Low Frequency Mechanical Modes of Viral Capsids: An Atomistic Approach

Eric C. Dykeman and Otto F. Sankey

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

(Received 24 August 2007; published 14 January 2008)

We present a method for the calculation of the low frequency vibrational modes and frequencies of viral capsids, or other large molecules, where the modes are modeled with atomic detail. Extending ideas from electronic structure theory, an energy functional is used to find modes of a classical dynamical matrix below a fixed (pseudo-Fermi) level. The icosahedral satellite tobacco necrosis virus is modeled as an example. We find that atoms around the $C5$ and $C3$ axis have small relative displacement while the β sheet body shows gliding motion.

DOI: 10.1103/PhysRevLett.100.028101 PACS numbers: 87.15.La, 87.64.kp

Viruses are a major threat to both public health and to agricultural production. They are responsible for many diseases that inflict harm or death in the human population and also infect plants and animals damaging cultivated plants and livestock. Viruses protect themselves with a protein coat (capsid) that has a well-defined structure. They are the smallest living creature and the capsid encapsulates the genetic material inside. Chemotherapies (drugs) may interfere with the virus life-cycle after infection, but the host may suffer harmful side effects. Rapid mutations often quickly render a drug ineffective.

Attacking the virus through mechanical means to produce damage is an attractive proposal which is just beginning to be considered. It has been suggested [1] that the AIDS virus can be attacked with 10 GHz (0.3 cm^{-1}) high frequency ultrasound or hypersound excitation. Another means [2] has recently been proposed using impulsive stimulated Raman scattering (ISRS) with visible light [3]. Novel potential therapies based on mechanical resonance can best be advanced if a detailed picture of the viral capsid vibrational patterns and frequencies are understood. We provide an atomic level description of mechanical vibrations of a model virus, the satellite tobacco necrosis virus (STNV), and identify Raman active modes that can be excited by ISRS excitation.

A viral capsid is a large molecule consisting of one or more protein units that are arranged in either a helical (tubular) or icosahedral (spherical) fashion, with many viral capsids being icosahedral. The total number of atoms can be enormous (10^5–10^6 atoms), even for small capsids. This illuminates the central problem of a fully atomistic normal mode analysis (NMA)—the memory required to store the force constant matrix and the time required to diagonalize it quickly make it unfeasible.

Methods used to study the normal modes of viral capsids include continuum elastic theory (CET) [4], elastic network models (ENMs) [5], and the rotation translation block method (RTB) [6]. These models vary in their description of individual atomic motion ranging from none (CET) to a coarse grained description (ENM and RTB model). Prior NMA of viral capsids have used coarse graining with group theory for the icosahedron [7], or have only considered rotations and translations of each protein unit (treated as a ridged block) [6,8].

We solve for the normal modes using a full basis set (i.e., the eigenvectors are constructed from the full Cartesian space of 3 degrees of freedom for each atom and thus will give the correct individual atomic displacements) and use an iterative method where only the operation of the dynamical matrix ($D = M^{-1/2}FM^{-1/2}$) on a vector is required. Here F is the force constant matrix and M is the diagonal mass matrix. The F matrix is sparse [9] and its operation on a vector can be calculated in Order (N) steps where N is the number of atoms in the molecule.

Our basic strategy is to use techniques from electronic structure (ES) theory which require states below a Fermi level. In ES problems, a few (perhaps a hundred) occupied states are required, but there are many more (perhaps many thousands) of unoccupied levels that do not affect the total energy. The ''band-structure'' energy G (neglecting spin) is given by $G = \sum_i \lambda_i$, where the sum is over occupied (lowest lying) electron orbital energies satisfying $\hat{H}\psi_i = \lambda_i\psi_i$. An expression for G is

$$G = \mathrm{Tr}[\hat{H} f(\hat{H}, \mu)] \tag{1}$$

where f is the Fermi function, $f(\hat{H}, \mu) = 1/(e^{\beta(H-\mu)} + 1)$, and the trace is over any complete basis and is invariant to the orthonormal basis used. At low temperature $f(\lambda, \mu)$ is 1 (occupied) or 0 (unoccupied) depending on λ. This form for G is not often used in practice because the Fermi function acting on a general state ψ is difficult to compute. Instead an energy functional is introduced and is minimized with respect to the occupied basis states.

The connection of the ES problems with the low frequency viral vibrational problem should now be clear. We consider the M lowest frequency vibrational states to form an ''occupied'' subspace. The problem of finding the M *lowest* frequencies (eigenvalues) and eigenvectors of the dynamical matrix D can be recast as the minimization of an effective band-structure energy G. The minimization pro-

ceeds through a set of M vectors $|u_i\rangle$, which are initially chosen at random and need be neither orthogonal nor even normalized. Several choices for G exist [10,11] and deep connections between forms is evident. We choose perhaps the simplest functional of Ordejon-Drabold-Martin-Grumbach (ODMG)[10],

$$G = \min\{\mathrm{Tr}_M[H + (1 - S)H]\}. \quad (2)$$

The ODMG functional produces the correct global minimum and the desired orthogonality for the M lowest frequency states. The "Hamiltonian" H and overlap S are $M \times M$ matrices with elements given by $H_{ij} = \langle u_i|\hat{D}|u_j\rangle$ and $S_{ij} = \langle u_i|u_j\rangle$. After minimization, the vectors $|u_i\rangle$ are not eigenvectors but are a complete orthonormal set spanning the space of the lowest states of $\hat{D}$. The true eigenstates are easily obtained by diagonalizing the *small* $M \times M$ matrix equation, $HC(\lambda) = \lambda SC(\lambda)$. The eigenvector $|e(\lambda)\rangle$ for eigenvalue λ is $|e(\lambda)\rangle = \sum_i C_i(\lambda)|u_i\rangle$.

To guarantee that the ODMG functional, [Eq. (2)], has the correct global minimum, the spectrum of $\hat{D}$ must be entirely negative. Thus we redefine H as $H_{ij} = \langle u_i|\hat{D}_s|u_j\rangle$, where $\hat{D}_s = \hat{D} - \lambda_L\hat{I}$ is the shifted dynamical matrix and λ_L corresponds to the largest eigenvalue of $\hat{D}$. The largest eigenvalue is calculated with a few ($\approx$10) Lanczos recursions with $\hat{D}$ operating on a random vector.

The energy functional Eq. (2) is minimized using the conjugate gradient (CG) method of Pollak and Riberi [12]. The necessary gradient at step n,

$$|g_k^n\rangle = 4\hat{D}_s|u_k^n\rangle - 2\sum_i S_{ki}^n\hat{D}_s|u_i^n\rangle - 2\sum_i H_{ki}^n|u_i^n\rangle, \quad (3)$$

is used to form the CG search direction $|p_k^n\rangle$. By operating $\hat{D}_s$ on the CG search directions, the vectors $|u_i\rangle$ and $\hat{D}_s|u_i\rangle$ can be iteratively updated,

$$\begin{aligned} |u_i^{n+1}\rangle &= |u_i^n\rangle + \delta^n|p_i^n\rangle \\ \hat{D}_s|u_i^{n+1}\rangle &= \hat{D}_s|u_i^n\rangle + \delta^n\hat{D}_s|p_i^n\rangle. \end{aligned} \quad (4)$$

The line minimization is performed *analytically* by requiring the new gradient to be perpendicular to the old search direction, which yields a cubic equation for δ^n. This CG procedure requires M dynamical matrix operations per step and a storage of $5M$ vectors. The density of states is small at low frequencies and we find that for molecules with $N \approx 10^4$, $M \approx 100$ is usually sufficient to find all eigenvectors with $\omega < 30$ cm^{-1} and a storage requirement of roughly 114 MB (double precision).

We now discuss results for the NMA of STNV. Atomic coordinates were obtained from the protein data bank (2BUK), which were determined from x-ray crystallography [13]. STNV is a naked capsid (i.e., lacking a lipid membrane envelope) icosahedral T1 virus having 60 copies of each unit. A unit contains a single protein, 158 water molecules and 3 Ca^{2+} (Fig. 1) totaling 3324 atoms. Group theory for the icosahedron was used to reduce the full problem to five smaller problems related to the irreducible representations (irrep.) of the icosahedron (A, $T1$, $T2$, G, and H) where the matrices have sizes of 9972, 29916, 29916, 39888, and 49860, respectively.

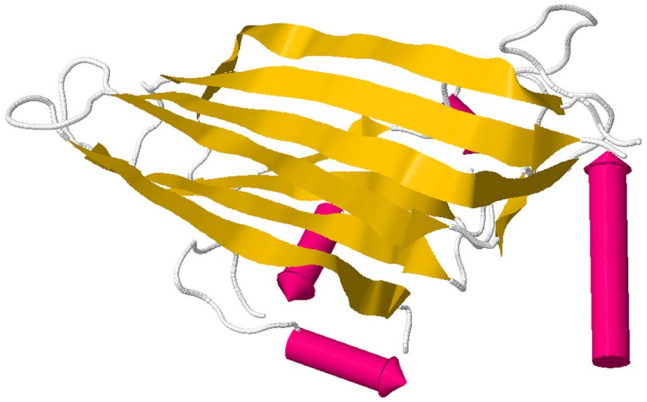

FIG. 1 (color online). Structure of a single protein building block of the satellite tobacco necrosis virus (2BUK). Not shown are the 3 Ca^{+2} atoms that bind the protein with itself, H_2O, and symmetry related copies.

The AMBER 6 [14] classical force field with generalized Born (GB) Coulomb interactions [15,16] was used to energy minimize STNV and calculate the dynamical matrix operating on a vector. The GB interactions included neighbors within 10 Å. The GB and van der Waals parameters for the Ca^{2+} ions were obtained from Babu *et al.* [17]. Although interactions with genetic material and explicit water were ignored, the methodology is adaptable to incorporate these interactions. The methodology is also capable of using other potential energy models.

A single unit of 2BUK was energetically minimized while enforcing group theory, stopping when the rms force was $<$0.001 eV/Å. The root mean square deviation between the optimized structure and the x-ray structure was 0.7 Å. This minimization procedure is consistent with other work [18]. Since the rms force was not exactly zero, negative eigenvalues ($\lambda = \omega^2 < 0$) can and do appear. Analysis of these negative vibrational modes show that they are localized on a few atoms, usually a "dangling" side chain or a single water molecule, and their frequency is small in magnitude ($<$13 cm^{-1}). We ignore them in what follows as none appear to be physically relevant.

Equation (2) was CG minimized using coordinates from the optimized structure for 100 vectors (M) in each irrep. Minimization continued until all residual vectors ($|r_i\rangle = \hat{D}|e_i^*\rangle - \lambda_i^*|e_i^*\rangle$) had magnitude less than 10^{-3}. The vectors $|e_i^*\rangle$ are the approximate eigenvectors of $\hat{D}$ and $\lambda_i^* = \langle e_i^*|\hat{D}|e_i^*\rangle$ are the approximate eigenvalues. All calculations were performed on a single Intel Xenon desktop processor, with 2 GB system memory. For each irrep., the number of CG steps was of order 10^3 which for $M = 100$ required approximately 10^5 dynamical matrix opera-

TABLE I. Frequencies (ω_i in cm^{-1}) and participation numbers (W) per protein unit of STNV for each of the irreducible representations (group I). Full participation is 3324 atoms. The degeneracies of each irrep. are shown, and A and H are Raman active.

Symmetry Degeneracy	ω_1 (W_1)	ω_2 (W_2)	ω_3 (W_3)	ω_4 (W_4)	ω_5 (W_5)
A (Raman)	2.4	4.7	5.3	6.1	7.1
1	(2610)	(142)	(1607)	(1752)	(104)
$T1$	2.8	3.5	3.6	4.7	4.8
3	(2689)	(457)	(603)	(833)	(562)
$T2$	2.0	3.42	3.43	3.8	4.3
3	(2740)	(1714)	(2459)	(497)	(616)
G	2.3	2.6	3.5	3.9	4.1
4	(2605)	(2523)	(2801)	(848)	(412)
H (Raman)	1.9	2.3	2.8	3.5	3.8
5	(2856)	(2951)	(2065)	(1210)	(543)

tions. This is the rough equivalent of 10^5 molecular dynamics steps on a single protein unit of STNV.

Table I lists the lowest five frequencies of vibration of STNV for each symmetry. The lowest frequency mode is 1.9 cm^{-1} and of the H irrep. We expect that the low frequency modes are acoustic-like with large regions swaying in concert. To quantify this, we compute the participation number $W_\lambda = e^{S_\lambda}$ for each mode. Here, S_λ is the (information) entropy of the mode given by $S_\lambda = -\sum p_i(\lambda) \ln(p_i(\lambda))$. The probabilities $p_i(\lambda)$ are the squared component of the normalized relative displacements for each atom/direction i, $p_i(\lambda) = |\eta_i(\lambda)|^2$, where $\eta_i(\lambda) \propto M^{-1/2} e_i(\lambda)$. An eigenvector equally distributed over each atom (such as a translation) produces a participation number equal to the number of atoms in one unit of the virus, or 3324 for STNV. Most states have half or more of the total atoms participating. Low participation is a signal that the mode is not global. For example the A state at 4.2 cm^{-1} with 142 atoms participating is found to be localized on the tail of the peptide string that is directed inward toward the center of the virus.

The only Raman active modes are A and H and only these will be sensitive to ISRS stimulation. The first A mode located at 2.4 cm^{-1} corresponds to a expansion or contraction (breathing) of the virus. Figure 2 shows the displacement pattern of the breathing mode for several views. The next important A mode is at 6.1 cm^{-1} and corresponds to a "puckering" about the $C5$ axis. Upon close examination of these two modes (2.4 and 6.1 cm^{-1}) we conclude that they are not uniform rotational or translational motion of rigid protein blocks. Instead they are closer to a gliding motion of the beta sheets over each other. The strongest Raman active H mode located at 4.1 cm^{-1} is shown in Fig. 3. Because of the complexity of the H modes, we illustrate only the center of mass motion of each peptide.

A NMA of STNV has not been previously studied, so a direct comparison of our results with other methods cannot be made. The major difference between our method and the ENM and RTB methods is in the size of the basis set used. In a study of Dialanine, van Vlijmen, and Karplus [7] demonstrate the effect of using a reduced basis set on the normal modes and frequencies. Their results show that the overlaps of eigenvectors calculated with a reduced basis and the full basis are around 60%.

Finally we calculate the combined Raman intensity of the viral A and H modes using a bond polarizability model (BP) [19]. The BP gives a semiquantitative picture of the relative Raman intensities by capturing the important sym-

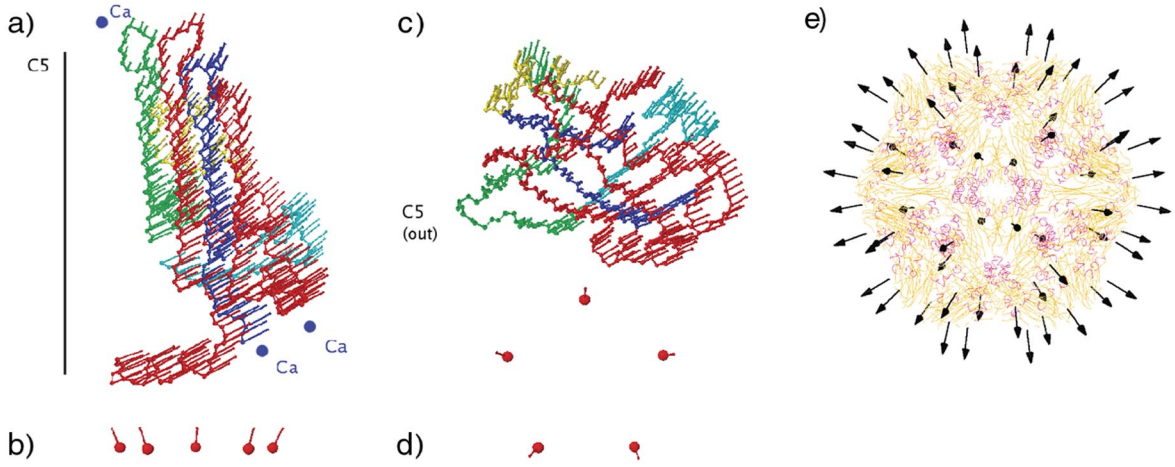

FIG. 2 (color online). Displacement pattern of the breathing A mode at 2.4 cm^{-1} of STNV. Displacements are shown using arrows. Each beta sheet is shaded (colored online) differently for clarity. (a) The backbone displacements within a single protein unit with the $C5$ axis vertically aligned. (b) The displacement of the center of mass of the entire protein unit for the five units around the $C5$ axis. (c) Similar to (a) but looking down the $C5$ axis. (d) Similar to (b) but looking down the $C5$ axis. (e) Displacement of the center of mass of each peptide in the full virus.

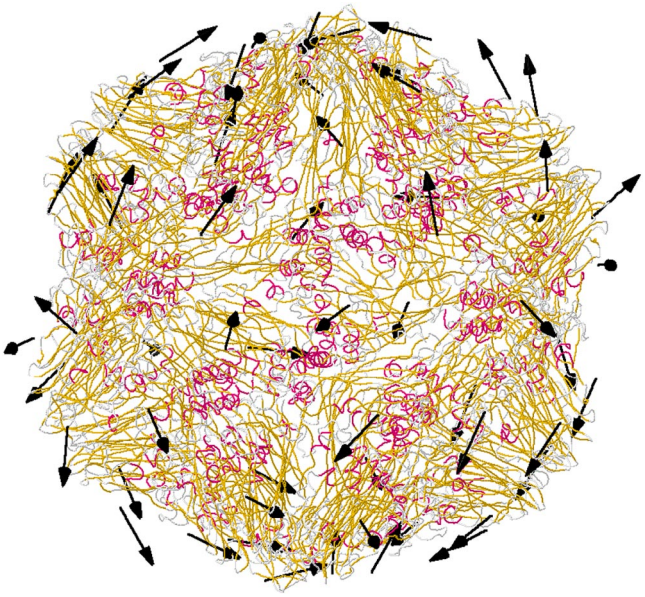

FIG. 3 (color online). Low frequency irrep. H mode of STNV located at 4.1 cm^{-1}. Arrows indicate the center of mass motion of each peptide in the full virus.

metries of the viral vibrations. We model the polarizability of each bond from parameters that will reasonably reflect most bonds, $(\alpha'_{\parallel} - \alpha'_{\perp})/d^2 = 1.2$, $(2\alpha'_{\perp} + \alpha'_{\parallel})/d^2 = 1.7$, and $(\alpha_{\parallel} - \alpha_{\perp})/d^3 = 0.5$. Figure 4 shows the relative Raman intensities for the A modes (solid line) and the H modes (dashed line). The relative Raman intensity of the H modes are much stronger than the A modes, partially due to the fivefold degeneracy. Experiments to resonantly pump vibrations by Raman scattering (ISRS) will likely excite these the strongest.

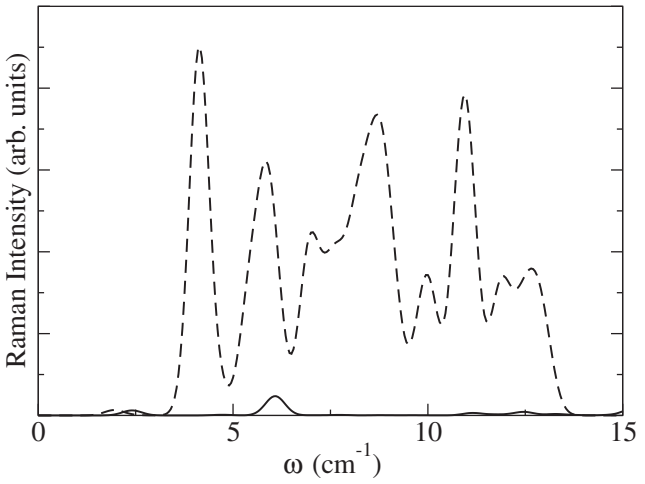

FIG. 4. The relative Raman intensity of A (solid line) and H (dashed line) vibrational modes of STNV calculated from a bond polarizability model.

We have demonstrated the minimization of an electronic structure based energy functional [Eq. (2)] provides a method for determining the low frequency vibrational modes of very large systems, such as viruses, atomistically. The method is iterative and requires the calculation of the dynamical matrix operating on a set of vectors. The resulting analysis of STNV modes shows a detailed, and fairly complex, picture of its low frequency mechanical vibrations.

[1] M. Babincova, P. Sourivong, and P. Babinec, Med. Hypotheses **55**, 450 (2000).
[2] K.-T. Tsen *et al.*, Virology J. **4**, 50 (2007).
[3] Y. X. Yan, E. B. Gambel, Jr., and K. A. Nelson, J. Chem. Phys. **83**, 5391 (1985).
[4] E. C. Dykeman, O. F. Sankey, and K.-T. Tsen, Phys. Rev. E **76**, 011906 (2007).
[5] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[6] F. Tama, F. X. Gadea, O. Marques, and Y. H. Sanejouand, Proteins: Struct., Funct., Genet. **41**, 1 (2000).
[7] H. W. T. van Vlijmen and M. Karplus, J. Chem. Phys. **115**, 691 (2001).
[8] F. Tama and C. L. Brooks III, J. Mol. Biol. **318**, 733 (2002).
[9] The range of the dynamical matrix depends on the force field used. The generalized Born model used has dielectric screening allowing a truncated range of interaction.
[10] P. Ordejon, D. A. Drabold, R. M. Martin, and M P. Grumbach, Phys. Rev. B **51**, 1456 (1995).
[11] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993).
[12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes* (Cambridge University Press, Cambridge, 1986), p. 303.
[13] T. A. Jones and L. Liljas, J. Mol. Biol. **177**, 735 (1984).
[14] W. D. Cornell *et al.*, J. Am. Chem. Soc. **117**, 5179 (1995).
[15] V. Tsui and D. A. Case, Biopolymers **56**, 275 (2000).
[16] D. Bashford and D. A. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).
[17] C. S. Babu, T. Dudev, R. Casareno, J. A. Cowen, and C. Lim, J. Am. Chem. Soc. **125**, 9318 (2003).
[18] Q. Cui, G. H. Li, J. P. Ma, and M. Karplus, J. Mol. Biol. **340**, 345 (2004).
[19] S. Go, H. Bilz, and M. Cardona, Phys. Rev. Lett. **34**, 580 (1975); D. W. Snoke and M. Cardona, Solid State Commun. **87**, 121 (1993).