a row m of irrep. p must also satisfy Eq. (7). Since the dynamical

matrix is invariant under any group operator, i.e.,

$$\hat{D} = R^\dagger(T)\hat{D}R(T) \quad (8)$$

for every group element T , it is easy to show that any basis set of orthogonal vectors $\vec{v}_{m\alpha}^p$ of length N_f which transform according to Eq. (7) block diagonalizes the Cartesian dynamical matrix to the form

$$\hat{D} = \begin{bmatrix} \hat{D}^A & & & \\ & \hat{D}^{T_1} & & \\ & & D^{T_1} & \\ & & & \ddots \end{bmatrix}, \quad (9)$$

where the smaller matrices \hat{D}^p have dimension $n_p \times n_p$ and occur exactly d_p times along the diagonal, giving rise to the degeneracy of eigenvectors from irrep. p . The power of group theory is shown by the fact that it reduced the size of the full $N_f \times N_f$ dynamical matrix to a set of 5 (for group I) smaller $n_p \times n_p$ dynamical matrices. The smaller dynamical matrices are diagonalized by the symmetry eigenvectors \vec{q}_α^p (no m index) of length n_p , with components that give the linear combination of symmetry basis vectors $\vec{v}_{m\alpha}^p$ that construct the full eigenvector $\vec{e}_{m\alpha}^p$,

$$\vec{e}_{m\alpha}^p = \sum_{\beta} \vec{v}_{m\beta}^p (q_\alpha^p)_\beta. \quad (10)$$

Note that no m index is required for the symmetry eigenvector \vec{q}_α^p as the other degenerate orthogonal eigenvectors for irrep. p can be constructed by using orthogonal symmetry basis vectors that transform as a different row m , or rotating the eigenvector using Eq. (7). As can be seen in Eq. (10), a set of orthogonal symmetry basis vectors $\vec{v}_{m\beta}^p$ of length N_f is required for both the construction of the full eigenvectors of the Cartesian dynamical matrix and the operation of a smaller symmetry related dynamical matrices on a vector. The next section discusses a method that uses symmetry basis vectors from the atomic icosahedron C_{60} (which have a length of 180) to perform the operation of the smaller symmetry related dynamical matrix on a vector.

B. Dynamical matrix operator in symmetry coordinates

In light of the previous section, the symmetry related dynamical matrix for irrep. p can be constructed from the following formula:

$$\vec{\vec{D}}^p = \vec{V}^\dagger \vec{\vec{D}} \vec{V}, \quad (11)$$

where the matrix \vec{V} is the $N_f \times n_p$ matrix containing a set of orthogonal symmetry basis vectors which transform according to Eq. (7). Only basis vectors that transform as a single row are needed in Eq. (11) since symmetry basis vectors from other rows will give an identical $\vec{\vec{D}}^p$.

There are many ways to construct symmetry basis vectors that transform according to Eq. (7). A simple scheme makes use of the projector operator in group theory,

$$\hat{\mathcal{P}}_{mn}^p = \frac{d_p}{n_g} \sum_{T \in \mathcal{G}} \Gamma_{mn}^{p*}(T) \hat{R}(T), \quad (12)$$

by projecting a set of random n_p vectors onto the basis space. Afterward, the Gram-Schmidt orthogonalization of the projected vectors results in a set of vectors that span the space of basis vectors that transform as a single row. Despite the effectiveness and simplicity of this method, this scheme is particularly cumbersome to use here for two reasons. First, the computer memory required to store the full $N_f \times n_p$ basis vector matrix is on the scale of the dynamical matrix. Second, the Gram-Schmidt procedure required to produce an orthogonal set of vectors is time consuming for large viruses since $n_p = 3N_b d_p$ scales with the number of basis atoms.

Instead, we construct a set of orthogonal symmetry basis vectors by a simpler scheme which uses symmetry basis vectors from the atomic icosahedron C_{60} . For the atomic icosahedron there are $3d_p$ orthogonal symmetry basis vectors of length 180 to construct for each irrep. p , giving 180 vectors in total. The symmetry basis vectors for the atomic icosahedron can be represented as vectors of length 3 labeled by cell number, $J=[1,60]$, irrep. p , row m , and appearance $\beta = [1, 3d_p]$, $\vec{v}_{J,m\beta}^p$. The full set for all irreps. can be easily stored in a small 180×180 matrix. The set of $n_p = 3N_b d_p$ orthogonal symmetry basis vectors for the virus can then be constructed by applying each C_{60} symmetry basis vector to each atom in each cell of the protein coat, one at a time. This results in a full set of symmetry basis vectors of length N_f that are automatically orthogonal and satisfy Eq. (7). In addition there may be some advantages to choosing the symmetry basis vectors to be the eigenvectors of C_{60} . If the low-frequency modes of the virus are expected to have similarities to the modes of C_{60} , then using the eigenvectors of C_{60} as symmetry basis vectors for the virus could precondition the resulting symmetry related dynamical matrix to something that is nearly diagonal. This possibility will be explored further in the future. For now we simply construct symmetry basis vectors for the point group using random vectors and the projection with the Gram-Schmidt orthogonalization procedure which proceeds rapidly for the point group.

Using the symmetry basis vectors from the atomic icosahedron $\vec{v}_{J,m\beta}^p$ and Eq. (11), the operation of the $n_p \times n_p$ symmetry related dynamical matrix on one of the $M=100$ symmetry vectors of length $n_p = 3d_p N_b$ (q_β^p with $\beta=[1, n_p]$ labeling the components of the vector) can be computed in three steps. First, $m=[1, d_p]$ vectors of length $3N_b$, \vec{z}_m^p , are formed using the point group symmetry basis vectors from $J=1$ only,

$$\vec{z}_m^p = \sum_{\beta=1}^{3d_p} \vec{v}_{1,m\beta}^p q_\beta^p. \quad (13)$$

It is important to note that the first $\beta=3d_p$ components of the vector \vec{q}^p are multiplied by the symmetry basis vectors for the point group (of length 3) and summed in Eq. (13) to give the first three components of the vector \vec{z}_m^p . Since the symmetry vector \vec{q}^p has a total length $n_p = 3d_p N_b$, this can be done a total of N_b times to give all $3N_b$ components of the

vector \vec{z}_m^p . This performs the first multiplication $\vec{V}^p \vec{q}^p$ in Eq. (11) but keeps only the portion of the vector which corresponds to cell $J=1$. In the second step, the vectors \vec{z}_m^p are multiplied by the Cartesian dynamical matrix. This performs the next multiplication, i.e., $\vec{D}^p \vec{q}^p$. Since the vector \vec{z}_m^p is (by construction) a vector in cell $J=1$, only the portions of the dynamical matrix which describe coupling of a cell J with cell 1, $D_{J,1}$, contribute. The result is d_p vectors of length N_f , $\vec{w}_{J,m}^p$, which can be written as $J=[1, 60]$ vectors of length $3N_b$,

$$\vec{w}_{J,m}^p = \vec{D}_{J,1} \vec{z}_m^p. \quad (14)$$

Finally, the third step performs the last multiplication by \vec{V}^\dagger to give the $3d_p N_b$ components of the symmetry related dynamical matrix times the vector \vec{q}^p ,

$$\vec{D}^p \vec{q}^p = \frac{n_g}{d_p} \sum_{J,m} \vec{v}_{J,m\beta}^p \cdot \vec{w}_{J,m}^p. \quad (15)$$

The dot product of the symmetry basis vectors $\vec{v}_{J,m\beta}^p$ (of length 3) with the first three components of $\vec{w}_{J,m}^p$ gives the first $3d_p$ components of the operator in Eq. (15). Since $\vec{w}_{J,m}^p$ has length $3N_b$, this can be done N_b times to give all $3d_p N_b = n_p$ components of the vector. The third multiplication is done on the fly and added in to the final result as the components of $\vec{w}_{J,m}^p$ are constructed. This prevents the need to store all $J=[1, 60]$ vectors $\vec{w}_{J,m}^p$, which can be quite large. The entire computation of \vec{D}^p times a vector proceeds in roughly order ($d_p N_b$) steps and requires a work space of only d_p vectors of length $3N_b$, which is equivalent to the storage requirement for a single symmetry vector \vec{q}^p .

IV. LOW-FREQUENCY MODES OF THE POLIO VIRUS

The PV is a member of the picornavirus (small RNA viruses) family of viruses, which contain the genus members *enterovirus* and *rhinovirus*. The polio virus and the virus producing the common cold are related in that they both belong to the picornavirus family with PV in the genus *enterovirus* and the common cold virus in the *rhinovirus* genus. The polio virus is a severe pathogen in humans and is the cause of poliomyelitis, which is a disease that affects the central nervous system. It can produce paralysis and can be fatal if it produces paralysis in respiratory muscles. The polio virus historically has attacked in developed countries because of advances in hygiene. Natural infections in infants were greatly reduced which led to susceptible adults and an increased likelihood of epidemic spread of the disease. A polio vaccine was introduced by Salk in 1955, which consisted of injections of an inactivated virus to produce effective antibodies. The incidence of the disease has dropped precipitously and the disease has the potential to be eradicated worldwide.

The structure of the polio virus that we use is PV-type 1, Mahoney strain, and was obtained by x-ray crystallography by Grant *et al.* [25]. The coordinates are available from the protein data bank (PDB) [26] and the PDB file used is 1VBD. The virus is a $T=3$ like capsid assembled from three proteins (VP0, VP1, and VP3) at a single symmetry site of

the capsid. Upon maturation, VP0 is cleaved into two separate proteins (VP2 and VP4) via viral protease giving the structure found in the PDB file. The number of atoms in a primitive site including hydrogen is 13 074. The x-ray structure does not include H-atom positions. We assigned positions based on chemical considerations at a pH value of 7. Relaxation of the structure (see below) produces a structure where no two hydrogen atoms are too close. The total number of atoms in the full icosahedron is $60 \times 13\,074 = 784\,440$. The total number of normal modes is $3 \times 784\,440 = 2\,353\,320$. The phonon functional will find just a small subset of these; specifically the low-frequency modes from each irreducible representation of the group I .

The first step is to optimize the structure to its equilibrium zero force structure within the force field that we are using. We mention in passing that the phonon functional method does not mathematically require a zero force structure and that it can be used to calculate normal modes within the harmonic approximation when the net force on the structure is not zero. We relaxed the structure using the conjugate gradient technique. We performed approximately 4000 conjugate gradient steps, which resulted in the total energy of the structure changing by 0.0032 eV/atom (or about 40 K in temperature units). At equilibrium, the residual root-mean-squared (RMS) force is 3×10^{-5} eV/Å and the RMS deviation from the initial atomic positions is 1.78 Å.

Next, we compute the normal modes by minimizing the phonon functional. We choose to seek convergence of the lowest $M_p=50$ modes for *each* irrep. The convergence of the energy functional G_p with a conjugate gradient step is shown in Fig. 3 for the A modes of the polio virus. Because of the shift of the eigenvalues by $-\omega_{max}^2$, the value of the functional is shifted from its true value by $-M_p \omega_{max}^2$. This is a very large value and, at first, appears to make seeking convergence of the phonon functional to within a fraction of 1 cm^{-1} very difficult. The value of the functional upon convergence is $G_p = M_p [\omega_{RMS}^2 - \omega_{max}^2]$, where ω_{RMS} is the RMS value of the frequency of the modes ($\omega_{RMS}^2 = \sum_{i=1}^{M_p} \omega_i^2 / M_p$). Thus, instead of G_p we plot in Fig. 3 the quantity $\sqrt{G_p / M_p + \omega_{max}^2}$, which theoretically is the RMS frequency of the converged modes ω_{RMS} versus step number. The total number of conjugate gradient steps performed is 6000. It is evident that the convergence is monotonic and rapid. After 1000 steps ω_{RMS} is within 3.0 cm^{-1} of the converged result, and after 3000 steps it is within 0.3 cm^{-1} . The difference between ω_{RMS} at 5000 and 6000 steps is 0.04 cm^{-1} . Thus, the frequencies can be obtained with a very high accuracy. The method has very favorable convergence properties in spite of the ill-conditioned nature of the eigenvalue spectrum, which has a ratio of high to low eigenvalues of approximately 10^7 [$\omega_{min}^2 \approx (1 \text{ cm}^{-1})^2$, $\omega_{max}^2 \approx (4000 \text{ cm}^{-1})^2$].

A schematic spectrum for the first 50 modes of each symmetry is shown in Fig. 4 (the six zero-frequency modes from translation and rotation are not included). The A -symmetric modes are symmetric under any group operator, and they are most spread out in frequency ranging from about 1 to 12 cm^{-1} . Modes of other symmetries have frequencies that range from about 1 to 8 cm^{-1} . The mode of lowest frequency is of H symmetry. The values of the frequencies are listed in Table II for the first 20 modes. In addition, the

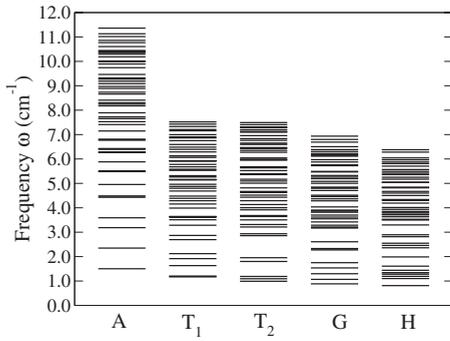


FIG. 4. Stick spectrum of the lowest 50 modes of the polio virus for each irreducible representation of the symmetry group I .

participation number (W) [18] is also given. The participation number measures the number of atoms participating in the displacement pattern for that mode and is obtained from the spatial extent of the eigenvectors. The number of atoms given in the participation number is given relative to the number of atoms in one cell (out of 60) of the icosahedral capsid. Most low-frequency modes are delocalized over a large fraction of the atoms, but occasionally there is a mode with just a few hundred atoms vibrating. These localized modes are located on “soft” spots of the capsids protein such as hairpin turns that are at the capsid surface or even a small segment of alpha helix that would be exposed to solvent. These modes are likely missed in an elastic network model which uses a simplified phenomenological spring potential between atoms within a cutoff distance.

TABLE II. Frequencies (ω_i in cm^{-1}) and participation numbers (W) per protein unit of polio for each of the irreducible representations (group I). Full participation is 13 074 atoms. Raman active modes are of A and H symmetries.

A		T_1		T_2		G		H	
ω	W								
1.50	10617	1.17	455	0.99	10347	0.88	5257	0.81	12446
2.35	10676	1.2	358	1.09	556	1.07	1301	1.1	1673
3.18	9313	1.63	9980	1.19	574	1.3	1721	1.19	3200
3.59	281	1.92	12099	1.8	10521	1.54	6980	1.28	1951
4.43	6529	2.12	10451	1.95	10528	1.75	11102	1.34	1022
4.49	6536	2.7	11816	2.86	10084	2.27	11432	1.44	2767
4.95	4048	2.87	11912	2.94	9462	2.33	12040	1.61	3265
5.48	5094	3.28	8567	3.21	7664	2.61	11761	1.99	8314
5.51	1296	3.51	3432	3.31	8454	3.17	9999	2.36	12224
5.88	6604	3.62	295	3.5	576	3.21	8177	2.46	10683
6.26	4715	3.63	410	3.64	495	3.29	6829	2.55	11472
6.29	3654	3.64	1240	3.68	547	3.42	1133	2.81	10628
6.39	8455	3.99	7265	3.92	9962	3.55	668	2.89	10903
6.43	6876	4.15	3061	4	6074	3.62	2565	3.3	7923
6.77	2082	4.29	1218	4.13	4131	3.7	641	3.5	2780
6.81	4525	4.39	5863	4.29	1407	3.82	2867	3.53	5349
7.15	4900	4.48	3367	4.43	4752	3.87	3419	3.63	2986
7.41	5114	4.69	6223	4.52	4302	3.9	8690	3.71	966
7.53	3500	4.8	8590	4.63	6312	4.04	3422	3.74	1851
7.65	3953	4.87	6539	4.67	7989	4.28	1896	3.81	1075

TABLE III. A comparison of the lowest three modes for the polio virus computed with all degrees of freedom (present work) and with dihedral degrees of freedom [19]. The frequencies are in cm^{-1} units and the values in parentheses are from Ref. [19].

A	T_1	T_2	G	H
1.50 (1.9)	1.63 (2.1)	0.99 (1.3)	0.88 (1.3)	0.81 (0.9)
2.35 (3.0)	1.92 (2.3)	1.80 (2.3)	1.54 (1.8)	1.19 (1.4)
3.18 (4.3)	2.12 (2.4)	1.95 (2.4)	1.75 (2.2)	1.44 (1.7)

The normal modes of the polio virus have been computed by van Vlijmen and Karplus [19] using a restricted basis set. They allowed only flexibility of the dihedral degrees of freedom, except that the dihedrals of the peptide bond were assumed rigid. A comparison is made between the two calculations for the three lowest modes of each frequency in Table III. Since our model includes all degrees of freedom, we find some modes of low frequency, which are fairly localized on a few hundred atoms. These are unlikely to be found using the smaller basis of only dihedral degrees of freedom and so we have included modes with a participation number of greater than 2000 from our analysis in the comparison. The comparison shown in Table III is striking in that the agreement between the two calculations is strong. Both calculations predict that the lowest-frequency modes are H , with G and T_2 just slightly higher, then A , and finally T_1 . There is a clear trend that the frequency of modes computed with all

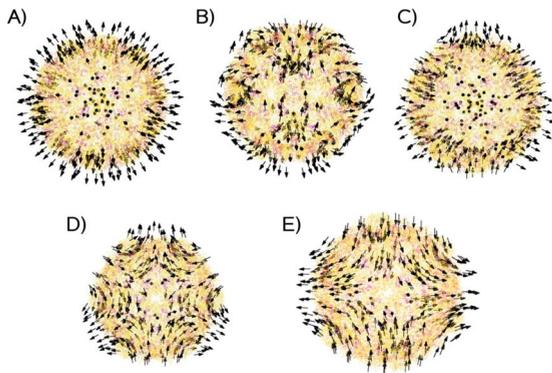


FIG. 5. (Color online) Displacement patterns of the polio virus for the lowest-frequency mode in each irreducible representation A , T_1 , T_2 , G , and H of the icosahedral group. Lowest-frequency (A) A , (B) T_1 , (C) T_2 , (D) G , and (E) H modes.

degrees of freedom is lower in frequency than if only dihedral degrees of freedom are included.

There are 15 modes in Table III and an average over these modes shows a percent frequency difference (defined as $\delta\omega/2\langle\omega\rangle$) of 13%. Here, $\delta\omega$ is the difference between the two frequencies and $\langle\omega\rangle$ is the average of the two frequencies. Part of this difference may be due to the difference in force-field models used. The authors of Ref. [19] used a CHARMM [27] based parameter set [28], while we use a AMBER [29] based generalized Born parameter set [30,31].

Figure 5 illustrates one displacement pattern from each of the irreducible representations A , T_1 , T_2 , G , and H . Figure 5(A) shows the lowest A mode at 1.50 cm^{-1} , which corresponds to the breathing mode. The T_1 mode at 1.92 cm^{-1} is shown in Fig. 5(B), the lowest T_2 mode at 0.99 cm^{-1} in Fig. 5(C), the lowest G mode at 0.88 cm^{-1} in Fig. 5(D), and the lowest H mode at 0.81 cm^{-1} in Fig. 5(E). The arrows in each figure represent the center-of-mass motion of one of the 4×60 proteins (VP1, VP3, or the cleaved VP2 and VP4) in the capsid shell.

Using a simplified bond polarizability model [20,21], the atomic displacement patterns and frequencies can be used to predict a low-frequency Raman scattering profile for the polio virus. Although the polarizability of bonds vary, we use a single set of polarization parameters (specifically Ref. [23]) to simplify the analysis and produce a semiquantitative picture of the Raman profile. Earlier we have used the atomic displacement patterns from the phonon functional method to predict the low-frequency Raman scattering profile of the tubular M13 bacteriophage. The results showed a reasonably good fit between experiment and theory [32,33]. Figure 6 shows the predicted Raman scattering profile for the Raman active A and H modes for polio. The A modes are shown in solid black line, while the H modes are shown in the dashed line. The degeneracy of the H modes contributes to the large difference in the Raman scattering intensity. This is in contrast to C_{60} where A and H mode scatterings are comparable. The origin of the weaker A mode scattering in virus particles likely reflects that uniform A modes such as the breathing will result in very little covalent bond stretching since the entire protein units move in concert.

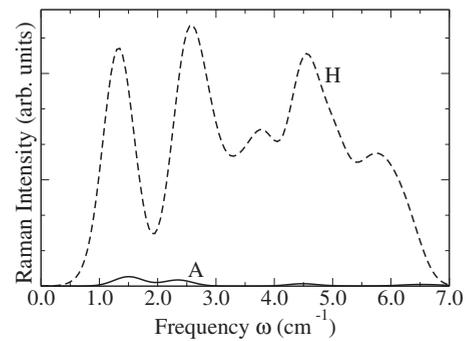


FIG. 6. The Raman intensity profile of the polio virus up to 7 cm^{-1} . Raman active modes are of A (solid line) and H (dashed line) symmetries.

V. LOW-FREQUENCY MODES OF HEPATITIS B

HBV is an infectious virus of vertebrates and is an exemplar of the hepadnaviridae (hepa+dna) family. The virus is a serious human pathogen that chronically infects over 350×10^6 people worldwide, which results in over 1×10^6 deaths annually [34]. Worldwide the virus infects 2×10^9 people. The HBV infection produces a serious inflammation of the liver (hepatitis) leading to liver failure or cirrhosis. The hepatitis virus also produces hepatocellular carcinoma in the liver, particularly in nonwestern countries.

The capsid of HBV surrounds DNA to protect it, while the capsid itself is surrounded with a lipid envelope which contains surface proteins encoded by the virus. The envelope is taken from its host. We model only the protein coat of the capsid and neglect the lipid membrane and the DNA contained inside. The capsid protein consists of replicas of a single polypeptide chain labeled HBcAg. The proteins arrange themselves in an icosahedral capsid that is dimorphic; it can be triangulated on the icosahedral faces as either a $T=3$ or a $T=4$ capsid [35]. The capsid has either 180 or 240 copies of the polypeptide HBcAg monomer. The assembly of either capsid occurs through dimer units. The $T=3$ capsid contains three symmetry-inequivalent core HBcAg particles A , B , and C . There are 60 AB dimer units and 30 CC dimer units. The $T=4$ capsid contains four symmetry-inequivalent core particles A , B , C , and D with 60 AB dimers and 60 CD dimers [36]. The switch determining which triangulation occurs is not yet clearly determined although proteins truncated near the C terminus have a strong influence on the final assembled structure. This region is rich in arginine which facilitates the electrostatic interaction of the capsid with the enclosed nucleic acids.

The local packing of subunits is very similar in the two sizes of particles. The arrangement is such that there exists a hydrophobic core that forms the base of the capsid, and the dimers produce spikes projecting out of the capsid. The spikes contain a four-helix bundle. The outermost tips of these spikes are believed to be the location of the primary antigenic site [37]. The arrangement of the proteins in either the $T=3$ or $T=4$ virus is such that there are holes in the capsid allowing the exchange of small molecules.

The particular capsid we study is a $T=4$ variety. The full monomer has a very basic C-terminal region, and this region

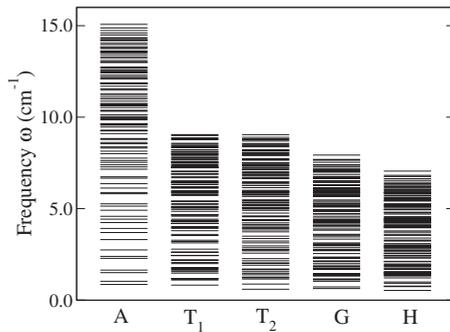


FIG. 7. Stick spectrum of the lowest 100 modes of the hepatitis B (HBV) virus for each irreducible representation of the symmetry group I .

interacts strongly with the charged DNA contained within the capsid. The gene can be expressed within *E-coli*, and this gene has been altered, so that the C-terminal sequence is trimmed from the monomer. These monomers assemble predominately into $T=4$ capsids. Specifically, the capsid we use to determine the vibrational modes includes the amino-acid sequence from residues 1–149 out of the 183 amino-acid sequence, which has been crystallized and its structure determined by x-ray diffraction [38], PDB code 1QGT. The capsid assembly is very similar to those found in an infected liver.

The HBV $T=4$ virus has four identical proteins in the unit cell, where each of the four proteins in the cell is the same. The total number of atoms in a unit cell is 8980. The full

icosahedron thus produces a total of more than 1.6×10^6 modes. The procedure for obtaining the vibrational modes is similar to that used for the polio example. The PDB structure was first relaxed to a zero force equilibrium structure. The RMS deviation from the input structure and the relaxed structure was 2.41 Å and the residual forces are reduced to an RMS value of 3.7×10^{-5} eV/Å.

We choose to determine the small subset of M_p modes with $M_p=100$. A schematic stick spectrum of the lowest 100 modes for each irreducible representation is shown in Fig. 7. The A -symmetric modes are the most spread out in frequency ranging from 0.9 to 15 cm^{-1} . Modes of other frequencies range from 0.5 to 9.0 cm^{-1} . The lowest-frequency modes are of T_2 , G , or H symmetry, being within 0.1 cm^{-1} of each other.

Table IV gives the frequencies of the first 20 modes for each irreducible representation and indicates localization (or lack thereof) by the participation number W . Full participation of all atoms in a mode (such as a translation) is 8980, corresponding to the number of atoms in a fundamental unit of the icosahedron. One notices that the first two modes of each symmetry are highly cooperative with $W=6500$ –8900. In fact, the first 20 modes of every symmetry are cooperative. The lowest value of W is the 17th A -symmetric mode with $W=1939$. This indicates that there are no localized floppy regions on the capsid within this manifold of modes.

Displacement patterns of the lowest modes for each symmetry are shown in Fig. 8. The arrows of the displacement pattern represent the center-of-mass motion for each one of the four core proteins in the unit. Interestingly, the lowest-

TABLE IV. Frequencies (ω_i in cm^{-1}) and participation numbers (W) per protein unit of hepatitis B virus for each of the irreducible representations (group I). Full participation is 8980 atoms. Raman active modes are of A and H symmetries.

A		T_1		T_2		G		H	
ω	W								
0.86	6512	0.82	7803	0.60	8538	0.64	8185	0.54	8906
1.03	8018	1.10	6918	0.88	8033	0.72	8172	0.74	8853
1.50	2086	1.17	7005	1.16	6867	1.03	7579	0.77	8852
1.64	3280	1.18	7906	1.23	7274	1.11	7804	0.92	8378
2.29	5804	1.48	2849	1.26	6367	1.16	6906	0.99	8423
2.39	4844	1.56	4618	1.39	3447	1.32	4601	1.22	6044
2.74	6580	1.64	4066	1.47	3602	1.38	4540	1.32	4925
3.31	6565	1.69	3553	1.56	2942	1.39	4220	1.36	6011
3.70	4966	1.76	4120	1.66	4212	1.41	5693	1.40	4622
3.92	2383	1.77	4539	1.70	3409	1.46	3512	1.43	4833
4.24	5203	2.00	6363	1.91	6711	1.69	5449	1.46	4061
4.43	6025	2.06	6539	2.01	6103	1.71	3857	1.50	3506
4.58	4516	2.18	7201	2.03	7802	1.76	4166	1.52	4800
4.91	5505	2.20	7395	2.15	7172	1.84	5714	1.56	3917
5.14	5777	2.43	5649	2.27	6957	1.92	5826	1.68	4525
5.26	3385	2.45	5722	2.57	3435	2.02	6289	1.72	4676
5.81	1939	2.61	7451	2.69	7491	2.11	7561	1.81	4540
5.88	2484	2.63	6981	2.76	4243	2.15	6892	1.87	5814
6.13	5296	2.79	6984	2.77	6435	2.28	7518	1.92	5839
6.36	4526	3.13	6464	2.86	6310	2.32	7073	1.98	6643

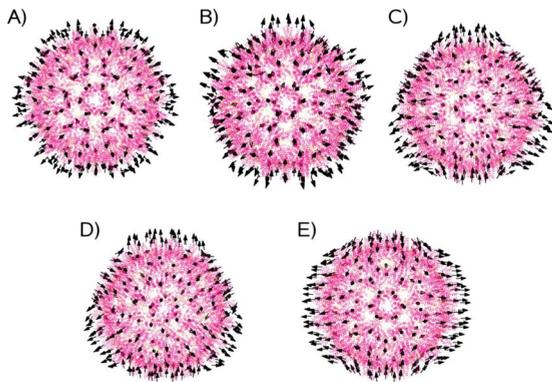


FIG. 8. (Color online) Displacement patterns of the hepatitis B virus for the lowest-frequency mode in each irreducible representation A , T_1 , T_2 , G , and H of the icosahedral group. Lowest-frequency (A) A , (B) T_1 , (C) T_2 , (D) G , and (E) H modes.

frequency A mode does not correspond to a breathing of the capsid, but instead is a more complex motion of the four proteins. A simplified description of the motion is that the five proteins around the fivefold axis expand outward while the six proteins around the threefold axis expand inward. The next three low-frequency A modes of hepatitis B (not shown) correspond to the breathing, pentagonal rotation, (where proteins on the fivefold and threefold symmetry axes “twist” in opposite directions) and the pentagonal pinch where proteins on both the fivefold and threefold axes move toward the main symmetry axis (either fivefold or threefold). Interestingly, the first few (<5) low-frequency modes of hepatitis B in all symmetry groups have similarities to the vibrational modes of C_{240} and reflects how extreme coarse-graining models that replace whole proteins with point masses can still usually predict a few of the low-frequency displacement patterns [39,40]. However, while such coarse graining may be able to predict the center-of-mass motion of the proteins for a few modes, it cannot predict the soft spots of the capsid that are most important for allowing the capsid to (in the case of breathing) expand [18].

A comparison of the spectra of the polio virus and the HBV reveals a surprising result. The polio virus has an overall higher frequency spectrum than the HBV even though the polio virus has nearly 50% more atoms than HBV. By higher frequency we mean that the extreme lowest-frequency values for polio are higher than HBV and that the density of levels (levels per cm^{-1}) is greater for HBV than polio. [Note that Fig. 4 (polio) is for 50 modes per symmetry, while Fig. 7 (HBV) is for 100 modes.] The result is unexpected because larger objects tend to have lower frequencies. However, the outer radii of the two viruses are comparable; the outer radius of polio is approximately 30–31 nm, while the outer radius of HBV is about 30 nm neglecting the spikes and is about 32.5 nm including the spikes. Many models of capsids use continuum elasticity theory to determine the modes of the capsid, and we believe that a comparison of these two capsids within an elastic model would be enlightening. It is unclear if the trends observed here would persist in a continuum elastic model.

A qualitative relative Raman intensity spectrum is determined using a bond polarizability model as was done for

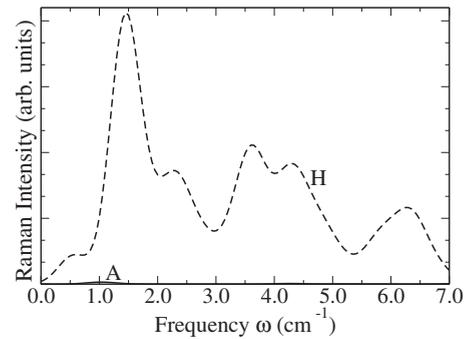


FIG. 9. The Raman intensity profile of the hepatitis B virus up to 7 cm^{-1} . Raman active modes are of A (solid line) and H (dashed line) symmetries.

polio. The spectrum is shown in Fig. 9. The spectrum is completely dominated by the H -symmetric modes with the A modes producing very little intensity. The intensity spectrum is heavily weighted toward very low frequency with additional intensity more evenly spread out.

Having a low-frequency spectrum with all degrees of freedom and a reasonably accurate force field allows one to get a sense of the robustness of more approximate methods. In this spirit we have computed the vibrational spectrum of HBV using an elastic network model (ENM). In the ENM, one treats only the C_α atoms and connects them together with springs. The springs connect C_α atoms within a cutoff distance. Here, we choose the cutoff to be 8 \AA and the spring constant k is the same independent of distance within the cutoff. The potential energy between nodes i and j is $U_{ij} = (k/2)(d_{ij} - d_{ij}^0)^2$, where d_{ij} is the distance between atoms i and j , and d_{ij}^0 is the equilibrium distance. The mass of each node is the mass of carbon. We choose the spring constant k so as to reproduce the frequency of the lowest-frequency A mode (0.86 cm^{-1}). Even with the ENM approximation the dynamical matrix is $101\,880 \times 101\,880$ ($566 C_\alpha$ atoms per site $\times 60$ sites $\times 3$ degrees of freedom), too large to directly diagonalize. A typical solution to this problem is to treat each of the virus’ proteins as a rigid mass [15,39], which would reduce the dynamical matrix for a $T=4$ capsid to 1440×1440 . We use the phonon functional method developed here (with symmetry) to determine the ENM mode frequencies and eigenvectors for the full $101\,880 \times 101\,880$ matrix, much like we do for the all-atom (AA) model. This shows the utility of the phonon functional method in that it can find the lowest-frequency modes of very large (and possibly ill-conditioned) matrices.

We compare the frequencies and displacement eigenvectors of the AA method with the ENM method for the A -symmetric modes. The eigenvectors of the AA calculation include far more atoms than the C_α atoms. We determine the all-atom displacement vectors $|\eta_v\rangle$ then remove all components other than those of C_α , and finally renormalize the vector to unity. Figure 10 shows a spectrum of the lowest 20 vibrational frequency modes for both the AA method and the ENM. Although the lowest ENM mode frequency was fit to the AA result, the remaining mode frequencies are in relatively good agreement with the AA results.

The correspondence of the first few ENM eigenvectors compared to those of the AA model is remarkable. We com-

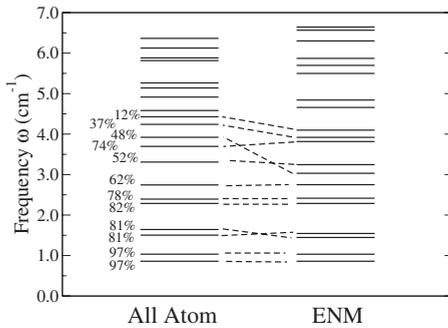


FIG. 10. A comparison of the frequency spectrum of the HBV capsid for the AA model and for an elastic network model (ENM). Only the lowest 20 modes are shown. The frequency of the lowest mode of ENM model has been adjusted to produce the AA result. The percentages refer to the largest projection of an ENM state onto the AA state, and the dashed lines show the mapping of the ENM state onto the AA state with largest projection.

pute the overlap of an ENM eigenvector $|\eta_i^{ENM}\rangle$ with an all-atom eigenvector $|\eta_j^{AA}\rangle$ to give a projection probability, $P_{ij} = |\langle \eta_i^{ENM} | \eta_j^{AA} \rangle|^2$. By completeness, this overlap is between zero and 1. We multiply by 100 to make it a percentage. Figure 10 shows that the eigenvectors of the ENM for the lowest few modes (~ 5) faithfully reproduce the displacement pattern of the all-atom results. For example the lowest ENM mode projects 97% onto the lowest all-atom mode. Higher modes become less faithful (e.g., the seventh ENM mode projects 62% onto the seventh all-atom mode). Considering that the ENM mode takes into account no specific binding or interactions, the close correspondence is surprising. Even the loss of fidelity at higher-frequency mode levels is not as sharp as Fig. 10 indicates. For example, ENM mode 12 shows a 12% projection probability onto mode 12 of the all-atom model. However, the sum of the projection onto the first 15 all-atom modes is 70%—thus, the ENM displacement pattern is still a faithful representation of a low-frequency distortion, but it is not a true eigenmode.

VI. LOW-FREQUENCY MODES OF CCMV

The CCMV is a $T=3$ nonenveloped icosahedral virus from the bromoviridae family that infects plants such as the black eyed pea plant. Infection occurs through a vector, such as an aphid or other small insect that carries the virus from infected plants to uninfected ones when the insect feeds on leaves or other tissues of the plant. Probably one of the most interesting features of the CCMV capsid is its ability to undergo a reversible swelling, which opens and closes 60 pores on the capsid surface. In the normal CCMV virus, swelling is prevented by placing 180 calcium ions at key electronegative areas of the capsid. However, *in vitro* calcium ions are typically removed and the swelling can then be induced by a simple change in the pH of the buffer solution. This reversible swelling has allowed for nanoparticles to be encapsulated by the CCMV protein capsid. This raises the possibility of using viruses such as CCMV as targeted drug delivery systems, where the virus capsid can be used to encapsulate a toxic drug and deliver it to a very specific target (e.g., can-

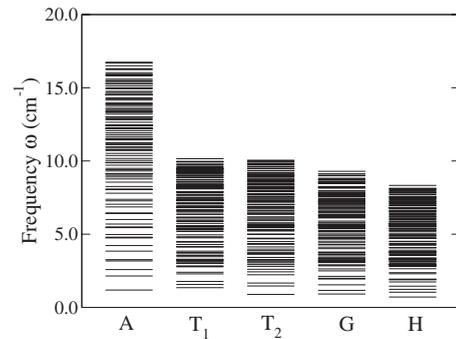


FIG. 11. Stick spectrum of the lowest 100 modes of the cowpea chlorotic mottle virus (CCMV) for each irreducible representation of the symmetry group I .

cerous cell) thereby minimizing the damage and side effects that would occur from typical administration of drug treatments. Understanding the swelling phenomenon and its cause shows how atomistic normal-mode analysis of viral capsids can impact these emerging fields.

The basic building block of the virus during assembly appears to be a dimer of coat proteins, which first assemble into a pentamer of dimers [41]. Experimental studies have shown that *in vivo* the CCMV virus assembles into exclusively $T=3$ particles formed from 90 dimers; however, *in vitro* the virus can be manipulated to assemble into $T=1$ or even $T=2$ particles which violate the Casper-Klug model [42]. The formation of $T=1$ and $T=2$ particles can be controlled by deleting the N -terminal region of the capsid protein ($N\Delta 34$), which forms a β annulus region from the six capsid proteins that meet at the threefold axes of the virus [42].

Coordinates for the CCMV virus were obtained from the protein data bank (PDB code 1CWP). In addition to the three capsid proteins, the crystal structure also contained a small segment of ssRNA which was noncovalently bonded to the capsid. We added three calcium ions with charge +2 to the three capsid proteins at the calcium binding sites. In total, a single icosahedral site of the virus (1 of 60) contained a total of 7520 atoms (three capsid proteins, ssRNA segment, and three calcium ions) for a total of $N=451\,200$ atoms in the complete virus capsid. The structure was minimized using icosahedral symmetry to a RMS force value of 6.6×10^{-5} eV/Å. The resulting RMS deviation of the atoms from x-ray structure was 3.7 Å.

The total number of mechanical modes in the CCMV virus is $3 \times N = 1.3M$. We solve for the lowest $M_p = 100$ frequency modes of the CCMV virus for each irrep. A stick spectrum of the lowest-frequency modes for each irrep. is shown in Fig. 11. As with polio and hepatitis B, CCMV has the symmetric A modes distributed over a larger frequency spectrum (1.19–17 cm^{-1}) compared with the T_1 , T_2 , G , and H modes.

Table V gives the frequencies of the first 20 modes of CCMV for each irreducible representation and indicates localization (or lack thereof) by the participation number W . Full participation of all atoms in a mode (such as a translation) is 7520, corresponding to the number of atoms in a fundamental unit of the icosahedron. The first three modes of

TABLE V. Frequencies (ω_i in cm^{-1}) and participation numbers (W) per protein unit of CCMV virus for each of the irreducible representations (group I). Full participation is 7520 atoms. Raman active modes are of A and H symmetries.

A		T_1		T_2		G		H	
ω	W								
1.19	7168	1.35	7169	0.89	6950	0.91	6706	0.71	7381
2.14	5585	1.56	6628	1.46	6187	1.16	6792	1.04	7364
2.58	5152	1.77	6350	1.66	6955	1.54	6591	1.24	7183
3.17	816	2.29	6356	2.23	5654	1.94	5059	1.45	6672
3.26	903	2.40	6775	2.43	5014	1.99	6137	1.74	6825
3.84	3606	2.77	5278	2.60	4997	2.12	5307	1.88	7005
4.22	3084	2.83	4685	2.80	4825	2.49	5102	1.93	7128
4.75	3142	3.01	3017	2.92	3001	2.63	4979	2.30	6122
4.79	4183	3.04	2782	3.04	3614	2.84	5873	2.37	6778
4.98	3608	3.14	1472	3.12	1641	2.91	5234	2.64	4881
5.44	4118	3.21	2495	3.22	583	3.06	2598	2.80	4684
5.50	421	3.25	1223	3.26	1086	3.11	1660	2.84	4376
5.69	3250	3.29	2068	3.41	956	3.13	1973	2.90	3726
6.00	4002	3.46	2275	3.41	3404	3.16	2584	2.91	3497
6.42	4535	3.54	2522	3.64	3599	3.20	2004	2.94	4103
6.44	4264	3.66	3404	3.67	3659	3.24	1412	3.01	3481
6.86	1817	3.75	4700	3.76	4524	3.31	2496	3.10	2359
7.05	4877	3.84	4465	3.88	3415	3.36	2202	3.16	1902
7.27	2506	4.11	4511	3.98	4710	3.44	1099	3.23	1932
7.38	3207	4.16	4888	4.13	5329	3.58	4654	3.27	671

each symmetry are highly cooperative with $W=5150-7381$. However, the fourth and fifth low-frequency A modes at 3.17 and 3.26 are very localized compared with the other symmetries with $W < 900$. These modes are localized on the ssRNA segments that are noncovalently bound to the capsid proteins.

As with the polio and hepatitis B, we show the displacement patterns of the lowest modes for each symmetry in Fig. 12. The arrows of the displacement pattern represent the center-of-mass motion for one of the three proteins in the icosahedral unit. The lowest mode of A is the breathing

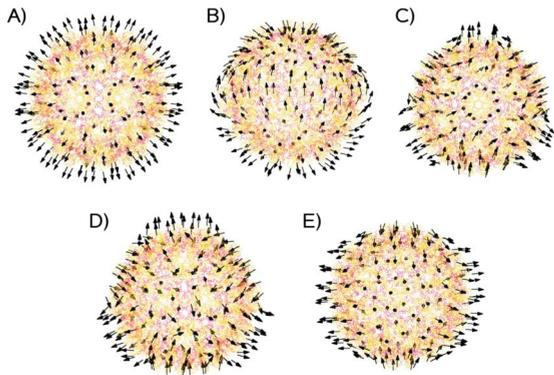


FIG. 12. (Color online) Displacement patterns of the cowpea chlorotic mottle virus for the lowest-frequency mode in each irreducible representation A , T_1 , T_2 , G , and H of the icosahedral group. Lowest-frequency (A) A , (B) T_1 , (C) T_2 , (D) G , and (E) H modes.

mode, while the lowest T_1 mode is a dipolelike motion. These first few lowest modes of CCMV have similarities to C_{180} and again show that simple models of viral capsid mechanical modes can be reproduce at least the very fundamentals. But other low-frequency modes (such as the fourth and fifth lowest A modes) would be missed in such simplified models due to their localization on the ssRNA.

Finally, we show the predicted Raman spectra for CCMV in Fig. 13. The H modes completely dominate the contribution to the low-frequency Raman scattering profile of CCMV. The only contribution from the A modes to the low-frequency scattering spectrum is the mode near 5.0 cm^{-1} .

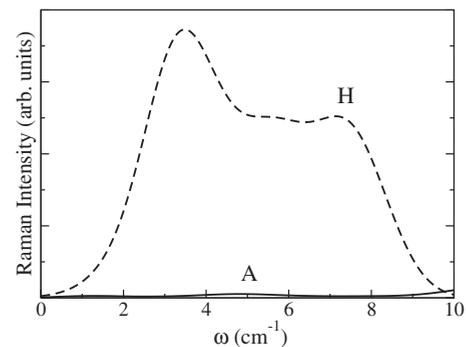


FIG. 13. The Raman intensity profile of the cowpea chlorotic mottle virus up to 10 cm^{-1} . Raman active modes are of A (solid line) and H (dashed line) symmetries.

TABLE VI. Estimates of the damping of three low-frequency modes of the CCMV virus. The damping estimates are based on the continuum elastic theory work of Murray and Saviot [44]. Frequencies ω and periods T are given for the lowest-frequency mode from the $l=0, 1, 2$ angular momentum numbers. The value τ is the approximate lifetime of the mode, i.e., the time at which the amplitude is decreased by \exp^{-1} .

	$l=0$ (A modes)	$l=1$ (T_1 modes)	$l=2$ (H modes)
ω^a	1.19 cm^{-1}	1.35 cm^{-1}	0.71 cm^{-1}
T^a	28.0 ps	24.7 ps	46.9 ps
ω^b	1.28 cm^{-1}	0.84 cm^{-1}	0.56 cm^{-1}
T^b	25.9 ps	39.6 ps	56.8 ps
τ^b	6.87 ps	27.3 ps	65.5 ps

^aThis work.

^bFrom Murray and Saviot [44].

VII. SOLVENT DAMPENING

The question naturally comes up concerning the dampening of the vibrational modes. Specifically are they overdamped, slightly underdamped, or highly underdamped? The phonon functional fails to answer this question since it has no explicit coupling with the solvent.

Experimentally there are some hints. Recently, the very-low-frequency vibrational modes of a virus capsid in solution have been measured by Raman scattering for a tubular virus, M13 bacteriophage. Tsen *et al.* [33] observed a clear peak at low frequency ($\sim 7 \text{ cm}^{-1}$). The intensity of the peaks scale with the viral concentration providing strong evidence that, at least in this case, dampening does not destroy mechanical modes. However, not all attempts have been successful. Stephanidis *et al.* [43] performed Brillouin scattering on a spherical virus, the satellite tobacco mosaic viruses, and did not detect virus particle mechanical modes. Although there has not been an extensive experimental search for Raman spectra of viruses, the two reported experiments (on two different viruses) yield quite different results. This perhaps suggests that some systems are overdamped while others have at least some underdamped modes.

The two experiments described above differ in that one experiment was on a tubular virus while the other was on a spherical virus. To determine the damping characteristics specifically of icosahedral viruses, we rely on estimates taken from continuum elasticity theory from the work of Murray and Saviot [44]. In their model, icosahedral viruses are modeled as spheres of radius R in a viscous fluid (water). The complex frequencies are determined as a function of radius. From this work we estimate the lifetime τ of the lowest modes for CCMV, which we take to be approximately 30 nm in diameter or 15 nm in radius. Table VI lists estimates for the damping of the lowest-frequency modes in the A , T_1 , and H irreps. It is clear that while modes such as the A irrep. breathing modes are highly damped, other modes such as the lowest-frequency H mode of CCMV are slightly underdamped. Generalizing these results to other low-

frequency modes besides just the lowest gives Q values from about 0.5 to 5, with H symmetry being less damped than A symmetry. This is consistent with the work of Talati and Jha [45] who found for a 100 nm diameter spherical capsid a value of $\omega\tau \approx 5$ and $Q \approx 1$. Thus, overall it appears that viral modes of spherical capsids are near the overdamped to slightly underdamped border with a trend that the highest symmetry (e.g., A symmetric) modes lie on the overdamped side, while less symmetric (e.g., H symmetric) modes are on the slightly underdamped side.

An appealing means to incorporate dampening into the phonon functional, at least approximately, is by means of Langevin modes introduced by Lamm and Szabo [46]. This model includes phenomenological friction with the solvent through a Stokes viscous drag interaction of solvent accessible atoms on the virus capsid. The method produces a non-Hermitian Langevin matrix that is $6N$ square instead of $3N$ square (N is the number of atoms). The eigenvalues have an imaginary part describing dampening. This method has yet to be implemented in the phonon functional method and represents a future and challenging problem.

VIII. CONCLUSIONS

We have presented a method for determining the low-frequency modes of large symmetric biological molecules (such as viruses) with atomic detail using the phonon functional method with a symmetry reduction in the dynamical matrix. As an illustration of the technique, we determined the low-frequency mechanical modes of three icosahedral viruses to atomic detail: the polio virus, the hepatitis B virus, and the cowpea chlorotic mottle virus. Additionally, the full atomic displacement pattern of the modes has allowed for the prediction of Raman spectra profiles of the viral capsids within a bond polarizability framework.

The ability to calculate the mechanical modes of viral capsids to atomic detail and by symmetry irreducible representation has allowed for comparisons of coarse-graining models of viral capsids, such as the elastic network models, with the fully atomic methods presented in this work. Comparison of the displacements patterns and frequencies show that coarse-graining models can produce good estimates of the displacement patterns and relative frequencies of at least the first few global modes (~ 5) that involve collective motions of protein subunits. The key approximation in the elastic network model is the building of a collective network of springs between C_α atoms. More (or fewer) connections between C_α atoms are how rigid (floppy) regions of the protein are described in the model. Thus, we expect that the elastic network model will do well in describing the first few modes when the strength of the atomic forces can be represented to a good degree by the number of connective springs. Thus, regions of the protein, which are weakly interacting but very compact (such as a weakly bound ligand), will produce a stiff area in the elastic network model with a corresponding high frequency, while the fully atomic method predicts a lower-frequency localized mode.

- [1] M. Babincova, P. Sorivong, and P. Babinec, *Med. Hypotheses* **55**, 450 (2000).
- [2] R. Cerf, B. Michels, J. A. Schulz, J. Witz, P. Pfeiffer, and L. Hirth, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1780 (1979).
- [3] K.-T. Tsen, S.-W. Tsen, T.-C. Wu, C. L. Chang, C.-F. Hung, and J. Kiang, *Virology Journal* **4**, 50 (2007).
- [4] K.-T. Tsen, S.-W. Tsen, O. F. Sankey, and J. Kiang, *J. Phys.: Condens. Matter* **19**, 472201 (2007).
- [5] K.-T. Tsen, S.-W. Tsen, C.-L. Chang, C.-F. Hung, T.-C. Wu, and J. G. Kiang, *J. Phys.: Condens. Matter* **19**, 322102 (2007).
- [6] K.-T. Tsen, S.-W. Tsen, C.-L. Chang, C.-F. Hung, T.-C. Wu, and J. G. Kiang, *J. Biomed. Opt.* **12**, 064030 (2007).
- [7] K.-T. Tsen, S.-W. D. Tsen, C.-F. Hung, T.-C. Wu, and J. G. Kiang, *J. Phys.: Condens. Matter* **20**, 252205 (2008).
- [8] B. Brooks and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4995 (1985).
- [9] W. Zheng, B. R. Brooks, and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7664 (2006).
- [10] P. Petrone and V. S. Pande, *Biophys. J.* **90**, 1583 (2006).
- [11] A. W. Van Wynsberghe and Q. Cui, *Structure (London)* **14**, 1647 (2006).
- [12] E. C. Dykeman, O. F. Sankey, and K.-T. Tsen, *Phys. Rev. E* **76**, 011906 (2007).
- [13] M. M. Tirion, *Phys. Rev. Lett.* **77**, 1905 (1996).
- [14] I. Jeong, Y. Jang, and M. K. Kim, *J. Mol. Graphics Modell.* **24**, 296 (2006).
- [15] F. Tama, F. X. Gadea, O. Marques, and Y. H. Sanejouand, *Proteins: Struct., Funct., Genet.* **41**, 1 (2000).
- [16] Z. Yang, I. Bahar, and M. Widom, *Biophys. J.* **96**, 4438 (2009).
- [17] P. Ordejon, D. A. Drabold, R. M. Martin, and S. Itoh, *Phys. Rev. Lett.* **75**, 1324 (1995).
- [18] E. C. Dykeman and O. F. Sankey, *Phys. Rev. Lett.* **100**, 028101 (2008).
- [19] H. W. T. van Vlijmen and M. Karplus, *J. Mol. Biol.* **350**, 528 (2005).
- [20] S. Go, H. Bilz, and M. Cardona, *Phys. Rev. Lett.* **34**, 580 (1975).
- [21] D. W. Snoke and M. Cardona, *Solid State Commun.* **87**, 121 (1993).
- [22] S. Guha, J. Menendez, J. B. Page, and G. B. Adams, *Phys. Rev. B* **53**, 13106 (1996).
- [23] D. Bermejo, S. Montero, M. Cardona, and A. Muramatsu, *Solid State Commun.* **42**, 153 (1982).
- [24] P. D. Ross, J. F. Conway, N. Cheng, L. Dierkes, B. A. Firek, R. W. Hendrix, A. C. Steven, and R. L. Duda, *J. Mol. Biol.* **364**, 512 (2006).
- [25] R. A. Grant, C. N. Hiremath, D. J. Filman, R. Syed, K. Andries, and J. M. Hogle, *Curr. Biol.* **4**, 784 (1994).
- [26] The protein data bank, <http://www.rcsb.org/>
- [27] M. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- [28] E. Neria, S. Fischer, and M. Karplus, *J. Chem. Phys.* **105**, 1902 (1996).
- [29] 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).
- [30] D. Bashford and D. A. Case, *Annu. Rev. Phys. Chem.* **51**, 129 (2000).
- [31] V. Tsui and D. A. Case, *Biopolymers* **56**, 275 (2001).
- [32] E. C. Dykeman and O. F. Sankey, *J. Phys.: Condens. Matter* **21**, 035116 (2009).
- [33] K.-T. Tsen, E. C. Dykeman, O. F. Sankey, S.-W. D. Tsen, N.-T. Lin, and J. G. Kiang, *Nanotechnology* **17**, 5474 (2006).
- [34] T. L. Wright, *Am. J. Gastroenterol.* **101**, S1 (2006).
- [35] R. A. Crowther, N. A. Kiselev, B. Bttcher, J. A. Berriman, G. P. Borisova, V. Ose, and P. Pumpens, *Cell* **77**, 943 (1994).
- [36] A. Zlotnick, N. Cheng, J. F. Conway, F. P. Booy, A. C. Steven, and S. J. Stahl, *Biochemistry* **35**, 7412 (1996).
- [37] B. Bttcher, S. A. Wynne, and R. A. Crowther, *Nature (London)* **386**, 88 (1997).
- [38] S. A. Wynne, R. A. Crowther, and A. G. Leslie, *Mol. Cell* **3**, 771 (1999).
- [39] F. Tama and C. L. Brooks III, *J. Mol. Biol.* **345**, 299 (2005).
- [40] F. Tama and C. L. Brooks III, *J. Mol. Biol.* **318**, 733 (2002).
- [41] J. M. Johnson, J. Tang, Y. Nyame, D. Willits, M. J. Young, and A. Zlotnick, *Nano Lett.* **5**, 765 (2005).
- [42] J. Tang, J. M. Johnson, K. A. Dryden, M. J. Yong, A. Zlotnick, and J. E. Johnson, *J. Struct. Biol.* **154**, 59 (2006).
- [43] B. Stephanidis, S. Adichtchev, P. Gouet, A. McPherson, and A. Mermet, *Biophys. J.* **93**, 1354 (2007).
- [44] D. B. Murray and L. Saviot, *J. Phys.: Conf. Ser.* **92**, 012036 (2007).
- [45] M. Talati and P. K. Jha, *Phys. Rev. E* **73**, 011901 (2006).
- [46] G. Lamm and A. Szabo, *J. Chem. Phys.* **85**, 7334 (1986).