All-atom normal-mode analysis reveals an RNA-induced allostery in a bacteriophage coat protein

Eric C. Dykeman[1,2] and Reidun Twarock[1,3,2]
[1]*Department of Biology, University of York, York YO10 5DD, United Kingdom*
[2]*York Center for Complex Systems Analysis, University of York, York YO10 5DD, United Kingdom*
[3]*Department of Mathematics, University of York, York YO10 5DD, United Kingdom*
(Received 6 November 2009; published 10 March 2010)

Assembly of the $T=3$ bacteriophage MS2 is initiated by the binding of a 19 nucleotide RNA stem loop from within the phage genome to a symmetric coat protein dimer. This binding event effects a folding of the FG loop in one of the protein subunits of the dimer and results in the formation of an asymmetric dimer. Since both the symmetric and asymmetric forms of the dimer are needed for the assembly of the protein container, this allosteric switch plays an important role in the life cycle of the phage. We provide here details of an all-atom normal-mode analysis of this allosteric effect. The results suggest that asymmetric contacts between the *A*-duplex RNA phosphodiester backbone of the stem loop with the EF loop in one coat protein subunit results in an increased dynamic behavior of its FG loop. The four lowest-frequency modes, which encompass motions predominantly on the FG loops, account for over 90% of the increased dynamic behavior due to a localization of the vibrational pattern on a single FG loop. Finally, we show that an analysis of the allosteric effect using an elastic network model fails to predict this localization effect, highlighting the importance of using an all-atom full force field method for this problem.

DOI: 10.1103/PhysRevE.81.031908 PACS number(s): 87.15.A−, 87.15.hp, 46.40.−f, 87.15.kp

I. INTRODUCTION AND BACKGROUND

Bacteriophage MS2 is a small $T=3$ icosahedral virus with a single-stranded (ss) RNA genome of 3569 nucleotides in length [1]. It is one of the most widely studied virus systems because it provides important insights into the roles of viral RNA during the assembly process [2,3]. X-ray crystallography reveals that 180 coat proteins occur as three quasi-equivalent conformers (*A*, *B*, and *C* subunits) in the $T=3$ capsid and are organized in 60 asymmetric A/B dimers [Fig. 1(a), upper] surrounding the fivefold axes and 30 symmetric C/C dimers [Fig. 1(a), lower] located on the twofold axes [4–6]. The main structural differences between the two dimers reside in the 17 amino acids of the FG loop (residues 66–82) which connect the *F* and *G* β strands in each coat protein subunit. In the *A* and *C* protein subunits (colored blue and magenta, respectively, in Fig. 1) the FG loops form an extended β sheet motif, while in the *B* subunit (colored green) the FG loop is in a distinct "folded" conformation. The folded FG loops of the *B* subunits meet at the fivefold axes of the $T=3$ shell, while the FG loops of *A* and *C* subunits interdigitate on the threefold axes [Fig. 1(b)]. The folded conformation of the *B* subunit is crucial to avoid steric clashes at the fivefold axes and is hence a prerequisite for formations of the particle.

Unlike viruses which preassemble their protein container and then subsequently package their genome using an adenosine triphosphatase packaging motor [7], MS2 *coassembles* its genome and coat protein dimers to form the final complete capsid. Moreover, cryoelectron microscopy studies of the wild-type MS2 have shown quasi-icosahedral symmetry in the RNA material residing inside the completely assembled capsids [8,9]. MS2 initiates the coassembly process by binding a small 19 nucleotide stem loop motif [Fig. 1(c)] present in its ss RNA genome to a single symmetric C/C dimer. The stem loop encompasses the start codon for the replicase gene and serves two purposes. First, it acts as a translational repressor (TR) preventing activity of the replicase gene [10]. And, second, the binding of the TR stem loop to a symmetric coat protein dimer biases a single FG loop in the dimer to fold, forming the assembly competent asymmetric A/B dimer with bound TR, which has been shown to greatly enhance assembly *in vitro* [11].

Nuclear-magnetic-resonance studies [11] suggest that the binding of a TR stem loop to a C/C dimer results in one of the FG loops to undergo a conformational switch to the asymmetric A/B dimer. Since the location of the TR stem loop binding site is >12 Å from the FG loops, the conformational switch appears to be due to an allosteric effect. Basnak *et al.* showed that in addition to the sequence specific TR stem loop, multiple other stem loops unrelated to TR can trigger the same effect suggesting that the allosteric effect is nonsequence specific [12]. This implies that multiple stem loops within the MS2 genome likely regulate the self-assembly of the capsid and encapsidation of the genome.

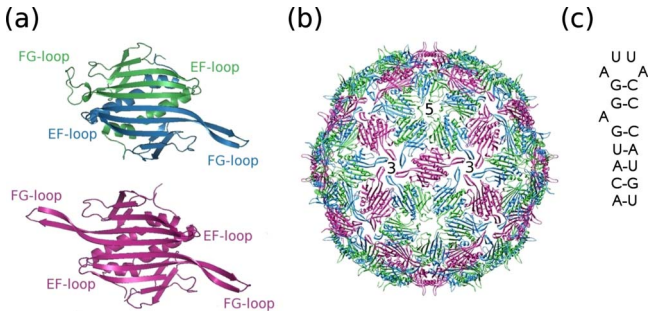

FIG. 1. (Color online) Structure of the MS2 bacteriophage and the TR stem loop. The three quasi-equivalent protein subunits *A*, *B*, and *C* are colored blue, green, and magenta, respectively, with the EF and FG loops labeled in each dimer. (a) The asymmetric A/B dimer (top) and symmetric C/C dimer (bottom). (b) The $T=3$ MS2 capsid (PDB code 1ZDH) viewed from a twofold axis. The folded FG loops of the *B* subunits meet at the fivefold axes, while the extended FG loops of *A* and *C* protein subunits interdigitate at the threefold axes. (c) Structure of the TR stem loop.

However, experimental evidence cannot exclude the possibility that some C/C dimers convert to an A/B dimer upon binding in an autosteric manner, i.e., through a conformational change that is induced through protein binding alone [13].

In earlier work, a nonsequence specific mechanism was proposed [14] where the contact of the phosphodiester backbone of the stem loop with the EF loop of a coat protein subunit was the critical trigger of the conformational switch. Is this work we further explore the physics of this allosteric effect, particularly the localization of the vibrational displacement. We find that the dynamic allostery of the MS2 dimer has similarities to other dynamical allosteries discussed for proteins such as the methionine (MET) repressor [15,16]. The coarse-grained models that have been developed for the dynamic allosteric properties of these proteins could provide additional insights also for the MS2 dimer. In addition, we test if an elastic network model (ENM) can provide similar predictions to the all-atom method used here. The all-atom normal modes in this work were calculated using the phonon functional method [17] which uses an empirical potential-energy model to obtain a few of the *low*-frequency modes of a very large biological structure, such as a virus, to atomic detail. This allows for a comparison of the all-atom normal modes of very large structures with the predictions from an elastic network model.

This paper is organized as follows. In order to render this paper self-contained, we start by presenting a review of the theoretical methods used in this work: (i) the phonon functional method for an all-atom normal-mode analysis of very large biological systems and (ii) the mathematics used to obtain B factors and correlation matrices within a classical harmonic framework. Next we present the computational results of the all-atom normal-mode analysis of MS2 with and without RNA. Although normal-mode analysis uses the harmonic approximation, the results should illustrate how the process of FG-loop refolding would be initiated from within the harmonic well after binding of RNA. The results show a dynamic mechanism for the allosteric effect which stems from a localization of the displacement pattern in the four lowest-frequency modes. In the following section, we perform the same normal-mode analysis using an elastic network model force field, and then compare with the all-atom normal-mode analysis. Finally, we discuss the results and their implications for MS2 assembly.

II. THEORETICAL METHODS

In this section we outline the theoretical methods used to determine the mean-squared fluctuations (B factors) and correlation matrices for the bound and unbound dimers using the low-frequency normal modes of each structure. This is followed by a brief description of the phonon functional method which can be used to compute the normal modes and frequencies of large biological structures to atomic detail.

A. B factor and correlation matrix

We seek to compute the B factor (Debye-Waller factor) and correlation matrices of protein using the lowest-frequency normal modes, which are calculated within a classical harmonic approximation. Both the B factor and correlation matrix depend on the average displacements of the atoms away from their equilibrium positions at temperature T. The B factor for amino acid residue R, denoted as B_R, can be written as [18]

$$B_R = \frac{1}{N_R}\sum_i B_i = \frac{1}{N_R}\sum_i \frac{8\pi^2}{3}\langle \vec{U}_i^2 \rangle, \tag{1}$$

where $\vec{U}_i^2 = u_{ix}^2 + u_{iy}^2 + u_{iz}^2$, and $u_{i\beta}$ is the β component of the displacement of atom i in the amino acid residue from its equilibrium position $\vec{R}_i$ ($u_{i\beta} = r_{i\beta} - R_{i\beta}$). The angular brackets denote an average computed at temperature T. One can see from Eq. (1) that the B factor for atom B_i is proportional to the mean-squared fluctuations of the atom away from its equilibrium position and thus is a useful measure of a protein's flexibility or rigidity. The symmetric correlation matrix (with elements C_{ij}) is computed as a normalized weighting of the average displacement overlap of two atoms at temperature T [19],

$$C_{ij} = \frac{\langle \vec{U}_i \cdot \vec{U}_j \rangle}{\sqrt{\langle U_i^2 \rangle \langle U_j^2 \rangle}}. \tag{2}$$

As a result of the normalization, the elements of the correlation matrix range from -1 to $+1$ and indicate if, on average, two particular atoms i and j move in concert. Thus, a correlation of $C_{ij}=0$ indicates that atoms i and j move perpendicular to each other, while a correlation of $C_{ij} = \pm 1$ indicates that the two atoms move parallel or antiparallel when averaged over many modes. As seen from Eqs. (1) and (2), a calculation of the average of the atoms' squared displacements from equilibrium is required for a computation of the B factors and correlation elements. Specifically two quantities are required: $\langle \vec{U}_i \cdot \vec{U}_j \rangle$ and $\langle U_i^2 \rangle$. Both of these quantities can be written in terms of classical harmonic normal modes, with the dominant contributions ($\sim 80\%$) due to the low-frequency modes (e.g., $<30\ \mathrm{cm}^{-1}$).

In a harmonic description of the dynamics of a protein, the position $\vec{r}_i(t)$ of atom i can be written as a linear superposition of the normal-mode oscillations about its equilibrium position $\vec{R}_i$, i.e.,

$$\vec{r}_i(t) = \vec{R}_i + \sum_{\nu=1}^{3N-6} Q_\nu(t)\,\vec{\eta}_i(\nu), \tag{3}$$

where the six uniform rotations and translations are removed from the sum. The time-dependent factor $Q_\nu(t)$ in Eq. (3) takes the form $Q_\nu(t) = A_\nu \cos(\omega_\nu t)$, where A_ν is the amplitude of the oscillation, ω_ν is the frequency, and $\vec{\eta}_i(\nu)$ is the displacement vector of atom i in mode ν. In the classical limit ($\hbar\omega_\nu \ll k_b T$) at thermal equilibrium, each harmonic oscillator ν contains $k_b T$ in total energy ($1/2 k_b T$ for both momentum and positional degrees of freedom) via the equipartition theorem. This can be related to the maximum potential energy of a normal mode ν via

$$E_\nu = k_b T = \tfrac{1}{2}\omega_\nu^2 A_\nu^2. \tag{4}$$

Equation (4) allows one to compute the amplitude of the mode ν at thermal equilibrium; $A_\nu^2 = 2k_b T/\omega_\nu^2$. Using Eq. (3) to compute $\vec{U}_i^2(t)$ gives

$$\vec{U}_i^2(t) = \sum_{\nu,\mu=1}^{3N-6} Q_\nu(t) Q_\mu(t) \vec{\eta}_i(\nu) \cdot \vec{\eta}_i(\mu). \tag{5}$$

Substituting $Q_\nu(t) = A_\nu \cos(\omega_\nu t)$ for Q's and time averaging Eq. (5) at temperature T gives the result

$$\langle U_i^2 \rangle = \frac{k_b T}{m_i} \sum_\nu \frac{e_i^2(\nu)}{\omega_\nu^2}, \tag{6}$$

where $\int Q_\nu(t) Q_\mu(t) dt = 1/2 A_\nu^2 \delta_{\nu\mu}$ has been used. The vector $\vec{e}_i(\nu)$ is the mass weighted displacement vector $[\vec{\eta}_i(\nu) = \vec{e}_i(\nu)/\sqrt{m_i}]$ and is the normalized eigenvector of the dynamical matrix [see Eq. (9) below]. Following the same procedure used to obtain Eq. (6), a similar result for the cross terms, $\langle \vec{U}_i \cdot \vec{U}_j \rangle$, is obtained

$$\langle \vec{U}_i \cdot \vec{U}_j \rangle = \frac{k_b T}{\sqrt{m_i m_j}} \sum_\nu \frac{\vec{e}_i(\nu) \cdot \vec{e}_j(\nu)}{\omega_\nu^2}. \tag{7}$$

Equations (6) and (7) show that the contributions to the averages from each mode depends on $1/\omega_\nu^2$, so that the low-frequency modes dominate strongly.

B. Phonon functional method for normal-mode analysis

Normal-mode analysis relies on the harmonic approximation whereby small displacements of the atoms about a local energy minimum are assumed, allowing for an expansion of the potential energy $V(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_n)$ to second order as

$$V(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_n) \approx V_{r=R} + \langle \eta | \vec{\vec{F}} | \eta \rangle. \tag{8}$$

Here, the $3N \times 3N$ matrix $\vec{\vec{F}}$ is the force matrix (Hessian) with elements $F_{ij} = \partial^2 V/\partial x_i \partial x_j$, and $|\eta\rangle$ is the vector of displacements from equilibrium position R. The Lagrangian ($\mathcal{L} = T - V$) leads to a system of $3N$ coupled equations which determine the normal-mode displacements $|\eta\rangle$ and can be written in matrix form as

$$\vec{\vec{D}} |e_i\rangle = \lambda_i |e_i\rangle. \tag{9}$$

The matrix $\vec{\vec{D}}$ is the mass weighted force matrix (dynamical matrix) and $|e_i\rangle = \vec{\vec{M}}^{1/2} |\eta_i\rangle$ is the mass weighted displacement vector. The mode frequency is obtained from the eigenvalue $\lambda_i = \omega_i^2$. The normal modes of a system of N atoms contain $3N$ possible displacements with six of the modes corresponding to the three uniform rotations and three translations.

A key problem in solving Eq. (9) for large systems with many atoms ($N > 1000$) is that the storage of the dynamical matrix can require (e.g., for viruses) many terabytes (10^{12} bytes = 1000 Gbytes) of computer memory. For a single MS2 dimer there are roughly 4000 atoms, which requires ~2.6 Gbytes of memory to store the dynamical matrix and all eigenvectors. Although the MS2 dimer problem is within current limits of computer memory, it is clear that the memory limit is nearly surpassed for even "small" biological structures such as a single coat protein dimer. In addition, the computation time to diagonalize large matrices using standard techniques increases roughly as N^3.

Often (as in the case here) only the low-frequency modes of the dynamical matrix are desired in normal-mode calculations since these are likely to describe the global motions of the atoms that account for large conformational changes. In addition, the low-frequency spectrum dominates the contribution to mean-squared deviations of atoms or correlation matrices due to the $1/\omega^2$ dependence [see Eqs. (6) and (7)]. Clearly, alternatives to standard diagonalization where only the low-frequency modes need be calculated while avoiding explicit storage of the dynamical matrix can be quite useful in a variety of normal-mode analysis problems in biological physics.

One common alternative approach for large systems is to use a coarse-graining method such as the ENM [20,21], which (for example) considers only motions of the C_α carbons of the protein or treats each amino acid (or entire protein chain) as a rigid block (rotational translational block method) [22]. While the ENM has been used quite successfully to predict overall low-frequency displacement patters of normal modes, the method is phenomenological and comparison of frequencies calculated with more degrees of freedom (i.e., a larger basis set) has shown that ENM frequencies can be too high [23]. Inaccurate frequencies could potentially lead to inaccurate descriptions of the flexibility and rigidity of a protein as well as the correlation matrices.

A second alternative is to use an iterative scheme such as the Lanczos algorithm, which can search for the largest and smallest eigenvectors and eigenvalues of a large matrix [24]. While the Lanczos algorithm and its descendants have enjoyed much success, such algorithms are extremely difficult to employ on normal-mode calculations of proteins where all degrees of freedom are taken into account. This is due to the highly ill-conditioned nature of the atomic dynamical matrix where the ratio of the largest eigenvalue to the smallest can be on the order of $\lambda_L/\lambda_S = 10^8$. The ill-conditioned nature causes "ghosts and clones," i.e., the unwanted largest eigenvalues and eigenvectors are repeatedly found [25], while the desired low-frequency modes are continually missed.

The phonon functional method was developed by Dykeman and Sankey [17] to avoid the ghost and clone problem of the iterative Lanczos schemes and to provide a set of low-frequency modes of large biological systems such as a complete viral capsid to atomic detail, while only storing a few (~100) full vectors of length $3N$. The method is based on electronic structure order N methods [26], where the lowest M eigenvectors and eigenvalues of a large matrix are found by minimizing a functional G that depends on the matrix that is to be diagonalized. In the study of mechanical modes of proteins, the method begins with the user choosing the number M of *lowest*-frequency modes of the atomic dynamical matrix to calculate. Next, an initial set of M (e.g., $M = 100$) vectors of length $3N$, $|x_i^0\rangle$ (with $i \epsilon [1, M]$), are chosen at random; they neither need be orthogonal nor normalized since the minimization of the functional will enforce orthonormality. Once the initial set of vectors $|x_i^0\rangle$ have been chosen, the functional

$$G = \mathrm{Min}[\mathrm{Tr}_M[\overleftrightarrow{H} + \overleftrightarrow{H}(\overleftrightarrow{I} - \overleftrightarrow{S})]] \tag{10}$$

is minimized resulting in new vectors $|x_i^n\rangle$ at each step n of the minimization procedure. The matrices $\overleftrightarrow{H}$ and $\overleftrightarrow{S}$ in Eq. (10) are $M \times M$ matrices with elements

$$H_{ij} = \langle x_i | \overleftrightarrow{D}_s | x_j \rangle,$$

$$S_{ij} = \langle x_i | x_j \rangle. \tag{11}$$

The matrix $\overleftrightarrow{D}_s$ is the shifted dynamical matrix defined as $\overleftrightarrow{D}_s = \overleftrightarrow{D} - \lambda_L \overleftrightarrow{I}$, where λ_L is the largest eigenvalue of the dynamical matrix. The shift insures that entire eigenvalue spectrum of the dynamical matrix is negative and is necessary to guarantee a minimum in the functional G [27]. Once Eq. (10) is minimized, the vectors $|x_i\rangle$ span the space of the M lowest eigenvectors of the dynamical matrix $\overleftrightarrow{D}$.

We use the method of conjugate gradients [28] to minimize Eq. (10), which performs a line minimization at each step n of the minimization along a search direction $|p^n\rangle$. The method of conjugate gradients allows the line minimization to be done *analytically* by solving a cubic equation for the step size along $|p^n\rangle$ [27]. The resulting algorithm requires only M operations of the dynamical matrix on a vector for each minimization step. The operation of the dynamical matrix on a vector is computed analytically in a similar fashion to the forces [27]. Once the minimization is complete, the small $M \times M$ generalized eigenvalue problem can be solved, i.e.,

$$\overleftrightarrow{H}\vec{C}(i) = \lambda_i' \overleftrightarrow{S}\vec{C}(i), \tag{12}$$

where $\vec{C}(i)$ denotes one of the M eigenvectors of the $M \times M$ matrix $\overleftrightarrow{H}$ with eigenvalue λ_i'. Once the eigenvectors $\vec{C}(i)$ of $\overleftrightarrow{H}$ are obtained by solving Eq. (12), they can then be used to generate the M lowest eigenvectors of the dynamical matrix, $|e_i\rangle$, via

$$|e_i\rangle = \sum_{j=1}^{M} C_j(i) |x_j\rangle. \tag{13}$$

The eigenvalue $\lambda_i' = \lambda_i - \lambda_L$ in Eq. (12) results from using the shifted dynamical matrix $\overleftrightarrow{D}_s$ instead of the unshifted dynamical matrix $\overleftrightarrow{D}$ to obtain the elements of the $\overleftrightarrow{H}$ matrix using Eq. (11). The eigenvalue λ_i is the true eigenvalue of $\overleftrightarrow{D}$. Naturally this shift does not affect the eigenvectors.

The minimization of the functional [Eq. (10)] using conjugate gradients converges rapidly, typically within $N_s \approx 1000-3000$ minimization steps, requiring on the order of (using $M=100$ as an example) $M \times N_s = 10^5$ *operations* of the dynamical matrix on a vector. This can be roughly thought of as computationally equivalent to 10^5 molecular-dynamics steps. An operation of the dynamical matrix on a vector can be computed in a similar fashion to the forces in a molecular-dynamics simulation. For each term in the energy equation, the second derivatives are calculated analytically, multiplied by the appropriate elements of the vector and then added into a work vector; i.e., a temporary storage array in computer memory which stores intermediate results. After all energy terms have been done, the work vector contains the operation of the dynamical matrix on the original vector.

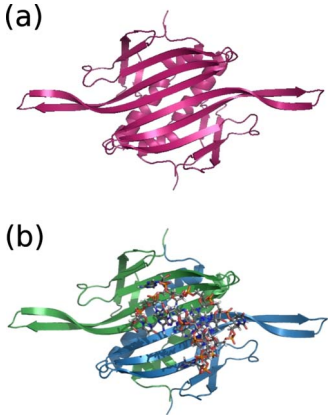

FIG. 2. (Color online) The two MS2 coat protein dimers studied with normal-mode analysis. The quasi-equivalent C subunits are colored magenta, while the modified A^* and B^* subunits are colored blue and green, respectively. (a) The unbound C/C dimer. (b) The asymmetric A^*/B^* dimer in complex with TR. The extended FG loops represent a state prior to FG-loop folding.

III. RESULTS OF THE ALL-ATOM NORMAL-MODE ANALYSIS

Coordinates for the MS2 C/C dimer and TR stem loop were extracted from x-ray crystal data at a resolution of 2.7 Å [6] [protein data bank (PDB) code 1ZDH]. These data were obtained from recombinant MS2 phage capsids soaked with RNA fragments encompassing the TR stem loop sequence and show RNA bound to both A/B and C/C dimers. From the data we constructed an RNA free symmetric C/C dimer (unbound dimer) and a modified C/C dimer with bound RNA stem loop (bound dimer). The bound dimer represents an intermediate state immediately after the TR binds to the symmetric C/C dimer but prior to FG-loop folding. For comparison of this intermediate state with the symmetric C/C dimer and the asymmetric A/B dimer, we will refer to the modified C protein subunits in the bound dimer as A^* and B^* to reflect the A and B subunits that they will eventually transform into. Hydrogen atoms were added to both structures following the known structures of amino acids at pH 7 before performing any minimization and normal-mode analysis.

The RNA free and RNA bound dimers used in the calculations are shown in Figs. 2(a) and 2(b), respectively. The bound dimer contained a total of $N=4275$ atoms and was minimized to a root-mean-square force (RMSF) of $<10^{-4}$ eV/Å using the AMBER energy model [29] in the SAGUARO biological simulation package [30]. Explicit water was not used in the calculation. Instead, we partially accounted for the energetic effects of the water using the generalized Born implicit solvent model [31–33]. The generalized Born model uses a dielectric screening methodology to approximately account for the electrostatic interactions of the protein and RNA with water as well as any effects of screen-

TABLE I. The four lowest-frequency modes and participation numbers W of the unbound and bound wild-type coat protein dimers in the all-atom method. Frequencies are given in cm^{-1} and the participation number W ranges from 1 to 3858. A participation of $W=3858$ corresponds to all atoms participating in the motion (such as in a uniform translation).

Unbound		Bound	
ω	W	ω	W
1.79	545	1.87	358
2.19	757	2.31	361
3.03	635	3.04	497
3.47	1042	3.43	663

ing between the protein and RNA due to water. This allows Coulomb terms to be truncated at a specified cutoff (10 Å is used in this work). Although the generalized Born model allows for the energetic effect of ions to be included in the model through a Debey-Hückel term, we did not include this effect. Also, no effects from an external heat bath are included in the model. Once the bound dimer was minimized, the TR stem loop consisting of 417 atoms was removed and the C/C dimer alone (unbound dimer) was minimized to an RMSF $<10^{-4}$ eV/Å under the constraints of $C2$ symmetry to yield the minimized unbound dimer with a total of $N=3858$ atoms. This minimization procedure provided bound and unbound dimers which were within similar energy minima, allowing for a comparison of the frequencies and displacement patterns. The all-atom root mean squared deviation comparing both unbound and bound dimers with the original 2.7 Å resolution x-ray structure was <1.53 Å, which showed good agreements between the PDB file and minimized structures.

We then performed an all-atom normal-mode analysis on both the bound and unbound dimers using the phonon functional method (see Sec. II B). For the RMSF tolerance of $<10^{-4}$ eV/Å all frequencies for both the bound and unbound dimers were positive. In both cases, $M=100$ normal modes were calculated which resulted in the frequencies ω of these modes, ranging from roughly $\omega=1.79$ to 29.9 cm^{-1} for both structures. The eigenvectors for the unbound dimer contained $3N=11\,574$ components, while those for the bound dimer contained $3N=12\,825$ due to the extra 417 atoms in the TR stem loop. When comparing normal modes between the bound and unbound dimers, we took the portion of the eigenvector which corresponds to the protein to give eigenvectors of identical length.

Table I lists the lowest four frequency modes and participation numbers W_ν for the bound and unbound dimers. The participation number is used as a measure of how many atoms are moving in a given normal-mode pattern and is calculated from the formula $W_\nu=e^{S_\nu}$, where S_ν is the (informational) entropy for mode ν given by $S_\nu=\Sigma_{i=1}^{N}\eta_i^2(\nu)$. Here, the displacement vector $|\eta_\nu\rangle$ of length $3N$ is normalized instead of the eigenvector of the dynamical matrix, so that W_ν ranges from 1 to $N=3858$. Thus, a small (large) participation number indicates a localized (global) mode. Remarkably, the frequencies of the lowest modes remain roughly the same despite the addition of the TR stem loop. For example, the frequency of the lowest mode at 1.79 cm^{-1} increases by 0.08 cm^{-1} with similar shifts for the other three modes. However, a dramatic shift in the participation numbers for the modes can be seen; the lowest-frequency modes of the bound dimer become *more* localized. In the case of the lowest-frequency mode, the participation number decreases by ~40% upon addition of the TR stem loop, with similar shifts for the other three modes.

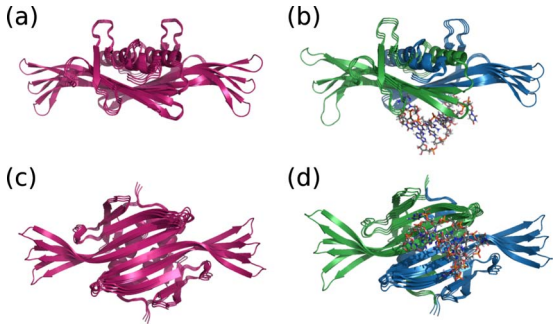

FIG. 3. (Color online) Displacement patterns for the [(a) and (b)] lowest- and [(c) and (d)] third-lowest-frequency modes of the MS2 bacteriophage coat protein dimers with and without bound RNA in the all-atom method. The quasi-equivalent C subunits are colored magenta, while the modified C subunits A^* and B^* are colored blue and green, respectively. Maximum and minimum displacements of the structure are shown overlaid with the equilibrium structure. (a) Lowest-frequency mode of the unbound C/C dimer at 1.79 cm^{-1} and (b) bound A^*/B^* dimer at 1.87 cm^{-1}. (c) Third-lowest-frequency mode of the unbound C/C dimer at 3.03 cm^{-1} and (d) bound A^*/B^* dimer at 3.04 cm^{-1}.

Figure 3 illustrates the displacement patterns for two of the four lowest-frequency modes in Table I (the first and third). Mode patterns for the unbound dimer are shown in Figs. 3(a) and 3(c), while mode patterns for the bound dimer are shown in Figs. 3(b) and 3(d). The modes are represented by a superposition of atomic structures that are obtained by displacing the atoms along the ± directions of the normal mode. For the purposes of visual clarity, we use an amplitude of 10 Å to obtain the maximum and minimum displacements, which is approximately 10 times greater than the thermal amplitude. The lowest-frequency modes are shown in Figs. 3(a) and 3(b) and have frequencies of 1.79 cm^{-1} for the unbound dimer and 1.87 cm^{-1} for the bound dimer. The low-frequency modes for both the unbound and bound dimers are predominantly localized on the FG loops which move simultaneously in the same direction. The distinct difference between the unbound and bound dimers can be found in the B^* protein subunit (colored green) of the bound dimer, which has a disproportionate amplitude (~40% larger) when compared with the A^* subunit (colored blue) and the C subunit [colored magenta in Fig. 3(a)].

Similarly, Figs. 3(c) and 3(d) show the third-lowest-frequency modes for both unbound and bound dimers with frequencies at 3.03 and 3.04 cm^{-1}, respectively. Again, both modes are predominantly localized on the FG loops which move simultaneously in the same direction. However, the direction they move toward is orthogonal to the direction

pictured for the lowest-frequency modes in Figs. 3(a) and 3(b), i.e., they move out of the plane with respect to Figs. 3(a) and 3(b). The relative amplitudes of the FG loops are identical in the symmetric unbound dimer [Fig. 3(c)] but different in the asymmetric bound dimer [Fig. 3(d)] where the B^* protein subunit has an increased amplitude when compared with the A^* protein subunit ($\sim$14% larger). Thus, in both low-frequency modes shown in Fig. 3, equal displacements of the FG loops in the symmetric unbound dimer shift, upon addition of TR, to favor displacements in the FG loop of the B^* protein subunit. This is a result of the localization of the displacement pattern (see Table I) onto the B^* protein subunit, which results from TR binding. The localization of vibrational motion on the FG loop of the B^* subunit seems to occur in other low-frequency modes but is most dominant in these two modes shown in Fig. 3.

To understand how this observed localization of the vibrational modes can affect the fluctuations of the atoms away from their equilibrium positions, we calculated the B factors for each residue using Eqs. (1) and (6). The B factors allowed for a determination of the change in FG-loop flexibility of a C/C dimer due to binding of the TR stem loop. Figure 4(a) plots the B factors for residues in the unbound C protein subunit (black line) along with the B factors for the subunits A^* and B^* (blue and red lines, respectively). Each protein subunit is 129 residues in length. It is clear from Fig. 4(a) that the B factors of most residues remain strikingly unaltered after TR binding. However, two regions of amino acids show significant changes to their B factors in the A^* and B^* subunits due to the presence of TR. The first is a noticeable decrease (increase) in the flexibility of residues 48–54 in the B^* (A^*) subunit, which correspond to the EF loops in these coat proteins [see Fig. 4(c)]. The second is an increase (decrease) in the flexibility in residues 66–82 in the B^* (A^*) subunit, which correspond to the FG loops [see Fig. 4(b)]. For further clarity, we show the *change* in B factors (B factor for an A^* or B^* subunit minus the B factor for the C subunit) plotted as a color code on a symmetric C/C dimer in Fig. 4(d). Sections that become more (less) flexible as a result of TR binding are colored red (blue), while sections that are unchanged are colored gray. The two main areas that are affected by TR, the EF and the FG loops, show either an increase or a decrease in flexibility.

In order to understand the role of TR in this localization of the displacement pattern on the FG loop of the B^* subunit, its location in the bound dimer needs to be taken into account. Figure 5(a) shows the contacts that the TR stem loop makes to an A/B dimer. The stem loop is shown colored gray in space filling representation, while the protein is shown as a ribbon diagram. The phosphodiester backbone of TR makes contacts to residues R49, S51, and S52 in the EF loop of the B protein subunit (colored green), while no contacts are made with the EF loop in the A protein subunit (colored blue). This asymmetry in the TR contacts along with the B-factor results, which show that only the EF- and FG-loop regions of the protein are significantly effected by TR binding, suggests that the EF loop in the B subunit may be instrumental for the conformational change to a folded FG loop, and that TR's contact with the EF loop triggers it.

To further corroborate that the TR contact to the EF loop can affect the flexibility of the FG loop, we computed the correlation matrix elements between residues for the symmetric C/C dimer using Eqs. (2) and (7). The correlation matrix, shown in Fig. 5(b), was computed using the 100 lowest mode vectors. To check for convergence of the matrix elements, we computed an additional 100 eigenvectors (for a total of 200 normal modes) and recomputed the correlation matrix. No significant changes were found suggesting that 100 normal modes are sufficient for convergence of the correlation matrix elements. The matrix is plotted by the residue number (1–129) for each of the two proteins in the dimer on a color scale, with blue representing negatively and red positively correlated regions. The two vertical black boxes identify the two FG loops of the dimer, while the two horizontal black boxes represent the EF loops. The correlation plot implies that the two EF loops (residues 48–54) are correlated with the FG loops (residues 66–82). This is most likely due to the fact that the EF and FG loops are connected by the F β strand (of approximately ten amino acids in length).

Although the correlation matrices do not conclusively show that EF-loop suppression enhances FG-loop motion, it does suggest that EF- and FG-loop motions may be coupled through many vibrational modes. Thus, RNA binding could potentially affect global modes involving FG-loop motions by strongly interacting with the EF loop. Earlier theoretical modeling performed by Hawkins and McLeish [15] suggested a similar mechanism for the MET repressor where allosteric effects were amplified by "enslaved fast modes," i.e., high-frequency modes that are coupled to the low-frequency global modes.

IV. DISCUSSION

The results of the atomistic normal-mode analysis of the unbound and bound dimers suggest a mechanism for how TR binding affects the folding of just one FG loop in the MS2 coat protein dimer via an allosteric effect. Upon binding of the TR stem loop to a symmetric C/C dimer, its phosphate backbone makes a contact with the EF loop in the B^* protein subunit, hence inhibiting motion of this loop [see Figs. 4(c) and 4(d)]. Since the EF loop is correlated with the FG loop due to their connection via the F β strand, the FG loop in the B^* subunit becomes more flexible. The physical basis for the allosteric mechanism can best be described in a "stringed instrument" analogy, where pressure from the phosphodiester contact on the EF loop in the B^* subunit "shortens the string," i.e., the F β strand, causing a small up-shift (0.08 cm^{-1} for the lowest-frequency mode) in the frequency and localization of the displacement pattern. Interestingly, the localization effect is strongest in the four lowest modes (two of which are shown in Fig. 3) which correspond to vibrational motions that are nearly completely localized on the two FG loops of the dimer. A simple calculation of the contribution of each mode to the *total* change in B factors observed in the FG loop of the B^* subunit show that the four lowest modes contribute over 90% to the total change.

At the present time, it is too computationally difficult to calculate how the increased amplitudes of the four lowest modes would contribute to the folding of the FG loop, and hence determine the exact pathway that each atom in the FG

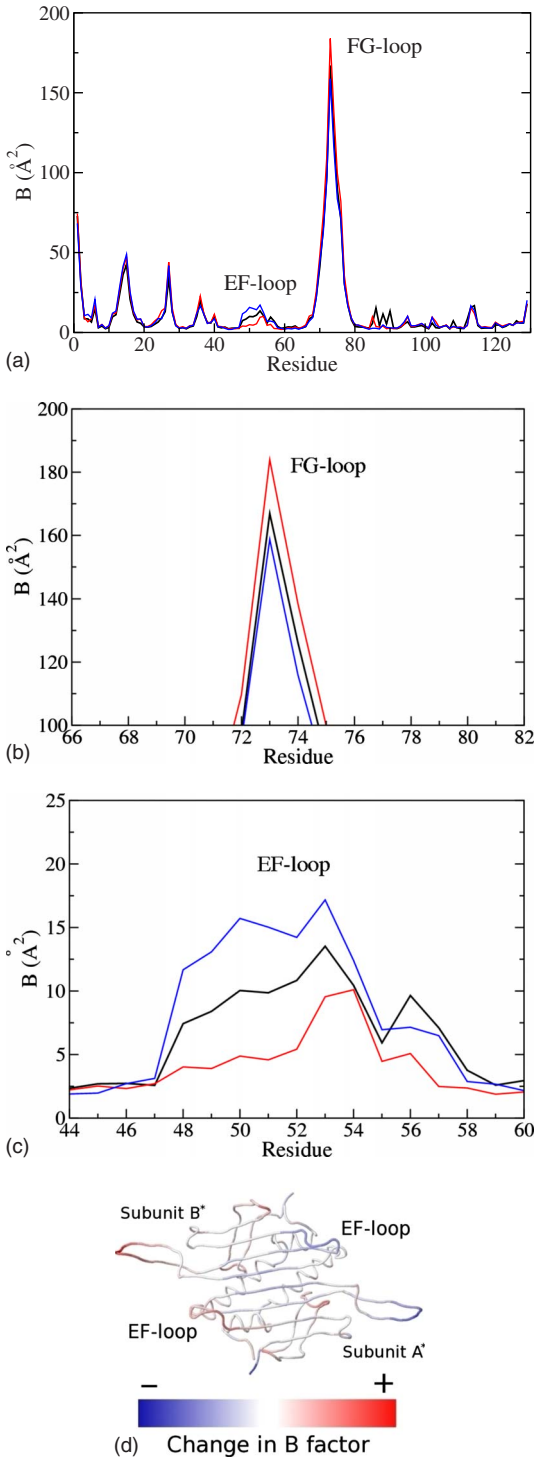

FIG. 4. (Color online) All-atom B-factor results for the MS2 coat protein dimers. (a) B factors B_R of residues R in the coat protein subunits before and after TR binding. Black lines represent the C protein subunit prior to TR binding, while the blue and red lines represent the A^* and B^* protein subunits after TR binding, respectively. (b) Zoomed in view of the FG-loop region showing the increase (decrease) in the B^* (A^*) subunits B factors. (c) Zoomed in view of the EF-loop region showing the decrease (increase) in the B^* (A^*) subunits B factors. (d) Change in B factors of the wild-type coat protein subunits plotted on a symmetric C/C dimer. Regions of increased flexibility (+) are shown in red, while regions of decreased flexibility (–) are shown in blue.

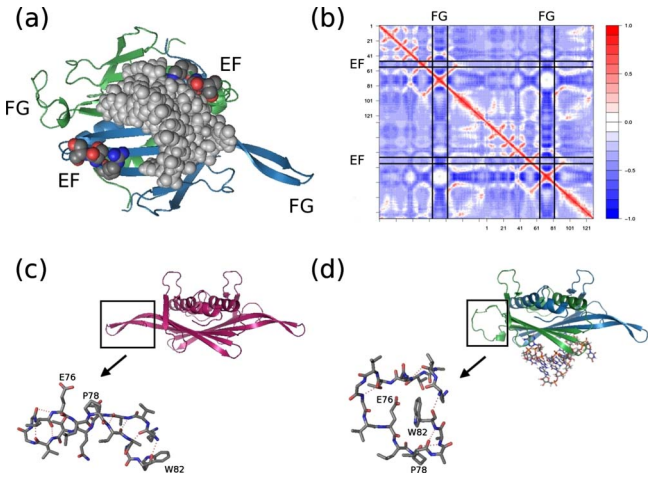

FIG. 5. (Color online) Mechanism of the allosteric effect. (a) Illustration of the asymmetric contacts made by the phosphodiester backbone of TR with the EF loop of the coat protein in an A/B dimer. The TR stem loop and residues R49, S51, and S52 are shown in space filling representation with the TR stem loop colored gray. The TR stem loop contacts the EF loop in the B subunit (green), while no contact is made in the A subunit (blue). (b) Correlation matrix for the symmetric C/C dimer computed from the lowest 200 normal modes. Red areas indicate positive correlations, while blue areas indicate negative correlations. Close-up views of the FG loops of (c) a C subunit and (d) a B subunit. Residues E76, P78, and W82 are labeled to show the structural rearrangements required to form a B conformer from a C conformer. Hydrogen bonds are shown as red dashed lines.

loop follows during the conformational switch. This is primarily due to difficulties in performing molecular-dynamics simulations on the FG-loop refolding, which appears to occur on the microsecond-millisecond time frame. In order to estimate how strong the influence of TR binding is on the four lowest-frequency modes, we computed the average amplitude of atoms within the FG loop of the B^* subunit before and after TR binding as well as the change in their total energy. Within a harmonic analysis, the total energy (or equivalently the maximum kinetic energy) of an atom i in the full protein dimer at temperature T is given by the sum over all normal modes ν,

$$E_i = k_b T \sum_{\nu} e_i^2(\nu). \tag{14}$$

This allows for a calculation of the change in energy of atoms within the FG loop of the B^* protein subunit due to TR binding using

$$\Delta E_{FG} = k_b T \sum_{i,\nu} [e_i'^2(\nu) - e_i^2(\nu)], \tag{15}$$

where the sum on atoms i is only over atoms in the FG loop of the B^* subunit and $e_i'^2(\nu)$ denotes the eigenvector of atom i in the TR bound dimer. By neglecting the sum over all modes ν in Eq. (15), one can calculate the contribution to the change in energy of the B^* subunit's FG loop from each mode. The four lowest modes are altered the most by TR binding, and their contribution to the change in energy and

TABLE II. Average increase in amplitude and change in kinetic energy of atoms in the FG loop of the B^* subunit after TR binding for the four low-frequency modes in the all-atom method. The average increase in amplitude is given as a percentage of the original amplitude, while the change in kinetic energy is given in units of k_bT. The total energy of a single mode at thermal equilibrium is k_bT.

ω (cm^{-1})	ΔA (%)	ΔKE (k_bT)
1.76	41.4	0.32
2.19	−39.2	−0.26
3.03	14.1	0.20
3.47	−11.8	−0.13

average amplitude (given as a percentage of the amplitude of the unbound dimer) of atoms in the B^* protein subunit's FG loop is shown in Table II. It is clear that the lowest mode redistributes a large amount of energy ($0.32k_bT \sim 33\%$ of its total energy) to the FG loop of the B^* subunit upon TR binding. This energy redistribution is significant and on the order of weak hydrogen bond energies in protein structures ($\sim 0.5k_bT$). Similarly, the amplitude of atoms in the FG loop of the B^* subunit of the TR bound dimer increases on average by 41.4% when compared with the amplitude of atoms in the FG loop of the unbound dimer. Similar results are seen for the third-lowest-frequency mode.

A peculiarity is that the second and fourth lowest modes actually exhibit a decrease in the amplitude of the B^* subunit, which is due to a localization of the displacement pattern on the A^* rather than on the B^* subunit. The second lowest mode is almost identical to the lowest-frequency mode, but it corresponds to FG loops moving simultaneously in *opposite* directions instead of the same direction. Similarly, the fourth lowest mode exhibits FG loops moving simultaneously in opposite directions when compared with the third-lowest-frequency mode. However, when averaged over all four modes (plus the higher-frequency modes), the result is a clear increase in the flexibility of the B^* subunit's FG loop and a decrease in the flexibility of the A^* subunit's FG loop (see the B-factor results in Fig. 4). As a result, atoms in the FG loop of the B^* subunit receive a net increase in energy on the order of $0.1k_bT$. This could potentially disrupt weak hydrogen bonds that hold the FG loop in an extended conformation. Figures 5(c) and 5(d) illustrate the reorganization of the FG loop that is necessary to convert a C subunit into a B subunit upon TR binding, requiring the breaking and formation of number of hydrogen bonds.

V. COMPARISON WITH ELASTIC NETWORK MODEL

An interesting question is whether simplified models, such as ENM [20,21], can reproduce the localization and B-factor results of the all-atom full force field normal-mode analysis above. The ENM uses the Tirion potential [20], a simplified spring potential in which pairs of atoms separated by a distance less than a user specified cutoff are connected by a spring of arbitrary spring constant. In the simplest ENM, each spring is identical. However, one could construct a more sophisticated ENM that uses a distance-dependent spring constant.

The original Tirion model [20] connected springs between all atoms within a specified cutoff. Our goal here is to compare the phenomenological Tirion potential with the more sophisticated empirical AMBER potential. To do this, we use the same procedure as Tirion to calculate the modes, but consider all $3N$ degrees of freedom when calculating modes instead of only dihedral or C_α motions. This allows for an easy comparison with the all-atom AMBER results above. All springs are given the same spring constant and we perform the ENM analysis twice using two different cutoffs for the spring network: a 5 Å cutoff that is roughly consistent with Tirion's original work [20] and a slightly larger cutoff of 8 Å that is comparable to the cutoffs used in other ENM studies of virus motions [21] and protein B factor and correlation computations [34]. Atom pairs outside the 5 or 8 Å cutoff are not connected via a spring. The phonon functional method is used to obtain the lowest 100 normal modes as in the all-atom full force field method above. For both these cutoffs used, the displacement patterns of the computed normal modes were essentially identical. The only difference between the results of the 5 and 8 Å cutoff calculations seemed to be in the overall relative values of the frequencies of the modes. Given the similarities between the two, we will only show displacement patterns and B factors for the 8 Å cutoff results in what follows.

Figure 6 shows the B factors computed using normal modes from the elastic network model for residues in the unbound C protein subunit (black line) along with the B factors for the subunits A^* and B^* (blue and red lines, respectively). Since the choice of units (and strength) of the spring constant is user specified, the results are shown in arbitrary units. What is important for the comparison to Fig. 4 (the all-atom results) are the relative heights of the various peaks. A key difference is that the ENM shows an increase in flexibility in *both* FG loops. This is contrary to our results, and also to experiment, which shows that TR binding biases *only* the B^* subunits FG loop to fold and leaves the A^* subunit unaffected.

A comparison of the lowest-frequency mode for the unbound [Fig. 7(a)] and bound [Fig. 7(b)] dimers predicted using the ENM shows no (or very little) localization of the displacement pattern on the FG loop of the B^* subunit. Specifically, we found that the lowest-frequency mode of the bound dimer only showed an increase of $\sim 0.87\%$ in the average amplitude of the atoms in the FG loop of the B^* subunit when compared with the unbound dimer. Using Eq. (15), this would be equivalent to a redistribution of $0.033k_bT$ in energy to the B^* FG loop, which is tenfold smaller than that predicted by the all-atom full force field method. Examination of the other low-frequency modes computed with the ENM also showed no strong localization effect on the B^* subunit. Hence, the increase in the B factors in the FG loop seen in Fig. 6(b) is likely a result of the combined effect of small differences in many modes unlike in the all-atom full force field method where the change in flexibility of the B^* subunit's FG loop is mostly due to the four lowest modes. It is clear that, at least in the case of MS2, the simplified

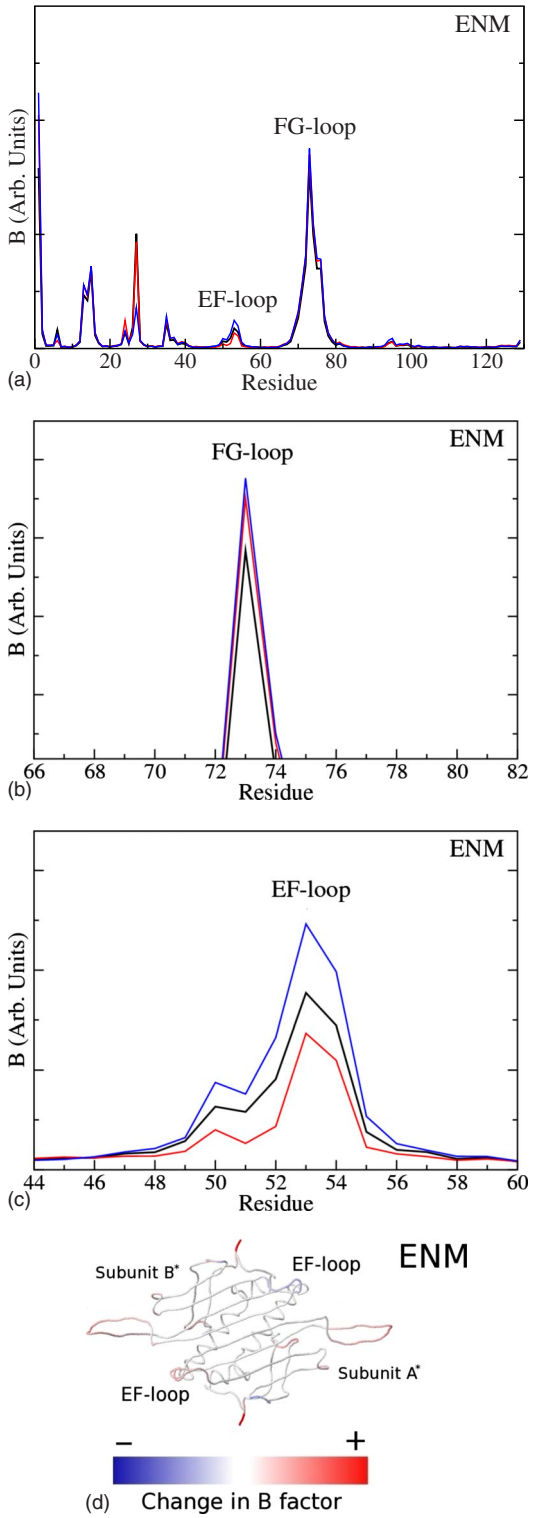

FIG. 6. (Color online) Elastic network model B-factor results for the MS2 coat protein dimers. (a) B factors B_R of residues R in the coat protein subunits before and after TR binding. Black lines represent the C protein subunit prior to TR binding, while the blue and red lines represent the A^* and B^* protein subunits after TR binding, respectively. (b) Zoomed in view of the FG-loop region. (c) Zoomed in view of the EF-loop region. (d) Change in B factors of the wild-type coat protein subunits plotted on a symmetric C/C dimer. Regions of increased flexibility (+) are shown in red, while regions of decreased flexibility (–) are shown in blue.

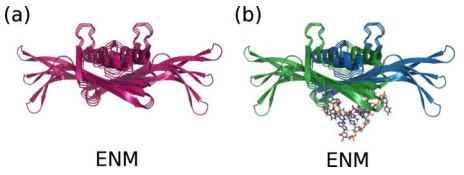

FIG. 7. (Color online) Displacement patterns for the lowest-frequency modes of the MS2 bacteriophage coat protein dimers with and without bound RNA in the elastic network model. The quasi-equivalent C subunits are colored magenta, while the modified C subunits A^* and B^* are colored blue and green, respectively. Maximum and minimum displacements of the structure are shown overlaid with the equilibrium structure. (a) Lowest-frequency mode of the unbound C/C dimer and (b) bound A^*/B^* dimer.

potential-energy model of identical springs used in the ENM is not sufficient to account for the subtle differences in normal modes that result from TR binding. A more sophisticated force field (such as AMBER) is hence required to explain the dynamic allostery in bacteriophage MS2.

VI. CONCLUSION

By examining the mechanical modes of the MS2 coat protein dimer in the presence and absence of a 19 nucleotide RNA stem loop to atomic detail, we have been able to suggest a mechanism that may be responsible for the conformational switch of the MS2 coat protein dimer from a symmetric to an asymmetric state. Our analysis suggests that this conformational switch results from the localization of only the lowest-frequency modes on just one of the FG loops, and this could be due to the repression of EF-loop motions in the B^* subunit, which is in direct contact with TR. An important consequence is that this allosteric effect is not sequence specific, i.e., it only requires contact of the phosphodiester backbone of the stem loop with the EF loop. It can therefore be triggered by other stem loops from within the MS2 genome, which further corroborates the view that the genome plays an active role in the assembly of the MS2 capsid [8].

Analysis of the allosteric effect using an ENM showed that the simplified single-parameter harmonic potential that is usually used in ENMs was not sensitive enough to pick up the subtle localization of the vibrational modes that was predicted in the all-atom full force field method. One possibility for this discrepancy could be that the 5–8 Å cutoff used in this work was inappropriate. Another possibility is that the single spring constant was insufficient to represent the physics of the system. In either case, the effect of TR binding on the normal modes seems to be incorrectly modeled within the ENM framework. We note that it may be possible, for example, to adjust the ENM cutoff to obtain results that are closer to those of the all-atom full force field method.

Our results are consistent with current views of allostery in which ligand binding is interpreted as resulting in a redistribution of protein conformational ensembles [35]. In the case of MS2, the ensemble consists of A/B and C/C states, with a shift of balance toward the C/C state when the coat protein dimer is RNA free. Upon RNA binding, the dimer experiences an increase in the kinetic energy (and hence free

energy) of one of its FG loops, resulting from a localization of the vibrational modes. This shifts the population probabilities of A/B and C/C states to favor the A/B state, in which one FG loop is folded. The TR stem loop may in addition influence the energy landscape, for example, by decreasing the barrier height between extended and folded FG-loop states; our proposed vibrational mechanism does not rule out a combined effect. In any case, our analysis shows that the contribution of the vibrational modes to the allosteric effect is significant. This suggests that the localization of vibrational modes and the resulting shift in vibrational displacement patterns observed here should be important also in understanding a variety of other allosteric effects. Future work on RNA-protein interactions in other viruses such as tobacco mosaic virus and cowpea chlorotic mottle virus will show if the type of dynamic allostery observed for MS2 corresponds to a generic mechanism that occurs also in other viruses.

ACKNOWLEDGMENT

R.T. would like to gratefully acknowledge financial support from the Leverhulme Trust.

[1] W. Fiers, R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, and M. Ysebaert, Nature (London) **260**, 500 (1976).
[2] P. G. Stockley, N. J. Stonehouse, and K. Valegard, Int. J. Biochem. **26**, 1249 (1994).
[3] H. Lago, A. M. Parrott, T. Moss, N. J. Stonehouse, and P. G. Stockley, J. Mol. Biol. **305**, 1131 (2001).
[4] R. Golmohammadi, K. Valegard, K. Fridborg, and L. Liljas, J. Mol. Biol. **234**, 620 (1993).
[5] K. Valegard, J. B. Murray, P. G. Stockley, N. J. Stonehouse, and L. Liljas, Nature (London) **371**, 623 (1994).
[6] K. Valegard, J. B. Murray, N. J. Stonehouse, S. van den Worm, P. G. Stockley, and L. Liljas, J. Mol. Biol. **270**, 724 (1997).
[7] R. L. Duda, P. D. Ross, N. Cheng, B. A. Firek, R. W. Hendrix, J. F. Conway, and A. C. Steven, J. Mol. Biol. **391**, 471 (2009).
[8] K. Toropova, G. Basnak, R. Twarock, P. G. Stockley, and N. A. Ranson, J. Mol. Biol. **375**, 824 (2008).
[9] R. Koning, S. van den Worm, J. R. Plaisier, J. van Duin, J. P. Abrahams, and H. Koerten, J. Mol. Biol. **332**, 415 (2003).
[10] D. S. Peabody and K. R. Ely, Nucleic Acids Res. **20**, 1649 (1992).
[11] P. G. Stockley, O. Rolfsson, G. S. Thompson, G. Basnak, S. Francese, N. J. Stonehouse, S. W. Homans, and A. E. Ashcroft, J. Mol. Biol. **369**, 541 (2007).
[12] G. Basnak, V. L. Morton, O. Rolfsson, N. J. Stonehouse, A. E. Ashcroft, and P. G. Stockley, J. Mol. Biol. **395**, 924 (2010).
[13] D. L. D. Caspar, Biophys. J. **32**, 103 (1980).
[14] E. C. Dykeman, P. G. Stockley, and R. Twarock, J. Mol. Biol. **395**, 916 (2010).
[15] R. J. Hawkins and T. C. B. McLeish, Biophys. J. **91**, 2055 (2006).
[16] R. J. Hawkins and T. C. B. McLeish, Phys. Rev. Lett. **93**, 098104 (2004).
[17] E. C. Dykeman and O. F. Sankey, Phys. Rev. Lett. **100**, 028101 (2008).
[18] J. Kuriyan and W. I. Weis, Proc. Natl. Acad. Sci. U.S.A. **88**, 2773 (1991).
[19] A. W. Van Wynsberghe and Q. Cui, Structure (London) **14**, 1647 (2006).
[20] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[21] F. Tama and C. L. Brooks III, J. Mol. Biol. **345**, 299 (2005).
[22] F. Tama, F. X. Gadea, O. Marques, and Y. H. Sanejouand, Proteins: Struct., Funct., Genet. **41**, 1 (2000).
[23] H. W. T. van Vlijmen and M. Karplus, J. Mol. Biol. **350**, 528 (2005).
[24] C. Lanczos, *Applied Analysis*, 2nd ed. (Dover Publications, New York, 1988).
[25] O. F. Sankey, D. A. Drabold, and A. Gibson, Phys. Rev. B **50**, 1376 (1994).
[26] P. Ordejon, D. A. Drabold, R. M. Martin, and M. P. Grumbach, Phys. Rev. B **51**, 1456 (1995).
[27] E. C. Dykeman and O. F. Sankey, J. Phys.: Condens. Matter **21**, 035116 (2009).
[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN 77: The art of scientific computing*, 2nd ed. (Cambridge University Press, Cambridge, England, 1986).
[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] E. C. Dykeman and O. F. Sankey (unpublished).
[31] D. Bashford and D. A. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).
[32] V. Tsui and D. A. Case, Biopolymers **56**, 275 (2001).
[33] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, J. Phys. Chem. **100**, 19824 (1996).
[34] I. Bahar, A. R. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997).
[35] K. Gunasekaran, B. Ma, and R. Nussinov, Proteins: Struct., Funct., Bioinf. **57**, 433 (2004).