Electrostatic Regulation of Genome Packaging in Human Hepatitis B Virus

Tao Jiang,[†] Zhen-Gang Wang,[‡] and Jianzhong Wu[†*]
[†]Department of Chemical and Environmental Engineering, University of California, Riverside, California; and [‡]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California

ABSTRACT Hepatitis B virus (HBV) is a contagious human pathogen causing liver diseases such as cirrhosis and hepatocellular carcinoma. An essential step during HBV replication is packaging of a pregenomic (pg) RNA within the capsid of core antigens (HBcAgs) that each contains a flexible C-terminal tail rich in arginine residues. Mutagenesis experiments suggest that pgRNA encapsidation hinges on its strong electrostatic interaction with oppositely charged C-terminal tails of the HBcAgs, and that the net charge of the capsid and C-terminal tails determines the genome size and nucleocapsid stability. Here, we elucidate the biophysical basis for electrostatic regulation of pgRNA packaging in HBV by using a coarse-grained molecular model that explicitly accounts for all nonspecific interactions among key components within the nucleocapsid. We find that for mutants with variant C-terminal length, an optimal genome size minimizes an appropriately defined thermodynamic free energy. The thermodynamic driving force of RNA packaging arises from a combination of electrostatic interactions and molecular excluded-volume effects. The theoretical predictions of the RNA length and nucleocapsid internal structure are in good agreement with available experiments for the wild-type HBV and mutants with truncated HBcAg C-termini.

INTRODUCTION

Hepatitis B virus (HBV) is a pararetrovirus responsible for up to 80% of all cases of hepatocellular carcinoma worldwide, second only to tobacco among known human carcinogens (1). Although the initial viral infection is now preventable through vaccination, the HBV vaccine has no effect on the chronically infected population that is at high risk of developing liver cancer. Current drugs for treating HBV target only a specific motif of the viral genome or the reverse transcriptase. Such treatments are often plagued by constant mutations of the viral genome in a changing environment. Although remarkable progress has been made in understanding the natural history and pathogenesis of HBV, little is known about the molecular details of regulating the viral replication (2).

HBV replication entails a cascade of self-assembly processes that depend on various specific and nonspecific intermolecular interactions among a large number of biomacromolecules within the infectious virion and their interactions with the host cellular environment (3,4). A central step during the formation of HBV nucleocapsid is spontaneous coassembly of a pregenomic (pg) RNA with multiple copies of the core antigens (HBcAgs). The self-assembly process hinges on strong electrostatic interactions between the pgRNA and oppositely charged ''protamine tails'' at the C-termini of the capsid or core proteins. Previous mutagenesis experiments have shown that without the protamine tails, HBV capsids fail to upload significant amounts of RNA; and mutations at the C-terminal region of HBcAgs result in partial encapsidation of the viral genome (5–8). Although pgRNA in the wild-type (WT) HBV consists of 3500 nucleotides (3.5 kb), mutants with truncated protamine tails encapsidate only spliced viral RNA that have much shorter lengths.

Partial encapsidation of the pgRNA in HBV mutants suggests the existence of a balance between the genome size and the net charge of the capsid proteins in a stable nucleocapsid (8). The nonspecific nature of electrostatic interactions implies that charge regulation may be universally applicable to ssRNA/ssDNA viruses including retroviruses and retrotransposons (9,10). Despite its broad biophysical implications, the theoretical basis of electrostatic regulation for nucleocapsid formation remains obscure. For viral replication in vivo, electrostatic interactions involve not only nucleic acids and proteins but also small ions that are ubiquitous in a cellular environment. In addition, viral capsids often carry a significant amount of net charge, which makes any conclusion based solely on the charge neutrality of macromolecules (10) rather problematic. From a theoretical point of view, electrostatic interactions must work in combination with the molecular excluded-volume effects and other nonspecific intermolecular interactions. A faithful description of electrostatic interactions therefore must also account for all these factors.

In this work, we examine the thermodynamic basis in electrostatic regulation of RNA packaging in WT HBV and mutants of the capsid proteins by using a coarse-grained molecular model. Unlike essentially all existing theories of DNA/RNA packaging, this work explicitly accounts for the valences and sizes of salt ions and macromolecules. It has been repeatedly demonstrated by molecular simulations and by sophisticated density functional theory calculations that the Debye screening parameter alone is insufficient to represent the effect of small ions on interactions between macroions, i.e., complexation of nucleic acids with oppositely charged

Submitted August 28, 2008, and accepted for publication January 8, 2009.

*Correspondence: jwu@engr.ucr.edu

Editor: Nathan Andrew Baker.

© 2009 by the Biophysical Society

0006-3495/09/04/3065/9 $2.00

doi: 10.1016/j.bpj.2009.01.009

polypeptides in this case. Except for in systems with only monovalent ions at low salt concentration, theoretical calculations based on the Debye-Hückel-like approximation may lead to erroneous results. Our molecular model accounts for all important factors underlying the equilibrium properties of RNA nucleocapsids, including capsid size, molecular excluded volume, and ion-explicit electrostatic interactions. Without any ad hoc adjustable parameters, the coarse-grained model is able to predict the RNA contents in WT and mutant HBV nucleocapsids in fair agreement with experiments. Whereas the genome size shows a strong linear correlation with the capsid charge, both the slope and intercept differ from those of previous predictions that were based purely on electrostatic interactions (10). This work demonstrates the potential of coarse-grained methods for quantifying nonspecific multibody forces important for understanding viral packaging.

Molecular model and theory

To formulate a theoretical model for describing HBV nucleocapsid formation and stability, we assume that pgRNA encapsidation can be represented by a thermodynamic process that involves the nucleic acids binding to a preformed capsid with flexible protamine tails grafted at the inner surface (Fig. 1 *a*). The most stable nucleocapsid contains all or a fraction of the pgRNA such that the genome content minimizes an appropriately defined thermodynamic potential. The thermodynamic approach to RNA encapsidation is justified by experiments on the formation of HBV nucleocapsids in *Escherichia coli* that lacks the pgRNA and the viral reverse transcriptase (11,12). The bacterial expression of HBcAg results in RNA-containing nucleocapsids with genome length and internal structure that closely resemble those of a WT virus (11), suggesting that the internal structure of a nucleocapsid is not sensitive to the atomic details of the pgRNA, its secondary structures, or the specific cellular environment. In other words, the final state of nucleocapsid formation can be described by a thermodynamic process such that the genome length and the internal structure satisfy an equilibrium criterion. Although the thermodynamic model provides no information on the kinetics of nucleocapsid formation, it captures the internal properties of encapsidated RNA and the protamine tails in a stable nucleocapsid.

The thermodynamic framework allows us to describe the key components of an HBV nucleocapsid by using coarse-grained molecular models. Similar representations have been extensively used in the literature to describe structural and thermodynamic properties of biomacromolecular systems including packaging of DNA/RNA, protein folding, and assembly of viral capsids (13–16). Specifically, the HBV capsid is depicted here as a rigid spherical shell with a uniform surface charge density. The unstructured amino-acid residues at the C-terminal domains of the capsid proteins are represented by tangentially connected chains of hard spheres, each standing for one amino acid residue. For simplicity, these coarse-grained segments are assumed to have the same size but are distinguishable by their valences ($Z_T = +1$, -1, and 0 for charged and neutral residues). The average segment diameter size, $\sigma_T = 5$ Å, is derived from the 3D structures according to diameter the Brookhaven Protein Data Bank (17). We assume that at physiological conditions, each arginine residue carries one positive charge, a glutamate residue carries one negative charge, and a negative charge is also assigned to each phosphorylated serine residue (155, 162, and 170) (18). All other amino acid residues at the C-terminus of HBcAg are free of electrostatic charge.

We represent RNA as a tangent chain of charged spheres with identical diameter and valence, $\sigma_R = 7.5$ Å and $Z_R = -1$, respectively. Similar parameters for RNA have been used in a recent simulation work (14). These parameters reflect the

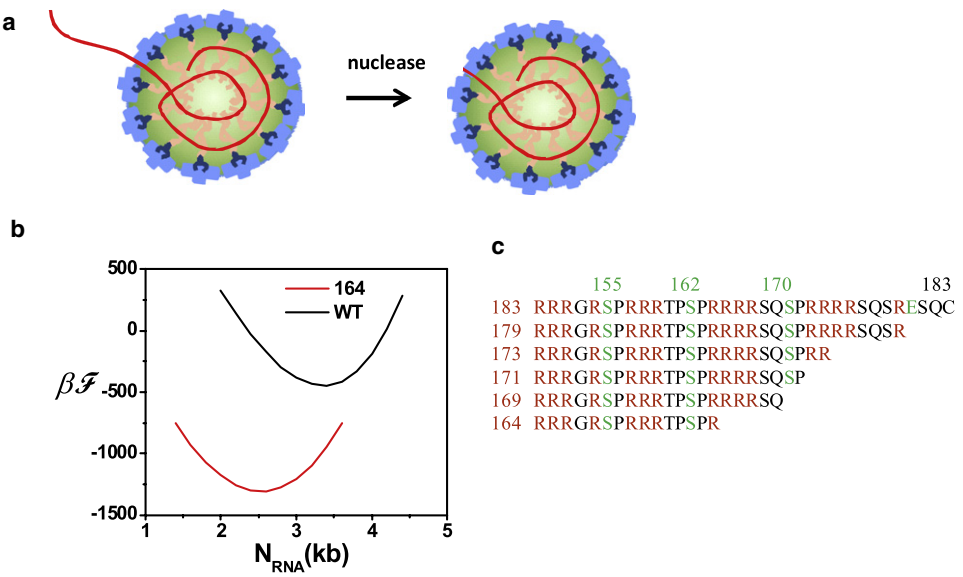

FIGURE 1 Thermodynamic representation of pgRNA stability. (*a*) The RNA fragment outside the capsid is subject to nuclease digestion. (*b*) The stable nucleocapsid yields a minimum thermodynamic potential defined by Eq. 4. Shown here are thermodynamic potentials corresponding to the WT HBV and mutant 164. (*c*) Sequences of amino acid residues at the arginine-rich domains of WT HBcAg (183 amino acids) and the mutants with truncated C-termini (residues 164–179). These mutants are employed in both experiments and theoretical calculations. The electrostatic charge of each amino acid residue is coded with a different color: *red*, +1; *teal*, −1; and *black*, neutral. Three serine sites (S155, 162, and 170) are phosphorylated.

average size and valence of nucleotides. Within the capsid lumen and host cell, the aqueous medium is assumed to be in equilibrium with a bulk NaCl solution at the physiological condition. As in a standard model of electrolyte solutions, the cations and anions are represented by spherical particles and water molecules by a continuous dielectric medium. In this work, the salt concentration is $C_S = 140$ mM , and the ionic diameters are $\sigma_{Na^+} = 3.9$ Å and $\sigma_{Cl^-} = 3.6$ Å, respectively. The diameters of small ions are taken from Simonin et al. (19), who obtained them by fitting the thermodynamic properties of the bulk electrolyte solutions. In addition to C-terminal tails and encapsidated nucleic acids, each HBV capsid contains a single copy of the reverse transcriptase. Its impact on the thermodynamic properties of RNA packaging is ignored, because the protein, ~3 nm in radius, is expected to have only a negligible effect on the equilibrium properties of the entire nucleocapsid.

In dimensionless units, the pair interaction potential between nonbonded polymer (RNA/polypeptide) segments includes a hard-sphere repulsion and Coulomb interaction

$$\beta u_{ij}(r) = \begin{cases} \infty, & r < (\sigma_i + \sigma_j)/2 \\ l_B \frac{Z_i Z_j}{r}, & r \geq (\sigma_i + \sigma_j)/2 \end{cases}, \qquad (1)$$

where r is the center-to-center distance, Z_i is the charge valence of segment i, $\beta = 1/k_BT$, k_B is the Boltzmann constant, and $T = 298.15$ K. The Bjerrum length is $l_B = 0.714$ nm, corresponding to that for pure water at 298 K. Equation 1 also applies to the interactions between small ions and to that between a small ion and a polymer segment. Without polyions, the potential becomes identical to that used in a standard theory of electrolyte solutions.

Within the tangent-chain model of polymers, the bonding potential $V_b(\mathbf{R})$ among coarse-grained RNA/polypeptide segments is described by the Dirac δ function

$$\exp[-\beta V_b(\mathbf{R})] = \prod_{i=1}^{M-1} \frac{\delta(|\mathbf{r}_{i+1} - \mathbf{r}_i| - \sigma_s)}{4\pi\sigma_s^2}, \qquad (2)$$

where σ_s denotes the segment diameter, M stands for the degree of polymerization, $\mathbf{R} \equiv (\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_M)$ specifies the molecular configuration or positions of individual segments, and δ stands for the one-dimensional Dirac δ function. The bonding potential merely imposes the chain connectivity of the coarse-grained segments. The tangent chain model ignores the bending energy and intramolecular interactions leading to the secondary structures of RNA and polypeptides. As discussed earlier, we assume that equilibrium properties of encapsidated RNA are not sensitive to the atomic details.

Because the HBV capsid is fenestrated with nonspherical holes as large as 2.4 nm × 1.8 nm (5), the capsid surface is permeable to small ions but not to the macromolecules. The $T = 4$ HBV capsid is thus approximated as a spherical shell with an inner radius $R_c = 13$ nm and a surface charge density $Q_C = 0.7/\text{nm}^2$ (20). The electrostatic charge of the HBV capsid arises from that of the core proteins, which in this work are represented by a spherical shell without consideration of the microscopic details. The charge density of the empty capsid (without C-terminal tails) was reported in an earlier publication (20), and was obtained by fitting experimental data for the equilibrium constant for self-assembly of the viral capsid. The inner surface of the HBV capsid is ''tethered'' with 240 copies of the protamine tails, one from each capsid protein. In a WT capsid, each protamine tail consists of 34 amino acid residues, among which 16 are positively charged (arginine residues), four are negatively charged (three phosphorylated serines and one glutamate), and the rest are neutral (18,21). HBV mutants consist of core proteins with partially truncated protamine tails. The protamine tails and encapsidated RNA are restricted to the interior space of the HBV nucleocapsid.

The interaction between an RNA/polypeptide segment and the capsid surface is described by a spherically symmetric potential,

$$\Psi(r) = \begin{cases} \infty & r > R_c - \sigma_i/2 \\ Z_i e \varphi_s(r) & r \leq R_c - \sigma_i/2 \end{cases}, \qquad (3)$$

where $\varphi_S(r)$ represents the electrostatic potential due to the surface charge. Unlike a nonelectrostatic external potential, the electrostatic potential of the surface is coupled with the overall electrostatic potential of the entire charge system. Except for the electrostatic potential, the capsid exerts no direct effect on small ions.

Although our coarse-grained description of RNA packaging provides no atomic details, it retains the appropriate size and charge of key components important for the thermodynamics of genome packaging. Because the numerical values for the electrostatic status and diameter of individual nucleotides and amino acids are well established, the coarse-grained model is free of any adjustable parameters. As a result, the theoretical predictions provide a stringent test of the thermodynamic approach to modeling nucleocapsid formation. The predictive method is particularly useful for describing the microscopic structure and properties of nucleocapsids in vivo.

Thermodynamic potential

Recent mutagenesis experiments suggest that upon nucleocapsid formation, an RNA chain binds the protamine tails inside the HBV capsid with the 3′ end slicing between the capsid lumen and the host cell (Fig. 1 *a*) (8,22). Because the RNA fragment exposed outside the capsid is subject to enzyme digestion, only the encapsidated RNA is biologically active, e.g., in its reverse transcription to DNA within the HBV nucleocapsid. Therefore, in this work, each nucleocapsid is considered as a semi-open system defined by the number of protamine tails grafted at the inner surface, a partially encapsidated RNA chain, and permeable small ions with chemical potentials defined by a bulk solution.

The optimal amount of RNA in a stable nucleocapsid is determined by minimization of a thermodynamic potential that depends not only on the internal structure of the nucleocapsid but also on the thermodynamic properties of the local cellular environment (23):

$$F = F_{\text{capsid}} + F_{\text{out}}, \tag{4}$$

where F_{capsid} stands for the semigrand-canonical potential of a nucleocapsid, and F_{out} is the thermodynamic potential of the RNA outside the capsid. Equation 4 reflects the partition of an RNA chain into a portion inside the capsid and the part that remains outside.

The first term on the right side of Eq. 4 accounts for RNA-capsid interactions; this term depends on the capsid size, the number densities of RNA and protamine tails, and the chemical potentials of small ions. Within the coarse-grained model, the thermodynamic properties of a nucleocapsid can be calculated by using a density functional theory (DFT) for inhomogeneous polyelectrolytes (13). The free energy of RNA outside the nucleocapsid accounts for RNA interaction with the cellular milieu. This term can be estimated from the reversible work to create a cavity within the host cell (to accommodate the unpackaged RNA segments),

$$F_{\text{out}} = \Pi v\left(N_{\text{pgRNA}} - N_{\text{RNA}}\right), \tag{5}$$

where Π stands for the osmotic pressure of the cellular environment, ν is the average volume of an RNA segment, N_{pgRNA} represents the total number of segments for the entire RNA chain (e.g., pgRNA), and N_{RNA} is the number of encapsidated nucleotides. Because of the compartments, the host cell does not have a uniform osmotic pressure; here, Π corresponds to a local osmotic pressure within the host cell in which RNA encapsidation occurs. Although the local osmotic pressure is unknown, we find that the osmotic pressure term has little effect on the total thermodynamic potential. In this work, we assume $\Pi = 10$ bar, which is reasonable for a cellular environment. The total free energy (Eq. 4) is minimized with respect to the length of RNA inside the capsid. Because of the linearity of the second term, the thermodynamic potential of RNA is independent of the length of RNA outside the capsid.

The free energy of the nucleocapsid corresponds to a semigrand potential that depends on the molecular composition of the grafted polypeptides, the length of encapsidated RNA, the chemical potential of salt ions, and the interaction potential between the capsid and its interior components,

$$\begin{aligned} F_{\text{capsid}} = & \sum_{\alpha=+,-} \int d\mathbf{r}\rho_\alpha(\mathbf{r})[\ln\rho_\alpha(\mathbf{r}) - 1] \\ & + \sum_{\alpha=+,-} \int d\mathbf{r}\rho_\alpha(\mathbf{r})\left[\Psi_\alpha(\mathbf{r}) - \mu_\alpha\right] \\ & + \sum_{M=\text{R,T}} \int d\mathbf{R}\rho_M(\mathbf{R})\Psi_M(\mathbf{R}) + F^{\text{ex}}. \end{aligned} \tag{6}$$

In Eq. 6, subscript α denotes small ions and M for macroions (R = RNA, T = polypeptide); $\rho_{\text{M}}(\mathbf{R})$ and $\rho_\alpha(\mathbf{r})$ stand for the density profiles of tethered polymers and small ions, respectively; and ψ represents the capsid external potential. The excess Helmholtz energy, F^{ex}, accounts for the thermodynamic nonideality of the system, i.e., the free energy arising from nonbonded interactions. In calculating ψ, we solve the electrostatic potential of the capsid along with the density profiles of all ionic species, including those from tethered polypeptides and encapsidated RNA.

Because both polypeptides and RNA are confined within the capsid, the semigrand potential (Eq. 6) excludes the translational entropies of the chain molecules. However, in excluding the ideal-gas part of the chain free energies for the tethered polypeptides and RNA, we have also discarded contributions due to both the translational and conformational entropies of the confined chains. Although these two contributions cannot be decoupled in the DFT calculations, we may estimate the effect of chain conformation on RNA packaging by using equations for an ideal Gaussian chain confined in a spherical cavity. In the ground-state dominance approximation, we can calculate the free energy for confining a chain of N Kuhn segments in a spherical cavity of radius R_{c} (24),

$$\beta F \approx \frac{\pi^2 a^2}{6R_{\text{c}}^2} N, \tag{7}$$

where a is the Kuhn length, taken to be the segment diameter for our tangentially connected sphere chains. Equation 7 provides an upper bound for the thermodynamic driving force of RNA packaging related to the chain confinement. Because the conformational entropy of the tethered polypeptides is insensitive to the RNA length, the reduced thermodynamic potential due to the RNA conformation is $\partial\beta F_{\text{R}}/\partial N \approx \pi^2 a^2/(6R_{\text{c}}^2) \approx 5 \times 10^{-3}$ regardless of the RNA chain length. In comparison with the other effects (see Fig. 2 *b*), the conformational entropy of polymers has little influence on RNA packaging. Therefore, the main effect of discarding the ideal-gas part of the free energy for RNA/polypeptide chains is the desired exclusion of the translational entropy of the chains.

The number of nucleotides in a stable nucleocapsid is found by minimizing the thermodynamic potential (23):

$$\frac{\partial F}{\partial N_{\text{RNA}}} = 0. \tag{8}$$

Because F_{out} is a linear function of N_{RNA} and N_{pgRNA} is a constant, Eq. 8 can be rewritten as

$$\frac{\partial F_{\text{capsid}}}{\partial N_{\text{RNA}}} = \Pi v. \tag{9}$$

Equation 9 suggests that at thermodynamic equilibrium, the reversible work to insert an RNA segment inside the nucleocapsid is the same as that to push the same RNA segment

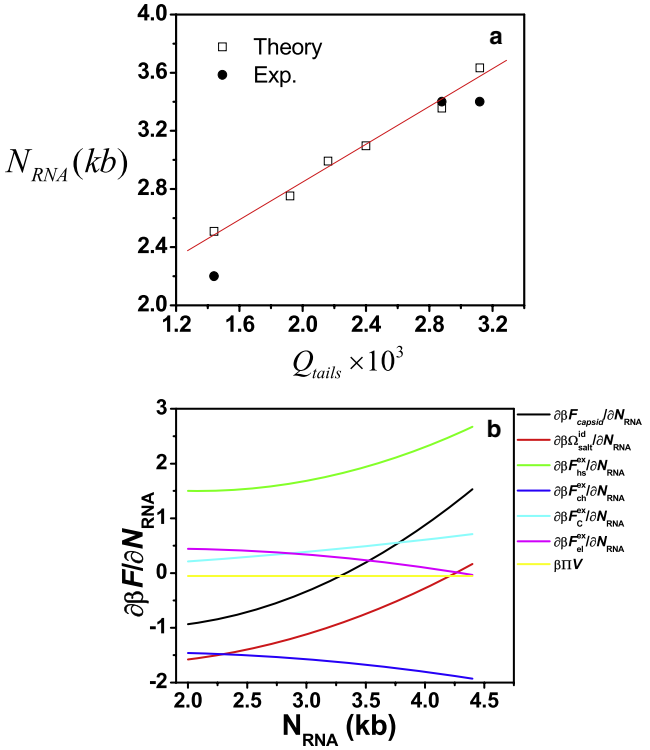

FIGURE 2 (*a*) Correlation between RNA length (in kilobases) and the net charge of the C-terminal tails in WT and mutated HBV capsids. The solid line is to guide the eye. (*b*) Individual contributions to the driving forces behind RNA packaging in WT capsids.

outside. This equivalence of reversible work can be intuitively understood as a mechanical ''force'' balance of RNA encapsidation.

We use a DFT developed in our earlier publications (25,26) to calculate the Helmholtz energy of nucleocapsids. The numerical performance of the DFT has been tested by extensive comparison of theoretical predictions with experiment and simulation results for both free and tethered polyelectrolytes. It has been shown that the DFT is accurate for both electrostatic interactions and excluded volume effects important for RNA packaging. We refer readers to earlier publications (25,26) for the technical details. Briefly, the DFT can be intuitively understood as an extension of the polyelectrolyte Poisson-Boltzmann equation; it incorporates the effects of ionic size, and the electrostatic and chain correlations that are ignored by the mean-field method. The DFT allows us to describe the local packaging of polyelectrolytes and the associated counterions and electrostatic and intrachain correlations. The numerical reliability of the non-mean-field methods is well documented for inhomogeneous electrolyte solutions and polymers (27,28).

A key component of the DFT for polymeric systems is formulation of an excess Helmholtz energy function that accounts for the thermodynamic nonideality due to the nonbonded inter- and intramolecular interactions and correlation effects. Within the coarse-grained model of the HBV nucleocapsid, the excess Helmholtz energy functional can be divided into contributions due to the excluded-volume effects, direct Coulomb interactions, and charge/chain correlations:

$$\beta F^{\mathrm{ex}} = \beta F^{\mathrm{ex}}_{\mathrm{hs}} + \beta F^{\mathrm{ex}}_{\mathrm{ch}} + \beta F^{\mathrm{ex}}_{\mathrm{C}} + \beta F^{\mathrm{ex}}_{\mathrm{el}}, \tag{10}$$

where

$$\beta F^{\mathrm{ex}}_{\mathrm{hs}} = \int d\mathbf{r}\Bigg\{ - n_0 \ln(1-n_3) + \frac{n_1 n_2 - \mathbf{n}_{V1}\mathbf{n}_{V2}}{1-n_3} + \frac{1}{36\pi}\left[n_3 \ln(1-n_3) + \frac{n_3^2}{(1-n_3)^2}\right]\frac{\left(n_2^3 - 3n_2\mathbf{n}_{V2}\mathbf{n}_{V2}\right)}{n_3^3}\Bigg\} \tag{11}$$

$$\beta F^{\mathrm{ex}}_{\mathrm{ch}} = \frac{1-M}{M}\int d\mathbf{r}\ \ n_{0p}\left(1 - \mathbf{n}_{V2p}\mathbf{n}_{V2p}/n_{2p}^2\right)\ln y(\sigma_p, n_\alpha) \tag{12}$$

$$\beta F^{\mathrm{ex}}_{\mathrm{C}} = \frac{l_B}{2}\sum_{i,j}\iint d\mathbf{r}d\mathbf{r}'\frac{Z_i Z_j \rho_i(\mathbf{r})\rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{13}$$

$$\beta F^{\mathrm{ex}}_{\mathrm{el}} E = \beta F^{\mathrm{ex}}_{\mathrm{el}}\left(\{\rho_i^b\}\right) - \int d\mathbf{r}\sum_i \Delta C_i^{(1)\mathrm{el}}\left[\rho_i(\mathbf{r}) - \rho_i^b\right] - \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}'\sum_{i,j}\Delta C_{ij}^{(2)\mathrm{el}}\left[\rho_i(\mathbf{r}) - \rho_i^b\right]\left[\rho_j(\mathbf{r}') - \rho_j^b\right]. \tag{14}$$

In Eqs. 10–14, $n_\alpha(\mathbf{r}),\alpha = 0,1,2,3,V1,V2$, are the scalar and vector-weighted densities from the fundamental measure theory (29); $\Delta C_i^{(1)\mathrm{el}}$ and $\Delta C_{ij}^{(2)\mathrm{el}}(r)$ are, respectively, the first-order and second-order electrostatic direct correlation functions of charged particles at bulk densities ρ_i^b; $F_{el}^{ex}(\{\rho_i^b\})$ is the corresponding electrostatic Helmholtz energy of the charged particles in the bulk; and $y(\sigma_p,n_\alpha)$ stands for the contact value of the cavity correlation function of the polyion segments.

In Eq. 10, $F^{\mathrm{ex}}_{\mathrm{hs}}$ is obtained from a modified fundamental measure theory for hard spheres (30,31); this term gives an accurate description of the free energy affiliated with the excluded volumes of nonbonded polymeric segments and small ions. $F^{\mathrm{ex}}_{\mathrm{ch}}$ represents the Helmholtz energy due to the intrachain correlation, which follows an extension of the first-order thermodynamic perturbation theory for polymeric fluids to inhomogeneous systems (25,33). $F^{\mathrm{ex}}_{\mathrm{C}}$ arises from the direct Coulomb interactions among charged species; this term is identical to that given by the Poisson-Boltzmann (PB) equation. As is well documented, the PB equation ignores ionic size and the electrostatic and intrachain correlations. The Helmholtz energy due to the electrostatic correlations, $F^{\mathrm{ex}}_{\mathrm{el}}$, is derived from a quadratic functional expansion of the electrostatic free energy (34,35), where the direct correlation functions of the uniform reference system are obtained from the mean spherical approximation (36). The

numerical performance of individual terms in Eq. 10 has been calibrated in our previous publications (25,26).

RESULTS AND DISCUSSION

To quantify the role of electrostatic interactions in nucleocapsid formation, we consider RNA encapsidation in WT HBV and various mutants in which the core protein contains truncated protamine tails. Fig. 1 *c* shows the amino acid sequences of the protamine tails considered in this work. These sequences are identical to those used in early mutagenesis experiments (8). Given a protamine tail length and sequence, we have calculated the thermodynamic potential of the nucleocapsid F with different amounts of encapsidated RNA and identified the optimal genome length in the capsid for the WT HBV and mutants.

Fig. 1 *b* shows representative examples of how the total free energy of the nucleocapsid varies with the number of encapsidated nucleotides for the WT HBV and mutant 164. Similar plots are obtained for all other mutants. The stable nucleocapsid contains an optimal number of nucleotides that minimizes the thermodynamic potential (Eq. 4). Table 1 presents theoretical predictions of the RNA lengths in HBV capsids in comparison with experimental data (8). The agreement is good for both the WT HBV and mutant 164, two nucleocapsids with known RNA contents. It is of interest that the theory predicts the RNA length in WT HBV, matching almost exactly the result obtained from *E. coli* (4). The success of our coarse-grained model suggests that the thermodynamics of RNA encapsidation is indeed dominated by nonspecific interactions, and that the genome content is little influenced by atomic details, including the secondary RNA structure.

Except for mutant 179, all mutants contain less RNA compared with the WT. Although in general the RNA content is increased upon retention of more amino acid residues at the C-terminus of the capsid protein, there is no direct correlation between the RNA length and the length of C-terminal tails. The lack of size correlation suggests that RNA encapsidation is not sensitive to the capsid volume (unlike DNA packaging in bacteriophages) or the surface area. Because mutant 179 lacks a glutamate and three neutral residues at the protamine tail, we predicted that it would be able to attain a larger amount of RNA compared with the WT. As observed in experiments (8), the RNA lengths in mutants 173 and 179 are close to that in the WT capsid.

TABLE 1 Number of amino acid residues at C-terminal tails of HBcAgs and electrostatic parameters for the WT and mutated HBV capsids

Mutants	M_T	q_T	η_T	Q_T	Q_R^{pre}	R^{pre}	Q_R^{exp}
183 (WT)	34	12	0.35	2880	3356	1.17	3400
164	15	6	0.4	1440	2508	1.74	2200
169	20	9	0.45	2160	2992	1.39	
171	22	8	0.36	1920	2752	1.43	
173	24	10	0.42	2400	3097	1.29	
179	30	13	0.43	3120	3632	1.16	~3400

M_T, number of amino acid residues at the C-terminal tails of HBcAgs; q_T, electrostatic charges on each tail; η_T, average number of charges per residue; Q_T, net charge of all C-terminal tails; Q_R^{pre}, predicted total charges of encapsidated RNA; R^{pre}, nominal charge inversion ratio; Q_R^{exp}, experimental values for the total charges of the encapsulated RNA.

Fig. 2 *a* shows that the RNA length in various mutants of HBV is approximately linearly correlated with the net charge of the C-terminal tails. The small deviation from the straight line is probably due to the specific sequences of the protamine tails. Although qualitatively the increase of the genome length with the capsid charge can be understood in terms of charge neutrality, the thermodynamic potential is affected not only by the net charge of the capsid but also by the charge distribution at the protamine tails and by the molecular excluded-volume effects. In addition, as Table 1 indicates, the ratio of these opposite charges is not fixed. Upon restoration of the amino acid residues from mutant 164 to WT, the charge ratio changes from 1.74 to 1.16.

Fig. 2 *b* shows various molecular driving forces for RNA encapsidation in the WT capsid. The different components of the excess thermodynamic potential are defined in Eqs. 10–14. Here the contribution due to small ions is obtained by using the grand potential for the salt ions, defined as

$$\beta\Omega_{salt}^{id} = \sum_{\alpha=+,-}\int d\mathbf{r}\rho_\alpha(\mathbf{r})[\ln\rho_\alpha(\mathbf{r})-1] + \sum_{\alpha=+,-}\int d\mathbf{r}\rho_\alpha(\mathbf{r})\left[\Psi_\alpha(\mathbf{r})-\mu_\alpha\right]. \quad (15)$$

It is clear that RNA binding to the protamine tails is most favored by the entropy effect affiliated with the release of small ions upon neutralization of the oppositely charged macromolecules. The origin of this entropic attraction has been discussed in earlier work on the binding of ligand and DNA (37), and more recently by a simulation of the interaction between oppositely charged micelles or globular proteins (38). As expected, an increase of the osmotic pressure outside the nucleocapsid favors RNA encapsidation. However, such an effect is negligible in comparison with contributions due to intermolecular interactions and entropy of salt ions. Fig. 2 *b* also indicates that the excluded-volume effects disfavor RNA packaging and that the opposite is true for the intrachain correlations. It is rather surprising that the electrostatic energy within the nucleocapsid, including the charge correlation effects, is against RNA encapsidation.

As indicated earlier, the osmotic pressure has only a minor effect on the optimal RNA length. For example, the equilibrium RNA length changes from 3307 for osmotic pressure set to 0 bar to 3356 for 10 bar. However, the effect of capsid charge on RNA packaging is nonnegligible. Compared to a neutral capsid, the surface charge density $Q_C = 0.7/\text{nm}^2$ of the capsid increases the optimal RNA length by ~30%. For example, the optimal RNA length in a WT HBV capsid is 2335 nucleotides if the surface is neutral and 3356 nucleotides for the capsid with surface charge density $Q_C = 0.7/\text{nm}^2$.

The coarse-grained model also allows us to quantify the inhomogeneous distributions of the amino acid residues and nucleotides within an HBV nucleocapsid. The microscopic information is important for understanding reverse transcription in HBV nucleocapsids from pgRNA to a partially double-stranded DNA, and the subsequent nucleocapsid envelopment and secretion. However, the internal structure is difficult to obtain from direct measurements. Fig. 3 shows the local densities of RNA and protamine tail segments in the capsid lumen for WT HBV and various mutants considered earlier. As observed in cryoelectron microscopy (cryo-EM) measurements (5,11,39), the encapsidated RNA is mainly located near the inner surface of the capsid and the packing density is extremely low at the center. A strong band of RNA between 90 and 125 Å in radius in the WT capsid is consistent with the density peak near the inner surface predicted by our theoretical calculations. In addition to the strong peak close to the inner surface, the density profile of RNA segments has a broad, though shallow, distribution tail extending to 40 Å in radius, roughly coinciding with the thickness of the C-terminal tails. Except for the peak near the inner surface of the capsid, where the RNA density remains nearly constant across different viral variants, the distribution of RNA segments within the capsid lumen closely follows that of the protamine tails. Beyond the brush region of the protamine tails, the density of RNA segments reaches a low plateau and remains approximately constant throughout the capsid center. It appears that the central region is mainly occupied by a small number of RNA segments and cations that help to maintain the charge neutrality.

The local densities of charged components inside the HBV nucleocapsid are closely correlated with their charges, positions along the polymer chain, and distribution of the mean electrostatic potential. Fig. 4 *a* shows the local packing density of amino acid residues with respect to their charges for a WT HBV virus. The strong density peaks and oscillations near the capsid shell are attributed to the tethering effect and chain connectivity. Whereas the positively charged arginine residues present a sharp peak near the inner surface and the density profile spans the entire polypeptide brush, the neutral and negative amino acids exhibit much broader distributions. Fig. 4 *b* shows the density distributions of three phosphorylated serine sites (155, 162, and 170), and the mean electrostatic potential in a WT HBV capsid. Although all serine residues exhibit Gaussian-like distributions, serine 155 is restricted to a much narrower region in comparison with the other two, because it is closest to the capsid surface.

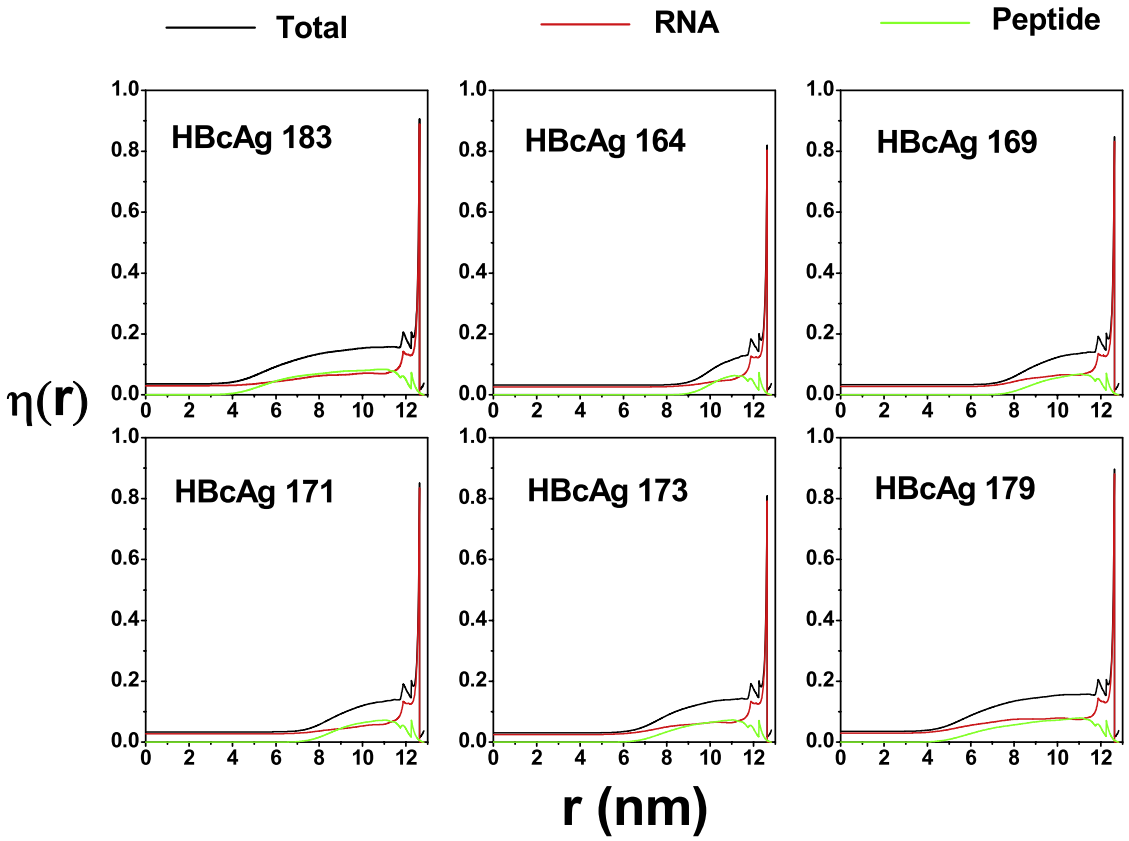

FIGURE 3 Local packing densities of RNA segments and C-terminal tails of HBV nucleocapsids. The local packing fraction for RNA/protamine tail segments is defined as $\eta(r) = \pi\rho(r)\sigma^3/6$, where $\rho(r)$ represents the local number density and σ the segment diameter.

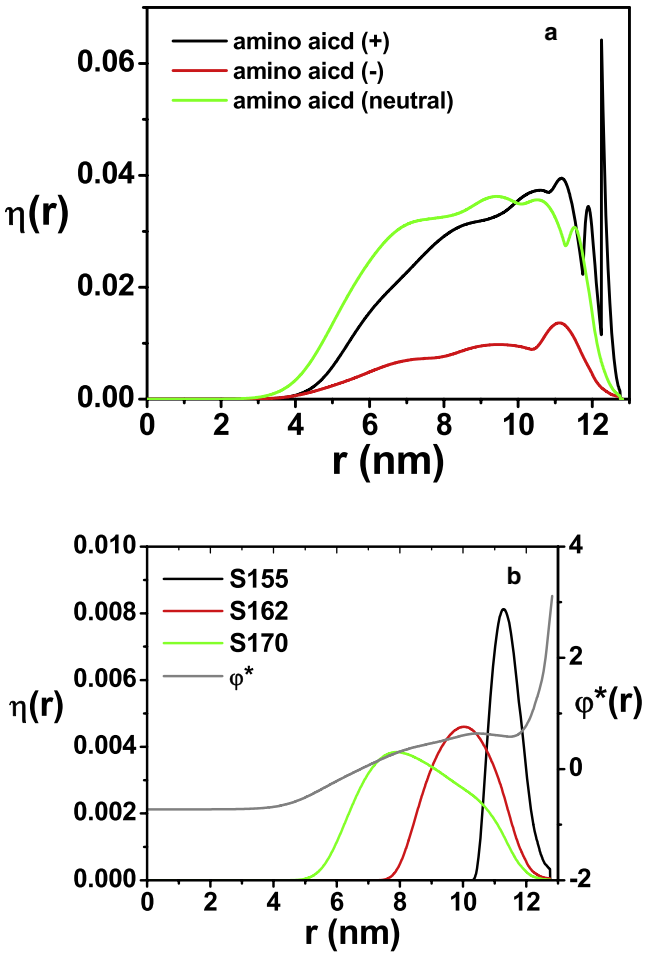

FIGURE 4 (*a*) Local packing fractions of amino acids in the C-terminal tail with different charges. (*b*) Distributions of the phosphorylated serine residues and the mean electrostatic potential inside the WT virus.

It is interesting to note that the electrostatic potentials at the peak positions of serine residues are close to each other, implying a correlation between the local electrostatic field and phosphorylation status of serine residues. Diverging from an earlier theoretical prediction (10), we find that the local electrostatic potential varies with RNA packaging and cannot be represented by a parabolic function.

CONCLUSIONS

We have shown that the RNA contents and the internal structure of WT HBV and mutants can be captured quantitatively by a coarse-grained model with explicit description of the electrostatic charge and excluded-volume interactions. The theoretically predicted RNA lengths are in good agreement with experiments. Although stable nucleocapsids maintain a proper balance between oppositely charged macromolecules, direct electrostatic interaction makes a relatively insignificant contribution to the overall molecular driving forces of RNA encapsidation. Furthermore, the molecular model is able to describe the internal structure of nucleocapsids, which is not directly accessible to conventional experiments. In addition to a strong shell of RNA segments next to the capsid inner surface, we find that within the capsid lumen, the RNA density closely follows that of the protamine tails. In light of the generic nature of the molecular model for nucleocapsids, we expect that the theoretical description of RNA encapsidation presented in this work is applicable not only to HBV but to a wide variety of ssRNA/ssDNA viruses that entail strong electrostatic interactions for nucleocapsid formation (40).

The authors are grateful to Dr. Jianming Hu for insightful comments on the biological aspects of HBV. For financial support and computer time allocations, we are grateful to the U.S. Department of Energy (DE-FG02-06ER46296) and to the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under contract No. DE-AC03-76SF0009.

REFERENCES

1. Ganem, D. 2004. Hepatitis B virus infection: natural history and clinical consequences. *N. Engl. J. Med.* 350:1118–1129. (Erratum in *N. Engl. J. Med.* 2004. 351:1268).
2. Nassal, M. 2008. Hepatitis B viruses: reverse transcription a different way. *Virus Res.* 134:235–249.
3. Bruss, V. 2007. Hepatitis B virus morphogenesis. *World J. Gastroenterol.* 13:65–73.
4. Beck, J., and M. Nassal. 2007. Hepatitis B virus replication. *World J. Gastroenterol.* 13:48–64.
5. Zlotnick, A., N. Cheng, S. J. Stahl, J. F. Conway, A. C. Steven, et al. 1997. Localization of the C terminus of the assembly domain of hepatitis B virus capsid protein: implications for morphogenesis and organization of encapsidated RNA. *Proc. Natl. Acad. Sci. USA.* 94:9556–9561.
6. Beames, B., and R. E. Lanford. 1993. Carboxy-terminal truncations of the HBV core protein affect capsid formation and the apparent size of encapsidated HBV RNA. *Virology.* 194:597–607.
7. Kock, J., M. Nassal, K. Deres, H. E. Blum, and F. von Weizsacker. 2004. Hepatitis B virus nucleocapsids formed by carboxy-terminally mutated core proteins contain spliced viral genomes but lack full-size DNA. *J. Virol.* 78:13812–13818.
8. Le Pogam, S., P. K. Chua, M. Newman, and C. Shih. 2005. Exposure of RNA templates and encapsidation of spliced viral RNA are influenced by the arginine-rich domain of human hepatitis B virus core antigen (HBcAg 165–173). *J. Virol.* 79:1871–1877.
9. van der Schoot, P., and R. Bruinsma. 2005. Electrostatics and the assembly of an RNA virus. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 71:061928.
10. Belyi, V. A., and M. Muthukumar. 2006. Electrostatic origin of the genome packing in viruses. *Proc. Natl. Acad. Sci. USA.* 103:17174–17178.
11. Roseman, A. M., J. A. Berriman, S. A. Wynne, P. J. G. Butler, and R. A. Crowther. 2005. A structural model for maturation of the hepatitis B virus core. *Proc. Natl. Acad. Sci. USA.* 102:15821–15826.
12. Bottcher, B., S. A. Wynne, and R. A. Crowther. 1997. Determination of the fold of the core protein of hepatitis B virus ky electron cryomicroscopy. *Nature.* 386:88–91.
13. Li, Z. D., J. Z. Wu, and Z. G. Wang. 2008. Osmotic pressure and packaging structure of caged DNA. *Biophys. J.* 94:737–746.
14. Zhang, D. Q., R. Konecny, N. A. Baker, and J. A. McCammon. 2004. 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.* 75:325–337.

15. Hagan, M. F., and D. Chandler. 2006. Dynamic pathways for viral capsid assembly. *Biophys. J.* 91:42–54.

16. Petrov, A. S., and S. C. Harvey. 2007. Structural and thermodynamic principles of viral packaging. *Structure*. 15:21–27.

17. Kolinski, A., A. Godzik, and J. Skolnick. 1993. A general method for the prediction of the 3-dimensional structure and folding pathway of globular proteins: application to designed helical proteins. *J. Chem. Phys.* 98:7420–7433.

18. Perlman, D. H., E. A. Berg, P. B. O'Connor, C. E. Costello, and J. M. Hu. 2005. Reverse transcription-associated dephosphorylation of hepadnavirus nucleocapsids. *Proc. Natl. Acad. Sci. USA*. 102:9020–9025.

19. Simonin, J. P., O. Bernard, and L. Blum. 1999. Ionic solutions in the binding mean spherical approximation: thermodynamic properties of mixtures of associating electrolytes. *J. Phys. Chem. B*. 103:699–704.

20. Kegel, W. K., and P. van der Schoot. 2004. Competing hydrophobic and screened-Coulomb interactions in hepatitis B virus capsid assembly. *Biophys. J.* 86:3905–3913.

21. Liao, W., and J. H. Ou. 1995. Phosphorylation and nuclear localization of the hepatitis B virus core protein: significance of serine in the 3 repeated Sprrr motifs. *J. Virol.* 69:1025–1029.

22. Ostrow, K. M., and D. D. Loeb. 2004. Underrepresentation of the 3′ region of the capsid pregenomic RNA of duck hepatitis B virus. *J. Virol.* 78:2179–2186.

23. Evilevitch, A., M. Castelnovo, C. M. Knobler, and W. M. Gelbart. 2004. Measuring the force ejecting DNA from phage. *J. Phys. Chem. B*. 108:6838–6843.

24. Casassa, E. F., and Y. Tagami. 1969. An equilibrium theory for exclusion chromatography of branched and linear polymer chains. *Macromolecules*. 2:14–26.

25. Li, Z. D., and J. Z. Wu. 2006. Density functional theory for polyelectrolytes near oppositely charged surfaces. *Phys. Rev. Lett.* 96:048302.

26. Jiang, T., Z. D. Li, and J. Z. Wu. 2007. Structure and swelling of grafted polyelectrolytes: predictions from a nonlocal density functional theory. *Macromolecules*. 40:334–343.

27. Wu, J. Z. 2006. Density functional theory for chemical engineering: from capillarity to soft materials. *AIChE J*. 52:1169–1193.

28. Wu, J. Z., and Z. D. Li. 2007. Density functional theory for complex fluids. *Annu. Rev. Phys. Chem.* 58:85–112.

29. Rosenfeld, Y. 1989. Free-energy model for the inhomogeneous hard-sphere fluid mixture and density-functional theory of freezing. *Phys. Rev. Lett.* 63:980–983.

30. Roth, R., R. Evans, A. Lang, and G. Kahl. 2002. Fundamental measure theory for hard-sphere mixtures revisited: the White Bear version. *J. Phys. Condens. Matter*. 14:12063–12078.

31. Yu, Y. X., and J. Z. Wu. 2002. Structures of hard-sphere fluids from a modified fundamental-measure theory. *J. Chem. Phys.* 117:10156.

32. Li, Z., and J. Wu. 2006. Density functional theory for planar electric double layers: Closing the gap between simple and polyelectrolytes. *J. Phys. Chem. B*. 110:7473–7484.

33. Yu, Y. X., and J. Z. Wu. 2002. Density functional theory for inhomogeneous mixtures of polymeric fluids. *J. Chem. Phys.* 117:2368–2376.

34. Yu, Y. X., J. Z. Wu, and G. H. Gao. 2004. Density-functional theory of spherical electric double layers and ζ potentials of colloidal particles in restricted-primitive-model electrolyte solutions. *J. Chem. Phys.* 120:7223–7233.

35. Li, Z. D., and J. Z. Wu. 2004. Density-functional theory for the structures and thermodynamic properties of highly asymmetric electrolyte and neutral component mixtures. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 70:031109.

36. Blum, L. 1975. Mean spherical model for asymmetric electrolytes. 1. Method of solution. *Mol. Phys.* 30:1529–1535.

37. Anderson, C. F., and M. T. Record. 1993. Salt dependence of oligoion polyion binding: a thermodynamic description based on preferential interaction coefficients. *J. Phys. Chem.* 97:7116–7126.

38. Wu, J. Z., D. Bratko, H. W. Blanch, and J. M. Prausnitz. 2000. Interaction between oppositely charged micelles or globular proteins. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*. 62:5273–5280.

39. Dryden, K. A., S. F. Wieland, C. Whitten-Bauer, J. L. Gerin, F. V. Chisari, et al. 2006. Native hepatitis B virions and capsids visualized by electron cryomicroscopy. *Mol. Cell.* 22:843–850.

40. Angelescu, D. G., and P. Linse. 2008. Viruses as supramolecular self-assemblies:modelling of capsid formation and genome packaging. *Soft Matter*. 4:1981–1990.