THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Impact of Hydrogen Bonding in the Binding Site between Capsid Protein and MS2 Bacteriophage ssRNA

Lokendra Poudel,[†] Reidun Twarock,[‡] Nicole F. Steinmetz,[§,∥,⊥,#,▽] Rudolf Podgornik,[○,◆] and Wai-Yim Ching*[,†]

[†]Department of Physics and Astronomy, University of Missouri-Kansas City, Kansas City, Missouri 64110, United States

[‡]Department of Mathematics and Biology and York Centre for Complex Systems Analysis, University of York, York YO10 5DD, United Kingdom

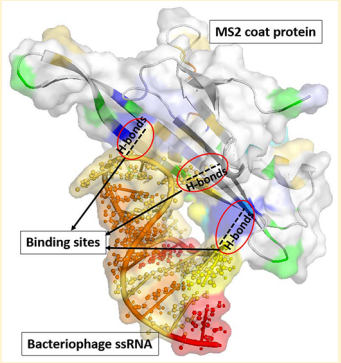
[§]Department of Biomedical Engineering, [∥]Department of Radiology, [⊥]Department of Materials Science and Engineering, [#]Department of Macromolecular Science and Engineering, and [▽]Case Comprehensive Cancer Center, Division of General Medical Sciences-Oncology, Case Western Reserve University, Cleveland, Ohio 44106, United States

[○]Department of Theoretical Physics, J. Stefan Institute, SI-1000 Ljubljana, Slovenia

[◆]Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia

Supporting Information

ABSTRACT: MS2 presents a well-studied example of a single-stranded RNA virus for which the genomic RNA plays a pivotal role in the virus assembly process based on the packaging signal-mediated mechanism. Packaging signals (PSs) are multiple dispersed RNA sequence/structure motifs varying around a central recognition motif that interact in a specific way with the capsid protein in the assembly process. Although the discovery and identification of these PSs was based on bioinformatics and geometric approaches, in tandem with sophisticated experimental protocols, we approach this problem using large-scale ab initio computation centered on critical aspects of the consensus protein−RNA interactions recognition motif. DFT calculations are carried out on two nucleoprotein complexes: wild-type and mutated (PDB IDs: 1ZDH and 5MSF). The calculated partial charge distribution of residues and the strength of hydrogen bonding (HB) between them enabled us to locate the exact binding sites with the strongest HBs, identified to be LYS43-A^{-4}, ARG49-C^{-13}, TYR85-C^{-5}, and LYS61-C^{-5}, due to the change in the sequence of the mutated RNA.



1. INTRODUCTION

One of the most persistent questions regarding the nature of life is the character of the driving forces in the process of molecular evolution, a question directly related to the essence of genetic mutation in the nucleotide sequence of the two fundamental nucleic acid types (DNA and/or RNA) involved in the distinct functional states comprising the replication, translation, and control of gene expression. Presently, there appear to be two approaches relevant to the study of this problem:[1] (i) large statistical data approaches in analyzing the mutation pairs to construct the fitness landscape for prediction of the evolutionary trajectory in the Darwinian sense,[2,3] and (ii) fundamental quantum mechanical approaches studying interactions among or between the three fundamental classes of biomolecules: nucleic acids, proteins, and lipids.[4−6] Although fundamentally different, these two approaches both rely on bioinformatics techniques that can thus be used as a bridge to provide mutual enhancement.

Viruses are nanoscale systems programmed to self-assemble into discrete particles encapsulating their genomic cargo. They contain a proteinaceous capsid that encloses the viral genome in the form of DNA or RNA, making a protective shell to transmit the infectious genetic cargo in a functionally intact state through space and time. This shell consists of a large number of copies of just a few types of capsid proteins (CPs) that usually but not always[7] exhibit icosahedral symmetry.[8] The process of packaging the viral genome is a critical step in the assembly of infectious viruses. There are strong indications from in vitro and in vivo studies, as well as mathematical modeling, that the capsid self-assembly proceeds through highly specific interactions between particular sections of RNA that are thought of as containing the so-called packaging signals (PSs, also defined as the origin of assembly sites, OASs), providing the specific binding sites between the CPs and viral RNA.[9−12] Experimental results on the structural determinants of the high affinity CP binding site (usually denoted as TR) and sequence variants thereof have been available from different experimental groups since the 1990s. However, when an all-atom normal-mode analysis identified the structural features of TR responsible for the allosteric switch between the two protein dimer conformations required for capsid assembly,[13,14] it became clear that multiple dispersed RNA structure elements

Received: March 18, 2017
Revised: May 3, 2017
Published: June 5, 2017

DOI: 10.1021/acs.jpcb.7b02569
J. Phys. Chem. B 2017, 121, 6321−6330

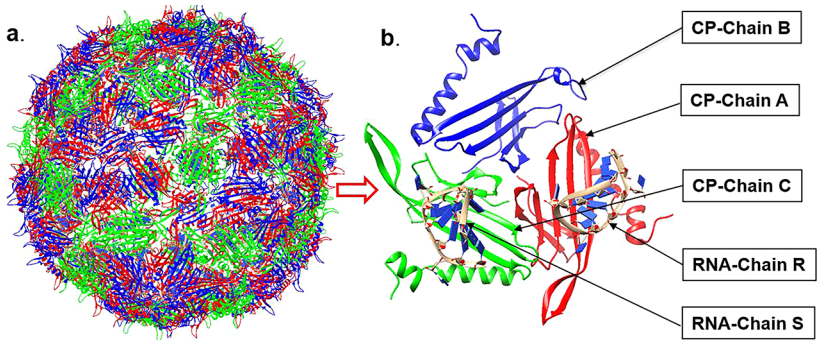


Figure 1. Structure of icosahedral viral capsid of the MS2 coat protein (CP)-RNA complex with (a) biological unit (60 copies of asymmetrical unit) and (b) asymmetrical unit. The ribbon denotes the MS2 CP, and the ribbon with tile denotes ssRNA.

in the MS2 genome other than TR could also trigger this allosteric switch and thus act as PSs. Here, allosteric switch (or allosteric control) is the regulation of an enzyme by binding an effector molecule at a site other than the enzyme's active site, which is now known as the packaging signal hypothesis. This prompted the development of an interdisciplinary approach to identify such PSs explicitly in the MS2 genome,[15] yielding results that are in agreement with CLIP-Seq experiments,[16] enabling a correct interpretation of sequence-specific collapse of the genomic RNA in the presence of cognate coat protein[17] as a result of multiple PSs. The primary CP−RNA interactions within the virus capsid are the key elements controlling its self-assembly.[18,19] Although they are *grosso modo* sequence specific, we do not fully understand the details as to how the specificity of RNA PSs is achieved and what exactly its nature is in terms of fundamental molecular interactions.

Understanding the roles of PSs in the process of virus self-assembly would provide important insights relevant also for drug design, control, and eradication of viral infections and epidemic threats (HBV, HIV, Ebola, Polio, Zika, etc.). Furthermore, detailed knowledge of the virus self-assembly processes would enable the synthesis of novel virus-inspired or templated materials with unique applications in medicine,[20] bionanotechnology,[21,22] and nanomaterials science.[19,23,24] The elucidation of the specific, sequence-dependent RNA−protein interactions would furthermore illuminate the fundamental aspects of gene regulation and possibly enhance the understanding of the physical basis of gene editing.[25] In principle, the equilibrium and kinetic aspects of viral self-assembly can be understood on the basis of molecular interactions between CPs or their subunits and the viral genome based on a combination of fundamental long(er)-ranged electrostatic and hydrophilic−hydrophobic interactions together with short(er)-ranged specific contacts between certain amino acids of the CPs and certain nucleotides of the RNA sequence.[26] Strong electrostatic attraction between the CP and the ssRNA viral genome have been shown to initiate the assembly of a number of viruses, as in the well-studied examples of the cowpea chlorotic mottle virus[27] or the brome mosaic virus,[28] where nonspecific electrostatic interactions between the negatively charged ssRNA and the positively charged arginine-rich motifs (ARMs) of the CP N-tails provide the thermodynamic driving force for the assembly.[29] Although electrostatic interactions are without any doubt among the fundamentals of the virus self-assembly process,[30] pure charge matching cannot explain all of the molecular details and specificities of the self-assembly process that are still not fully understood despite great advances in high-resolution structure determination of virus particles. There is thus an urgent need to understand the process of self-assembly and the precise role played in it by the viral RNA. High-level molecular modeling can make an important contribution to this endeavor, and accurate and realistic ab initio computations play a crucial role in unravelling the elusive nature of the specificity of the PSs in the protein−RNA interactions.

There presently exist few atomic-resolution simulations of viral capsids.[31−35] Currently, classical molecular dynamics (MD) based on empirical force fields has been the preferred method for studying RNA and proteins and has contributed greatly to the current level of understanding. However, classical simulations are as a rule based on fixed interaction potentials based on microscopic parameters such as the partial charge (PC) and the specific bonding geometry either calculated accurately on small fragments of local units or other criteria that are generally not strictly transferable. In this respect, details of the hydrogen bonding (HB) and quantification of HB-mediated interactions and their strengths at the specific binding sites are completely lacking. Such information is only accessible by full-scale quantum mechanical calculations on sufficiently large and often very complex structural models that require large computational resources. Detailed microscopic evaluations of the most relevant molecular interaction parameters are consequently an area recently witnessing rapid maturation due primarily to the availability of peta-scale supercomputing and the development of density functional theory (DFT)[36] for ab initio calculations. Nevertheless, this methodology is to date still not sufficiently efficient or accurate, and most computational studies have focused on the presumed geometric arrangements of the subunits[37] with their verification left to elusive experimental validation.

The present work is aimed to advance the ab initio quantum mechanical methodology in the context of advanced modeling of biomolecular assembly with the goal of understanding protein−RNA interactions in their simplest form, i.e., the assembly and packing of viral nucleo-components. We report the results of ab initio DFT calculations of the electronic structure and bonding using bacteriophage MS2 as a model system. Specifically, we considered the C-variant wild-type capsid protein−RNA complex (PDB ID: 1ZDH)[38] and its mutated form (PDB ID: 5MSF).[39] Because of computational limitations, these calculations are restricted to a single subunit of an asymmetrical unit of the virus, including an MS2 CP monomer and associated ssRNA. Still, this appears to be largest ab initio quantum computation performed on a complex biomolecular system to date. The high-quality quantitative results enable us to elucidate the molecular determinants of the

DOI: 10.1021/acs.jpcb.7b02569
J. Phys. Chem. B 2017, 121, 6321−6330

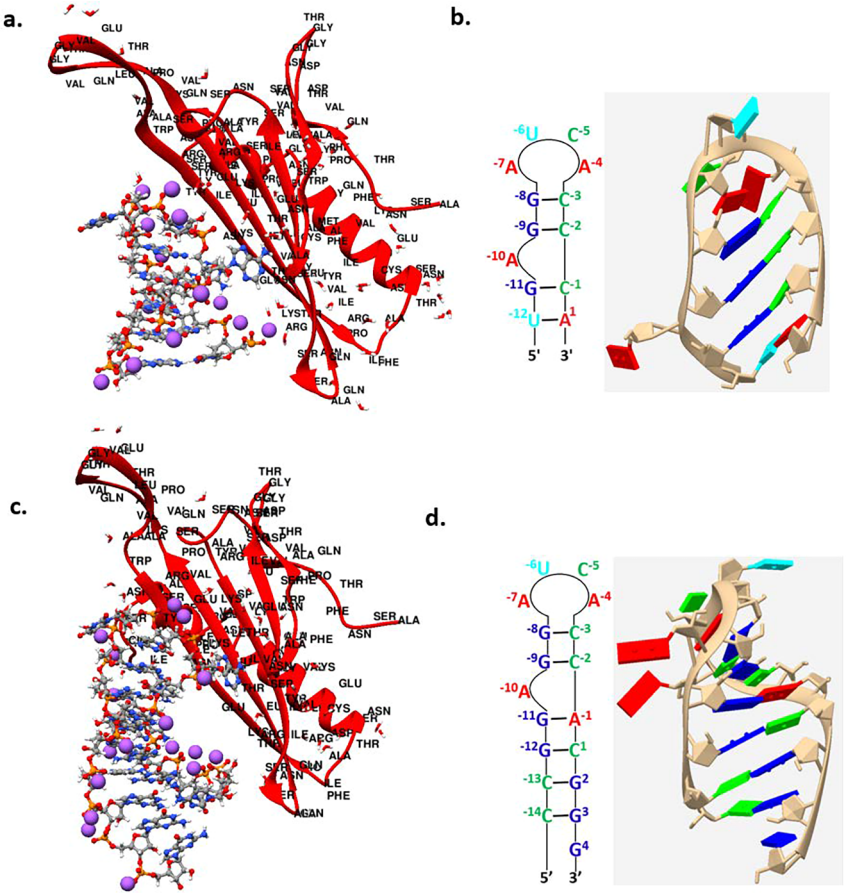


Figure 2. Relaxed structure of the MS2 CP−RNA complex with protein subunit A and RNA subunit R: (a) protein−RNA complex in 1ZDH, (b) structure and sequence of RNA in 1ZDH, (c) protein−RNA complex in 5MSF, and (d) structure and sequence of RNA in 5MSF. In (a) and (b), the red ribbon denotes the MS2 CP, the ball and stick denotes nucleotides of RNA, purple spheres denote Na, and stick represents the water molecules.

PS in viral assembly, allowing us to propose that they are based on the PC and the interfacial HB distributions. We address the issues of sequence-specific differences of CP−RNA complex formation, the distribution of partial charge, and the details of the interatomic HBs in the complex, providing much needed insights into the mode of action relevant to PSs in viral self-assembly through a purely fundamental computational route. Specifically, the difference between the two CP−RNA complexes studied stems from the differences in strength and number of HBs at the interface.

2. COMPUTATIONAL MODELING AND METHODS

2.1. Structural Modeling. The icosahedral viral capsid of MS2 with associated RNA stem-loops (biological unit) contains 60 identical units (called asymmetric units). An asymmetric unit consists of three identical MS2 CP chains (A, B, C) and two identical single-stranded RNA (ssRNA) chains (R, S). In our computational models, we have taken the MS2 CP chain 'A' and the single-stranded RNA R in both cases, which includes the loop binding site of RNA to protein. The models with subunits A/R and C/S would be computationally just as feasible, but we do not expect any fundamental differences and relegate the comparison of the details to a subsequent publication. The icosahedral viral capsid in the ribbon form is depicted in Figure 1(a) showing 60 copies of the subunits (Figure 1(b)) symmetrically replicated. We have considered two different structures of RNA stem-loops with the same loop motif, the highest affinity loop motif AUCA, with the same MS2 CP to investigate the sequence-specific CP−RNA interaction of this PS recognition motif. One structure is the C-variant wild-type stem-loop, and the other is the F5 aptamer (mutated type) structure. The C-variant wild-type MS2 operator has three unpaired adenines (−4 and −7 in the loop and −10 bulge in the stem) together with a pyrimidine (C^{-5}) (Figure 1(b)). The F5 aptamer−CP complex consists of a secondary RNA structure as a consequence of the "operator-like" conformation, which contains the non-Watson−Crick pair (G^{-11}-A^{1}) (Figure 2(b)), and the remaining base pairs are in the form of Watson−Crick pairs in the stem apart from a single overhanging 3′ guanine.[19] There are two important binding sites in the high-affinity PS TR (one in the loop and the other in bulge positions). Strictly speaking, a single ssRNA stem-loop and CP monomer within the asymmetric unit are not sufficient to include both sites, which require that two CP monomers forming a dimer in that unit be used to include both high affinity sites. Because of computational limitations, we have to restrict our two models to the CP monomer so that only the loop binding site (AUCA loop motif) can be probed. However, most PSs in the ensemble do not have the bulge binding site [A^{-10}] so that this restriction provides an appropriate and crucial scenario for the study of PSs overall.

The initial structures of the asymmetric unit of the C-variant wild-type and mutated type are taken from Protein data bank (PDB IDs: 1ZDH[38] and 5MSF,[39] respectively). The number and sequence of amino acids of MS2 CP are identical in 1ZDH and 5MSF. The 129 amino acids have the sequence:

DOI: 10.1021/acs.jpcb.7b02569
J. Phys. Chem. B 2017, 121, 6321−6330

ASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSR-SQAYKVTCSVRQSSAQNRKYTIKVEVPKVATQTVGGVEL-PVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN-PIPSAIAANSGIY. The sequences for nucleotides are UGAG-GAUCACCCA (13 bases) and CCGGAGGAUCACCACGGG" (18 bases) in 1ZDH and 5MSF, respectively. The specific residue K in the protein and A in the RNA sequence are the sites of the strongest HBs we identified, which will be discussed as follows. The PDB data also contain water molecules included in the calculations. To counterbalance the negatively charged phosphate (PO_4) group, the same number of Na atoms are added in the vicinity of each PO_4 group in accordance with the general scheme adopted for nucleobase biomolecules.[40−42] These initial structures for 1ZDH and 5MSF are then fully relaxed using the DFT-based package (VASP). They are used as input for the electronic structure calculation using another DFT-based method, the all-electron orthogonal linear combination of atomic orbital (OLCAO) method.[43] The relaxed structures are shown in Figure 2(a and c) and summarized in Table S1. It should be pointed out that the highly accurate ab initio DFT calculations for complex biomolecular systems with size of more than 2600 atoms are unprecedented and demand a huge amount of computational resources for quantitative analysis. There are other recent QM calculations on biomolecular systems with size approaching 1000 atoms.[44,45] However, there is a main difference from our work. These calculations took a smaller fragment of a larger system for full QM calculations with the rest treated by classical or semiclassical MM.

2.2. Structural Relaxation. After constructing the 1ZDH and 5MSF MS2 CP−RNA models, they are then fully relaxed using the density functional theory (DFT)[36,46]-based method. We have used the Vienna ab initio simulation package (VASP),[47] which has been highly effective for structure relaxation. We used the projector augmented wave (PAW) method with Perdew−Burke−Ernzerhof (PBE)[48] potential for the exchange correlation functional within the generalized gradient approximation (GGA). We employed a relatively high energy cutoff of 500 eV, and the electronic convergence criterion was set at 10^{-5} eV. The force convergence criteria for ionic relaxation was set at 10^{-3} eV/Å. We have used single k-point calculations because our models are in the form of large supercells; thus, a single k-point calculation at the zone center is sufficient. Similar structural relaxation for other large complex biomolecular systems has been successfully demonstrated in our other recent studies.[40,41,49] All VASP calculations were carried out at the National Energy Research Scientific Computing (NERSC) facility at Lawrence Berkeley Laboratory.

2.3. Electronic Structure Calculations. The ab initio OLCAO method[43] is used for electronic structure calculations for all of our models after VASP relaxation. There are many advantages of the OLCAO method, such as flexibility of the basis choice, lower computational cost, and ease of analysis using the Mullikan scheme. It is highly efficient for electronic structure calculations such as density of states, partial charge, and bonding properties of large complex biomolecules. The OLCAO is an all-electron method using local density approximation (LDA) of DFT. It employs Gaussian-type orbitals (GTO) for the atomic basis set. Depending on the nature of the investigation and the size of the model, three types of basis sets with different numbers of atomic orbitals can be used for the calculations. The minimum basis (MB) includes the core orbitals and the occupied or unoccupied orbitals in the valence shell. If additional empty orbitals of the next unoccupied shell are added to it, then this basis is referred to as the full basis (FB). In the present calculations, FB was used for the determination of the self-consistent potential. A minimum basis (MB) was used for the calculation of partial charge (PC) and bond order (BO) values. These data for the basis set are carefully constructed and well-tested for each atom within the database of the OLCAO package. More details can be found in ref 43. The combination of the VASP and OLCAO methods has been successfully employed in the study of many complex inorganic[50] and organic crystals[51] as well as biomolecules such as DNA,[40,41] collagen protein,[52,53] and a drug−DNA complex.[49]

The calculation of effective charge (Q^*) on each atom is a very important parameter for the partial charge distribution of a system. The deviation of Q^* or the charge transfer from the neutral atom (Q_0) is usually referred to as partial charge (PC) on that atom, or $\Delta Q = (Q_0 - Q^*)$ (i.e., $-\Delta Q$ = gain of electron or electronegative and $+\Delta Q$ = loss of electron or electro-positive). The Q^* on each atom in the molecule is calculated according to the Mulliken population analysis[54]

$$Q_\alpha^* = \sum_i \sum_{n_{occ}} \sum_{j,\beta} C_{i\alpha}^{*n} C_{j\beta}^{n} S_{i\alpha,j\beta} \tag{1}$$

where $C_{j\beta}^{n}$ are the eigenvector coefficients of the nth state, jth orbital, and βth atom. $S_{i\alpha,j\beta}$ are the corresponding overlap integrals. An accurate partial charge distribution of a molecule is an important ingredient for determining the intermolecular interaction potential. The partial charge on each amino acid and nucleotide can be obtained by adding the ΔQ values of all atoms within that group.

Another very important parameter is the bond order (BO) values $\rho_{\alpha\beta}$ for every pair of atoms. The precise quantification of bonding characteristics based upon quantum mechanical calculations and their relationship with electronic structure can then serve as a platform for understanding the structure of complex biomolecules. The total bond order (TBO) is the cumulative BO from all unique bond pairs among the constituent structural groups. The bond order quantifies the relative strengths of all types of bonds and generally scales with the bond length (BL) but also depends on the local environment of the bonding atoms. The BO values (in unit of electrons) for each pair of atoms α and β are calculated according to

$$\rho_{\alpha\beta} = \sum_{n_{occ}} \sum_{i,j} C_{i\alpha}^{*n} C_{j\beta}^{n} S_{i\alpha,j\beta} \tag{2}$$

We specifically explore the hydrogen bonding between the MS2 coat protein and the ssRNA that has not previously been analyzed. It should be pointed out that the calculations of PC and BO according to eqs 1 and 2, using the Mulliken scheme, are basis dependent. In the present study, as well as in many recent studies mentioned above, the same well-tested MB has been used for both models. Although there are other more accurate and elaborate methods for calculating PC or BO, they are by necessity limited only to small molecules with simpler structures. For the present large biomolecular system, our methods that provide the PC and BO values with accuracies up to two-to-three decimal points are sufficiently accurate for the proper quantitative description.

DOI: 10.1021/acs.jpcb.7b02569
J. Phys. Chem. B 2017, 121, 6321−6330

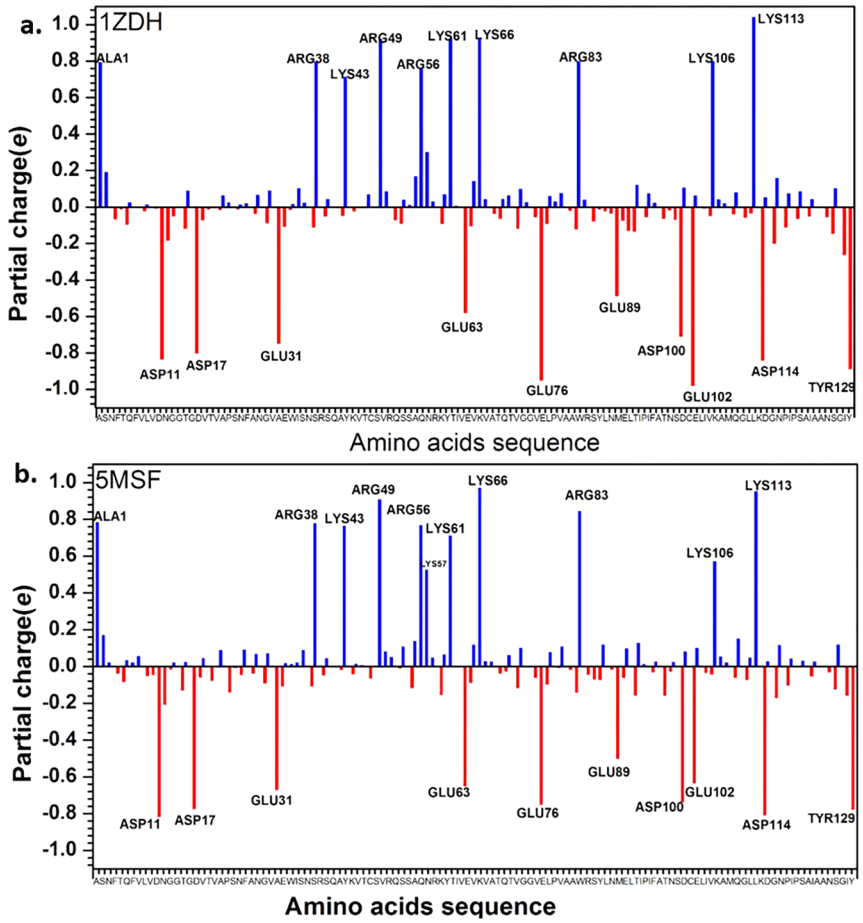


Figure 3. Calculated partial charge (PC) distribution in MS2 CP of (a) 1ZDH and (b) 5MSF.

3. RESULTS

3.1. Partial Charge (PC) Distribution. The PC on each atom in the two structural models are calculated using OLCAO.[43] The PCs for all of the atoms in the 1ZDH and 5MSF CP−RNA complexes are shown in Figure S1, and those resolved into each amino acid or nucleobase are obtained by adding the atomic PC in each unit shown in Figure S2, respectively. In Figure 3(a and b), we display the distribution of PCs for amino acid sequences in CPs in 1ZDH and 5MSF, respectively. It can be seen that the gross features are the same because they have identical sequences, but there are some important minor differences due to the different RNAs to which they bind. Most of the protein PCs stem from the canonical positively charged ARG and LYS and negatively charged ASP and GLU. The terminal amino acids ALA1 and TYR129 are also highly electropositive and -negative, respectively. The calculated PCs on different structural components are listed in Table 1. Obviously, the overall PC for CP is positive and that for ssRNA is negative. In the protein−ssRNA complex, water molecules gain some charge, whereas Na^+ ions lose charge mostly due to the negatively charged backbone PO_4 of the RNA as charge compensators. The total PC on the coat protein in the mutated type (5MSF) is more electropositive than in the wild-type (1ZDH). Water molecules in 5MSF have a larger negative PC than in 1ZDH even though they are fewer in number. The total PCs of RNA and Na^+ ions in 5MSF are higher because it contains more nucleotides and Na^+ ions.

Table 1. Total Partial Charge in Different Structural Sections of the MS2 CP−RNA Complex

model	coat protein	ssRNA	water	Na^+ ions
1ZDH	0.3013	−10.7660	−0.1533	10.6183
5MSF	0.6788	−14.6928	−0.1775	14.1928

To obtain more detailed insight into the difference between 1ZDH and 5MSF in their interactions between CP and RNA, we compare the PC distributions side by side for the protein in Figure 4(a and b) and nucleotides in Figure 4(c). In the protein, the sequence of the residues are the same, whereas in RNA, the arrangements for nucleotide bases differ. There are discernible differences in the PCs of the protein sequences between the mutated and C-variant wild-type cases with some residues actually changing sign (see Figure S2, these amino acids are marked green). We believe that the changes in the PC on the same amino acids in going from the C-variant wild-type to the mutated one is due to their different conformations and/or the presence of vicinal water molecules. For example, the residues SER51 and ASN87 change sign of their PCs and additionally form interfacial HBs (see below). The PCs of nucleotides in RNA are all negative in both cases but with different magnitudes. These differences are indicators of the packaging signal recognition motifs and will be discussed in detail later. The PC distribution plotted on the solvent-excluded surfaces for 1ZDH and 5MSF are displayed in Figure 5. This may be the first time such color-coded maps are displayed based on actual quantitative data for the PC and not

DOI: 10.1021/acs.jpcb.7b02569
J. Phys. Chem. B 2017, 121, 6321−6330

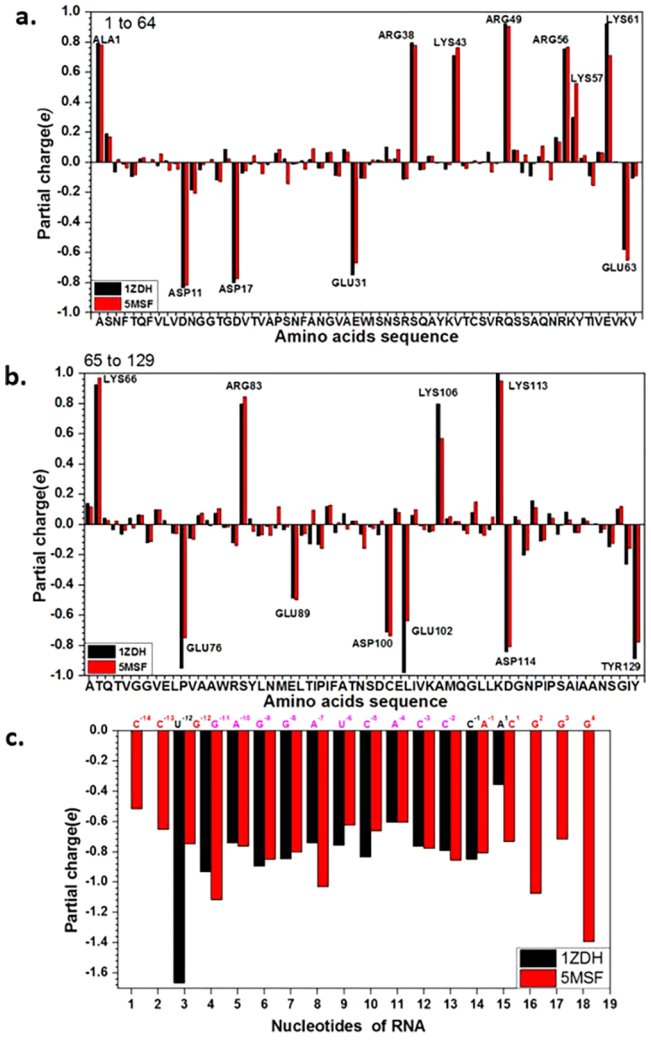


Figure 4. Comparison of partial charge in (a) MS2 CP in amino acid sequence 1−64, (b) MS2 CP in amino acid sequence 65−129, and (c) nucleotides in the RNA. In (c), those nucleotides existing only in 1ZDH are denoted in black and only in 5MSF in red, and common nucleotides are denoted in pink.

on the perceived charge or the charge inferred from experiments for different structural units, as routinely displayed in the literature for large complex biomolecules. We also note that there are differences in the PC distribution in the interfacial region of the protein and the RNA.

3.2. Interfacial Hydrogen Bond (HB) Distribution. It is well-known that HB holds the key to understanding many intriguing phenomena in biomolecular systems. Unfortunately, most of these explanations are based on the structural data of HB lengths and their locations without any quantitative information on the HB strength, which depends not just on the separation between H and anions (O or N) but also on their local environments. The existence of CP−RNA HB has been suggested based on the close contact between CP and RNA from high-resolution crystal structures determined experimentally.[55,56] We have obtained quantitative information for all HBs using the ab initio computational approach (see Table S2). The results for HBs at the interfacial region between the protein and RNA for both 1ZDH and 5MSF are shown in Figure 6 in the form of a bond order (BO) vs bond length (BL) plot. There are mainly two RNA nucleotides in the RNA loop (A^{-4}, C^{-5}) participating in interfacial HBs for both models, consistent with these sites expected to be the critical features of the PS loop recognition motif.[15] To a lesser extent, the HBs in the nucleotide at stem C^{-13} in 5MSF and U^{-12} in 1ZDH are also involved. Figure 6 shows 11 HBs for 1ZDH and 16 HBs for 5MSF up to an HB distance of 3 Å. The total bond order (TBO) values for HBs are 0.30 and 0.48 for 1ZDH and 5MSF, respectively; thus, not only is the number of HBs increased in 5MSF, but the TBO value is also increased by 55%. The strongest HB (LYS43-A^{-4}) in the C-variant wild-type with a BO value of 0.094 e^- and BL of 1.53 Å becomes much stronger in the mutated case with a BO value of 0.125 e^- and BL of 1.47 Å. This exceptionally strong HB is formed between an H atom from the highly electropositive amino acid LYS43 and the O atom of a negatively charged PO_4 group in the nucleotide A^{-4}. The strength of this HB also correlates with charge on the specific amino acid because the PC of LYS43 is more electropositive in 5MSF than in 1ZDH. Moreover, new HBs of considerable strength are formed (ARG49-C^{-13}, TYR85-C^{-5}, and LYS61-C^{-5}) in the mutated 5MSF. The amino acids GLU63, LYS61, and ASN87 form HBs with the C^{-5} nucleotide in both 1ZDH and 5MSF, but only TYR85 in the mutated type forms an HB with C^{-5}.

The mutation-dependent strengthening of the HB network at specific sites at the interfacial region demonstrates the molecular underpinning of the variation of the PS affinity for CP across the PS ensemble. Indeed, RNA and its sequence and

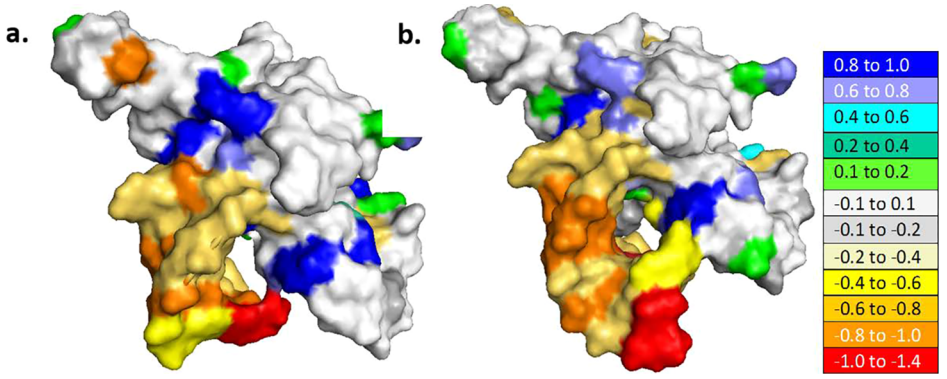


Figure 5. Partial charge distribution with solvent accessible surface on the MS2 CP−RNA complex (a) 1ZDH and (b) 5MSF. Both models represent chain A of the coat protein and chain R of RNA in the asymmetrical unit. The color bar on the side indicates the averaged partial charges from red to green to blue (RGB).

DOI: 10.1021/acs.jpcb.7b02569
J. Phys. Chem. B 2017, 121, 6321−6330

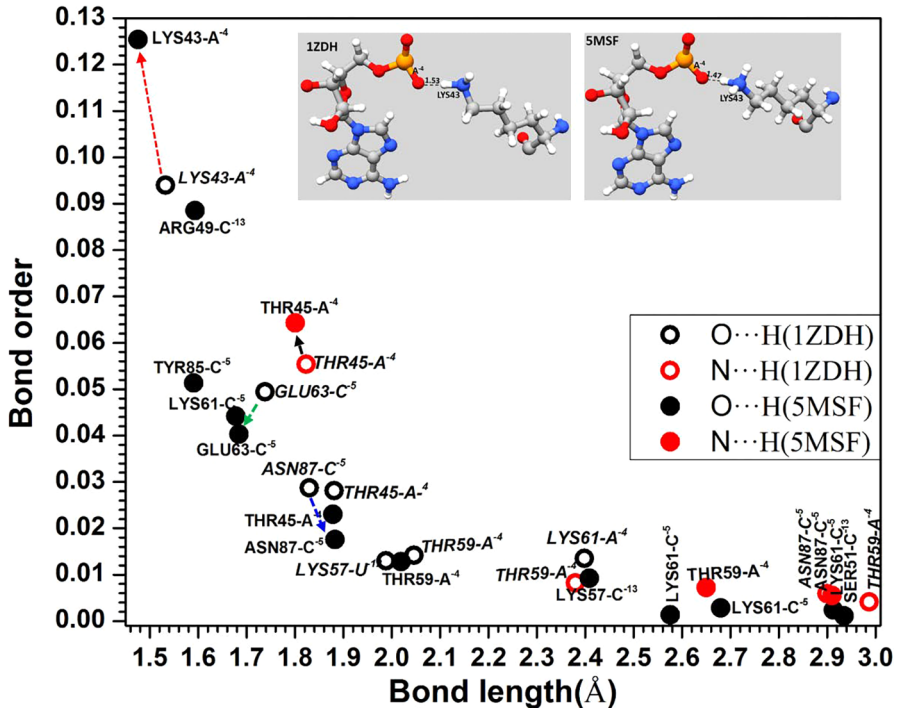


Figure 6. Hydrogen bonding distribution at the interface of MS2 CP and RNA. The open circle and label with italic text represent 1ZDH, and the closed circle and label with normal text represent 5MSF.

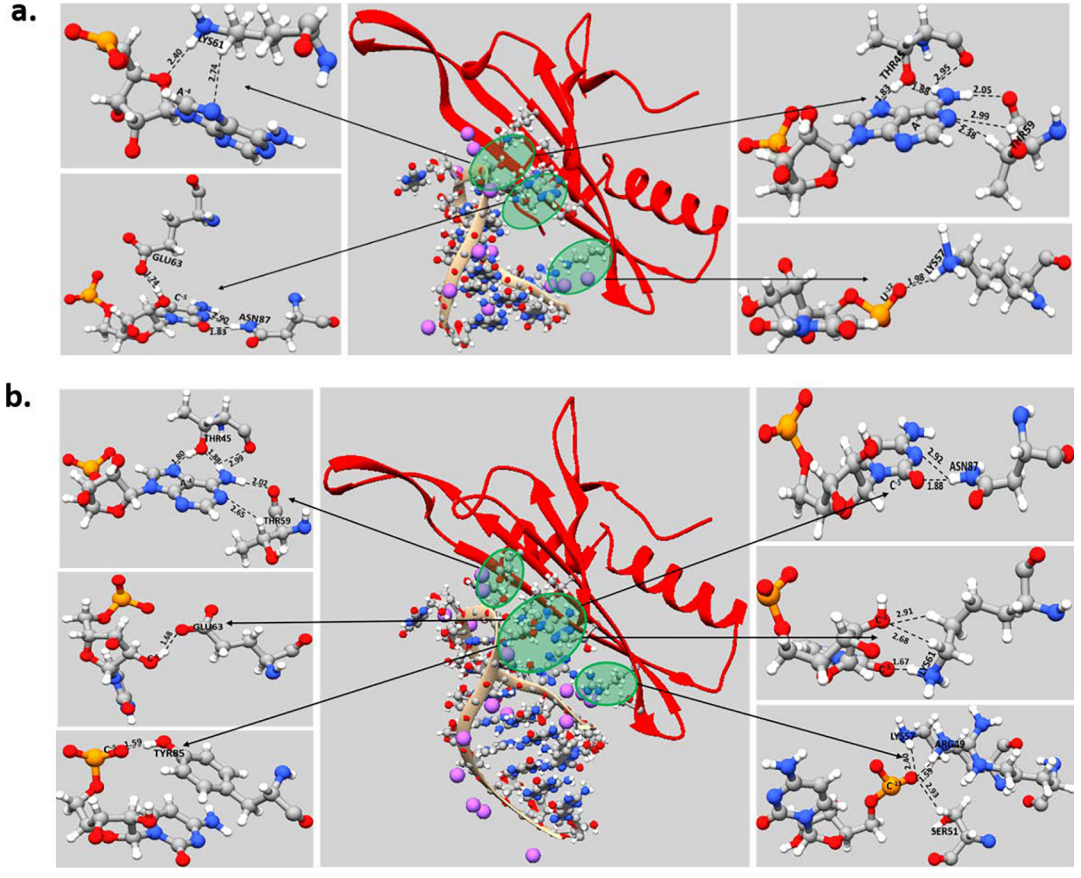


Figure 7. Hydrogen bonding network at the interface of MS2 CP and RNA of (a) 1ZDH and (b) 5MSF.

fold play critical roles in defining the PSs. To gain additional insights into the actual HB distribution in the interfacial region of the mutated (5MSF) and the C-variant wild-type (1ZDH) cases, we show in more detail the local geometry and residues associated with the key HBs in Figure 7(a) for 1ZDH and (b) for 5MSF, respectively. It can also be seen that, in 1ZDH, a

DOI: 10.1021/acs.jpcb.7b02569
J. Phys. Chem. B 2017, 121, 6321−6330

water molecule is present at the interface, which makes an HB with the nucleotide U^{-6} and amino acid TYR85 (see Figure S3), but no water molecule is visible in this region for 5MSF. All HBs are formed between amino acids of the CP and nucleobases of the RNA; they are indicated by dashed lines with the distances of separation marked. These HBs at the interface in the two structures, including their strength (BO) values up to a separation of 3.0 Å, are listed in Table S2.

4. DISCUSSION

Computationally based analysis of the interaction between the CP and RNA stem-loop in the MS2 phage provides an alternative route to the purely bioinformatics/experimental approach for the elucidation of molecular mechanisms involved in viral self-assembly capable of identifying molecular details of the RNA−CP interactions. It can provide more detailed, but most importantly quantitative, results not yet available by purely experimental protocols. In particular, it enables a comparative analysis of different variants of the PS recognition motif and a better understanding of the impact of mutations in the PS stem on the binding sites. We have identified strong hydrogen bonds (corresponding to the binding sites at the atomic level) in two representative variants of the MS2 PSs by means of accurate DFT[36] computations of their electronic structures. The key message is the enhancement of the HB between LYS43 in the CP and A^{-4} of the nucleotide at the 3′ end of the terminal loop and the formation of more interfacial HBs due to a mutation in the stem sequence.

Over the years, the notion that nucleotide C^{-5} in RNA plays a vital role has become well-established.[57] Our calculations indeed show that the C^{-5} residue makes stronger and more numerous HBs with the CP in both models, confirming the importance of C^{-5} for overall binding.[57] This is consistent with the observation that 16 PSs in the ensemble indeed have C^{-5} in that position.[15] The important role of HBs between the protein and an RNA hairpin is also consistent with the electron density map analysis of a similar MS2 CP−RNA system based on the data from high-resolution X-ray diffraction,[58] where specific HBs between CP amino acids and the nucleobases were proposed but not backed up by quantitative calculations. Similar studies on potential HBs at the RNA−protein interface based on high-resolution crystal structure analysis have been reported recently.[59]

The critical role played by the interfacial HBs in response to variations in the genomic sequences of the PSs is probably not limited to the case of MS2, and further investigations on other systems with putative PSs are needed for a firmer conclusion. Interestingly, the role of water molecules appears to be minimal in the case studied here because they are not in the vicinity of the interface where the strong HBs are formed. In principle, the presence of a solvent with dissolved salt ions could play a role in the overall environmental solvent effect.[60] Most of the current theoretical/computational research on CPs of RNA viruses are limited to coarse-grained models, focusing on the role of electrostatic interactions.[30,34,61] Such studies may provide a great deal of insight on the mechanism of virus self-assembly but cannot pinpoint the specific nature of packaging signals. In contrast, large-scale ab initio calculations of the electronic structure and interatomic bonding, especially HBs, reveal much more detailed information for the actual interaction at the CP−RNA interface by providing quantitative bond order values to characterize the strength of the bonds and their locations. The ability to obtain PCs on each atom in the model (see Figure S1) enables us to provide the PC distribution on each amino acid and nucleotide unit with great detail.

Quantitative results with sequence-specific PS information and detailed HB and PC distributions can certainly push forward the frontier of understanding for the fundamental mechanisms of PS action in viral research, enabling a mode-detailed connection between the details of the microscopic bonding properties and the mesoscopic theories of equilibrium (thermodynamic) and nonequilibrium (kinetic) phenomena in virus assembly. Our results pave the way for a better understanding of how subtle variations around the core PS recognition motif impact the affinity of the PS for CP, which in turn plays a crucial role in capsid assembly.[10] Once the interfacial hydrogen bonding between CP and RNA subunits is fully understood, the fundamental forces involving electrostatic, vdW, and steric interactions can be better characterized to develop a more detailed and nuanced approach to the problem of capsid assembly. To explore the universality of the PS mechanism, it is also desirable to extend the current investigation to other cases such as Satellite Tobacco Necrosis virus (SVNT)[62] or cowpea mosaic virus (CPMV),[63] where more experimental data and PS predictions can be compared with computations. Ideally, one could also envision an ab initio prediction of the candidate PS sites along the genome chosen by their bonding properties with the different CP regions and then test them more specifically either experimentally or with more detailed simulation studies.

5. CONCLUSIONS

We have identified the CP−RNA interactions in the MS2 phage by using ab initio quantum mechanical calculations, demonstrating the vital role of HBs at the interface between RNA and protein for virus assembly. Specific conclusions obtained are as follows: 1) The stem-loops studied here, which are representative of the MS2 PSs and share its core loop recognition motif XXYA (X denoting any nucleotide and Y a pyrimidine), are given by interactions between specific residues of the CP and nucleotides in the RNA identified with LYS43-A^{-4}, TYR85-C^{-5}, and LYS61-C^{-5} bonds in the loop area of the PS and a further interaction of ARG49-C^{-13} in its stem. 2) The protein is electropositive and the RNA is electronegative in the MS2 CP−RNA complex. ARG and LYS are highly electropositive and ASP and GLU are highly electronegative while the terminal amino acids ALA1 (+) and TYR129 (−) are oppositely charged, consistent with their known behavior.[60] 3) Strong hydrogen bonds exist at the interface between the CP and the RNA, and there are more HBs at the interface in the mutated case than in the wild-type case. 4) The total bond order value of HBs at the interface in 5MSF is much higher than in 1ZDH. Furthermore, this work also opens the door for systematic analysis of other complex biological systems such as protein−RNA, DNA−protein, protein−protein, and drug−protein systems undergoing specific mutations or exhibiting a natural sequence variation around a core recognition motif similar to the MS2 phage PSs, thereby providing additional information on the structure−function relationships in virus assembly. Finally, CP−RNA interactions are a dynamic process, and the transient interaction involving a conformational change cannot be easily explained by crystal structure information alone. Our methodology paves the way for a new route based on ab initio computations to understand the details of the binding properties in the CP−RNA nucleoprotein complex.

DOI: 10.1021/acs.jpcb.7b02569
J. Phys. Chem. B 2017, 121, 6321−6330

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.7b02569.

Summary of relaxed structure parameters for models 1ZDH and 5MSF, interfacial hydrogen bonding (<3.0 Å) of the MS2 coat protein and ssRNA for models, calculated atomic partial charge distributions on MS2 CP−RNA complexes for models, better display of comparison of PC distributions in MS2 coat proteins for models, schematic for interfacial HB contributed by water in 1ZDH, calculated BO distributions of HBs in 1ZDH and 5MSF (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: chingw@umkc.edu.

ORCID

Wai-Yim Ching: 0000-0001-7738-8822

Author Contributions

W.C., N.S., and R.P. initiated the project, and L.P. performed the calculations. R.T. provided crucial insights on PS-mediated virus assembly. All of the authors participated the discussion of the results. L.P. and W.C. wrote the paper. All authors edited and proofread the final manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported by US DOE-Office of BES, Division of Materials Science and Engineering under Grants DE-SC008176 and DE-SC8068. R.T. thanks the Wellcome Trust for funding via an investigator award. L.P. was supported by a research grant from the School of Graduate Studies at UMKC. This research used the resources of NERSC supported by the Office of Science of DOE under Contract DE-AC03-76SF00098 and the computing resources of Bioconsortium of the University of Missouri.

■ REFERENCES

(1) He, X.; Liu, L. Toward a prospective molecular evolution. *Science* **2016**, *352*, 769.

(2) Li, C.; Qian, W.; Maclean, C. J.; Zhang, J. The fitness landscape of a tRNA gene. *Science* **2016**, *352*, 837.

(3) Puchta, O.; Cseke, B.; Czaja, H.; Tollervey, D.; Sanguinetti, G.; Kudla, G. Network of epistatic interactions within a yeast snoRNA. *Science* **2016**, *352*, 840.

(4) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **2016**, *353*, 389.

(5) Bykhovski, A.; Zhang, W.; Jensen, J.; Woolard, D. Analysis of Electronic Structure, Binding, and Vibrations in Biotin−Streptavidin Complexes Based on Density Functional Theory and Molecular Mechanics. *J. Phys. Chem. B* **2013**, *117*, 25.

(6) Hirano, Y.; Takeda, K.; Miki, K. Charge-density analysis of an iron−sulfur protein at an ultra-high resolution of 0.48 Å. *Nature* **2016**, *534*, 281.

(7) Lorman, V.; Rochal, S. Landau theory of crystallization and the capsid structures of small icosahedral viruses. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 224109.

(8) Caspar, D. L.; Klug, A. In *Cold Spring Harbor symposia on quantitative biology*; Cold Spring Harbor Laboratory Press, 1962; Vol. *27*, p 1.

(9) Dykeman, E. C.; Stockley, P. G.; Twarock, R. Building a viral capsid in the presence of genomic RNA. *Phys. Rev. E* **2013**, *87*, 022717.

(10) Dykeman, E. C.; Stockley, P. G.; Twarock, R. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 5361.

(11) Patel, N.; Dykeman, E. C.; Coutts, R. H.; Lomonossoff, G. P.; Rowlands, D. J.; Phillips, S. E.; Ranson, N.; Twarock, R.; Tuma, R.; Stockley, P. G. Revealing the density of encoded functions in a viral RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2227.

(12) Simon, A. E.; Gehrke, L. RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2009**, *1789*, 571.

(13) Dykeman, E.; Stockley, P.; Twarock, R. Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. *J. Mol. Biol.* **2010**, *395*, 916.

(14) Dykeman, E. C.; Twarock, R. All-atom normal-mode analysis reveals an RNA-induced allostery in a bacteriophage coat protein. *Phys. Rev. E* **2010**, *81*, 031908.

(15) Dykeman, E. C.; Stockley, P. G.; Twarock, R. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *J. Mol. Biol.* **2013**, *425*, 3235.

(16) Rolfsson, Ó.; Middleton, S.; Manfield, I. W.; White, S. J.; Fan, B.; Vaughan, R.; Ranson, N. A.; Dykeman, E.; Twarock, R.; Ford, J. Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. *J. Mol. Biol.* **2016**, *428*, 431.

(17) Borodavka, A.; Tuma, R.; Stockley, P. G. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 15769.

(18) Stockley, P. G.; Ranson, N. A.; Twarock, R. A new paradigm for the roles of the genome in ssRNA viruses. *Future Virol.* **2013**, *8*, 531.

(19) Stockley, P. G.; Twarock, R.; Bakker, S. E.; Barker, A. M.; Borodavka, A.; Dykeman, E.; Ford, R. J.; Pearson, A. R.; Phillips, S. E.; Ranson, N. A. Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. *J. Biol. Phys.* **2013**, *39*, 277.

(20) Stray, S. J.; Johnson, J. M.; Kopek, B. G.; Zlotnick, A. An in vitro fluorescence screen to identify antivirals that disrupt hepatitis B virus capsid assembly. *Nat. Biotechnol.* **2006**, *24*, 358.

(21) Bittner, A. M.; Alonso, J. M.; Górzny, M. Ł.; Wege, C. In *Structure and Physics of Viruses*; Springer: 2013; p 667.

(22) Hesketh, E. L.; Meshcheriakova, Y.; Dent, K. C.; Saxena, P.; Thompson, R. F.; Cockburn, J. J.; Lomonossoff, G. P.; Ranson, N. A. Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. *Nat. Commun.* **2015**, *6*, 10113.

(23) Keef, T.; Twarock, R. Affine extensions of the icosahedral group with applications to the three-dimensional organisation of simple viruses. *Journal of mathematical biology* **2009**, *59*, 287.

(24) Twarock, R. A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *J. Theor. Biol.* **2004**, *226*, 477.

(25) Chen, X.; Gonçalves, M. A. Engineered viruses as genome editing devices. *Mol. Ther.* **2016**, *24*, 447.

(26) Garmann, R. F.; Comas-Garcia, M.; Knobler, C. M.; Gelbart, W. M. Physical Principles in the Self-Assembly of a Simple Spherical Virus. *Acc. Chem. Res.* **2016**, *49*, 48.

(27) Garmann, R. F.; Comas-Garcia, M.; Koay, M. S.; Cornelissen, J. J.; Knobler, C. M.; Gelbart, W. M. Role of electrostatics in the assembly pathway of a single-stranded RNA virus. *J. Virol.* **2014**, *88*, 10472.

(28) Ni, P.; Wang, Z.; Ma, X.; Das, N. C.; Sokol, P.; Chiu, W.; Dragnea, B.; Hagan, M.; Kao, C. C. An examination of the electrostatic interactions between the N-terminal tail of the brome mosaic virus coat protein and encapsidated RNAs. *J. Mol. Biol.* **2012**, *419*, 284.

(29) Belyi, V. A.; Muthukumar, M. Electrostatic origin of the genome packing in viruses. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 17174.

(30) Šiber, A.; Božič, A. L.; Podgornik, R. Energies and pressures in viruses: contribution of nonspecific electrostatic interactions. *Phys. Chem. Chem. Phys.* **2012**, *14*, 3746.
(31) Andoh, Y.; Yoshii, N.; Yamada, A.; Fujimoto, K.; Kojima, H.; Mizutani, K.; Nakagawa, A.; Nomoto, A.; Okazaki, S. All-atom molecular dynamics calculation study of entire poliovirus empty capsids in solution. *J. Chem. Phys.* **2014**, *141*, 165101.
(32) Arkhipov, A.; Freddolino, P. L.; Schulten, K. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure* **2006**, *14*, 1767.
(33) Hagan, M. F.; Zandi, R. Recent advances in coarse-grained modeling of virus assembly. *Curr. Opin. Virol.* **2016**, *18*, 36.
(34) Zhang, D.; Konecny, R.; Baker, N. A.; McCammon, J. A. Electrostatic interaction between RNA and protein capsid in cowpea chlorotic mottle virus simulated by a coarse-grain RNA model and a Monte Carlo approach. *Biopolymers* **2004**, *75*, 325.
(35) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **2013**, *497*, 643.
(36) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864.
(37) Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P.-S.; Ravichandran, R. Design of a hyperstable 60-subunit protein icosahedron. *Nature* **2016**, *535*, 136.
(38) Valegård, K.; Murray, J. B.; Stonehouse, N. J.; van den Worm, S.; Stockley, P. G.; Liljas, L. The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J. Mol. Biol.* **1997**, *270*, 724.
(39) Rowsell, S.; Stonehouse, N. J.; Convery, M. A.; Adams, C. J.; Ellington, A. D.; Hirao, I.; Peabody, D. S.; Stockley, P. G.; Phillips, S. E. Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nat. Struct. Biol.* **1998**, *5*, 970.
(40) Poudel, L.; Rulis, P.; Liang, L.; Ching, W. Electronic structure, stacking energy, partial charge, and hydrogen bonding in four periodic B-DNA models. *Phys. Rev. E* **2014**, *90*, 022705.
(41) Poudel, L.; Steinmetz, N. F.; French, R.; Parsegian, A. V.; Podgornik, R.; Ching, W.-Y. Implication of solvent effect, metal ions and topology in the electronic structure and hydrogen bonding of human telomeric G-quadruplex DNA. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21573.
(42) Taniguchi, M.; Kawai, T. Electronic structures of A-and B-type DNA crystals. *Phys. Rev. E* **2004**, *70*, 011913.
(43) Ching, W.-Y.; Rulis, P. *Electronic Structure Methods for Complex Materials: The orthogonalized linear combination of atomic orbitals*; Oxford University Press: Oxford, U.K., 2012.
(44) Hu, L.; Soderhjelm, P.; Ryde, U. Accurate reaction energies in proteins obtained by combining QM/MM and large QM calculations. *J. Chem. Theory Comput.* **2013**, *9*, 640.
(45) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. How large should the QM region be in QM/MM calculations? The case of catechol O-methyltransferase. *J. Phys. Chem. B* **2016**, *120*, 11381.
(46) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133.
(47) Kresse, G.; Furthmüller, J. Software VASP, vienna (1999). *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 169.
(48) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(49) Poudel, L.; Wen, A. M.; French, R. H.; Parsegian, V. A.; Podgornik, R.; Steinmetz, N. F.; Ching, W. Y. Electronic Structure and Partial Charge Distribution of Doxorubicin in Different Molecular Environments. *ChemPhysChem* **2015**, *16*, 1451.
(50) Ching, W. Theoretical studies of the electronic properties of ceramic materials. *J. Am. Ceram. Soc.* **1990**, *73*, 3135.
(51) Liang, L.; Rulis, P.; Kahr, B.; Ching, W. Y. Theoretical study of the large linear dichroism of herapathite. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2009**, *80*, 235132.
(52) Eifler, J.; Podgornik, R.; Steinmetz, N. F.; French, R. H.; Parsegian, V. A.; Ching, W. Y. Charge distribution and hydrogen bonding of a collagen α2-chain in vacuum, hydrated, neutral, and charged structural models. *Int. J. Quantum Chem.* **2016**, *116*, 681.
(53) Adhikari, P.; Wen, A. M.; French, R. H.; Parsegian, V. A.; Steinmetz, N. F.; Podgornik, R.; Ching, W.-Y. Electronic Structure, Dielectric Response, and Surface Charge Distribution of RGD (1FUV) Peptide. *Sci. Rep.* **2015**, *4*, 1.
(54) Mulliken, R. S. Electronic population analysis on LCAO−MO molecular wave functions. I. *J. Chem. Phys.* **1955**, *23*, 1833.
(55) Convery, M. A.; Rowsell, S.; Storehouse, N. J.; Ellington, A. D.; Hirao, I.; Murray, J. B.; Peabody, D. S.; Phillips, S. E.; Stockley, P. G. Crystal structure of an RNA aptamer−protein complex at 2.8 Å resolution. *Nat. Struct. Biol.* **1998**, *5*, 133.
(56) Horn, W. T.; Convery, M. A.; Stonehouse, N. J.; Adams, C. J.; Liljas, L.; Phillips, S. E.; Stockley, P. G. The crystal structure of a high affinity RNA stem-loop complexed with the bacteriophage MS2 capsid: further challenges in the modeling of ligand−RNA interactions. *RNA* **2004**, *10*, 1776.
(57) Lowary, P. T.; Uhlenbeck, O. C. An RNA mutation that increases the affinity of an RNA−protein interaction. *Nucleic Acids Res.* **1987**, *15*, 10483.
(58) Grahn, E.; Stonehouse, N. J.; Adams, C. J.; Fridborg, K.; Beigelman, L.; Matulic-Adamic, J.; Warriner, S. L.; Stockley, P. G.; Liljas, L. Deletion of a single hydrogen bonding atom from the MS2 RNA operator leads to dramatic rearrangements at the RNA−coat protein interface. *Nucleic acids research* **2000**, *28*, 4611.
(59) Chao, J. A.; Patskovsky, Y.; Almo, S. C.; Singer, R. H. Structural basis for the coevolution of a viral RNA−protein complex. *Nat. Struct. Mol. Biol.* **2008**, *15*, 103.
(60) Nap, R. J.; Božič, A. L.; Szleifer, I.; Podgornik, R. The role of solution conditions in the bacteriophage PP7 capsid charge regulation. *Biophys. J.* **2014**, *107*, 1970.
(61) Forrey, C.; Muthukumar, M. Electrostatics of capsid-induced viral RNA organization. *J. Chem. Phys.* **2009**, *131*, 105101.
(62) Jones, T. A.; Liljas, L. Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution. *J. Mol. Biol.* **1984**, *177*, 735.
(63) Lin, T.; Chen, Z.; Usha, R.; Stauffacher, C. V.; Dai, J.-B.; Schmidt, T.; Johnson, J. E. The refined crystal structure of cowpea mosaic virus at 2.8 Å resolution. *Virology* **1999**, *265*, 20.

DOI: 10.1021/acs.jpcb.7b02569
J. Phys. Chem. B 2017, 121, 6321−6330