### Title

A Thermodynamic Model for Genome Packaging in Hepatitis B Virus




### Journal

Biophysical Journal, 109(8)

### ISSN

0006-3495



### Authors

Kim, Jehoon

Wu, Jianzhong



### Publication Date

2015-10-01

### DOI

10.1016/j.bpj.2015.08.021

Peer reviewed

Article

# A Thermodynamic Model for Genome Packaging in Hepatitis B Virus

Jehoon Kim[1] and Jianzhong Wu[1,*]
[1]Department of Chemical and Environmental Engineering, University of California at Riverside, Riverside, California

ABSTRACT Understanding the fundamentals of genome packaging in viral capsids is important for finding effective antiviral strategies and for utilizing benign viral particles for gene therapy. While the structure of encapsidated genomic materials has been routinely characterized with experimental techniques such as cryo-electron microscopy and x-ray diffraction, much less is known about the molecular driving forces underlying genome assembly in an intracellular environment and its in vivo interactions with the capsid proteins. Here we study the thermodynamic basis of the pregenomic RNA encapsidation in human Hepatitis B virus in vivo using a coarse-grained molecular model that captures the essential components of nonspecific intermolecular interactions. The thermodynamic model is used to examine how the electrostatic interaction between the packaged RNA and the highly charged C-terminal domains (CTD) of capsid proteins regulate the nucleocapsid formation. The theoretical model predicts optimal RNA content in Hepatitis B virus nucleocapsids with different CTD lengths in good agreement with mutagenesis measurements, confirming the predominant role of electrostatic interactions and molecular excluded-volume effects in genome packaging. We find that the amount of encapsidated RNA is not linearly correlated with the net charge of CTD tails as suggested by earlier theoretical studies. Our thermodynamic analysis of the nucleocapsid structure and stability indicates that ~10% of the CTD residues are free from complexation with RNA, resulting in partially exposed CTD tails. The thermodynamic model also predicts the free energy of complex formation between macromolecules, which corroborates experimental results for the impact of CTD truncation on the nucleocapsid stability.

## INTRODUCTION

Hepatitis B virus (HBV) is a human pathogen that causes severe liver diseases such as hepatocellular carcinoma and liver cirrhosis (1,2). While >350,000,000 people alive today are chronically infected by HBV, contemporary treatments are effective mostly in the short term and often plagued by adverse reactions such as drug resistance (3). The lifecycle of HBV replication starts with a single-stranded pregenomic (pg) RNA binding with a large number of capsid proteins to form a fenestrated viral capsid. The encapsidated pgRNA undergoes reverse transcription, first to a single-stranded (ss) DNA and then a partially double-stranded DNA in a matured capsid. It has been recognized that pgRNA encapsidation is primarily driven by nonspecific electrostatic interactions with the C-terminal domains (CTD) of the capsid proteins (CPs). Each CP contains 183 amino-acid (aa) residues, which can be divided into an N-terminal assembly domain (140 aa), a nonpeptide linker (9 aa), and CTD (34 aa) tails. CTD is rich in arginine residues and is indispensable for proper pgRNA packaging and DNA polymerization (4–6).

Previous mutagenesis studies indicate that pgRNA packaging is strongly correlated with the net charge of the CTD tails, leading to the so-called "charge balance hypothesis" (7–9). A partial truncation of the flexible, highly charged CTD tails reduces the amount of encapsidated RNA, while restoring arginine residues to the truncated CTD tails increases the nucleic content. A positive correlation between the length of C-terminal tail and the amount of RNA was confirmed by cryo-electron microscopy measurements (10). The balanced electrostatic interaction between the positive charge from the CTD tails and the negative charge from the encapsidated nucleic acid could play an important role in many viral activities including capsid stability, assembly, RNA encapsidation, and DNA replication (7). The charge-balance hypothesis entails a stoichiometry-like correlation between basic residues and their binding partner of nucleic acids (8,9). While the experimental approaches have established near-quantitative charge-charge relationships (7,8), there still exists a unresolved question regarding how the electrostatic interaction between macromolecules is affected by solution conditions such as salt concentration, identities of low molecular weight ions, pH, and temperature (8,9). Besides, the charge distribution at the CTD tails may play an important role. It is not clear whether all arginine residues at the C-terminus make equal contributions to the electrostatic interaction with encapsidated RNA or DNA during the HBV lifecycle (9).

The regulative role of nonspecific electrostatic interactions has been identified for many ssDNA/RNA viruses that replicate through spontaneous molecular self-assembly

*Submitted March 18, 2015, and accepted for publication August 10, 2015.*
*Correspondence: jwu@engr.ucr.edu
Editor: Michele Vendruscolo.

0006-3495/15/10/1689/9
http://dx.doi.org/10.1016/j.bpj.2015.08.021

(11–19). It has been shown that that genome packaging is intimately affiliated with electrostatic interactions between the nucleic acids and oppositely charged, flexible domains of the capsid proteins (20). Previous theoretical studies suggest that viral packaging conforms a fixed charge ratio between nucleic acids and capsid proteins (21–25). In particular, a conserved ratio ($N/Q = 1.61$) of the genome size ($N$) to the overall charge of capsid proteins ($Q$) was proposed based on observations for >30 wild-type (WT) viruses, and the universality of a linear relation between capsid charge and genome size has been partially justified with a theoretical analysis (22). However, the applicability of a fixed charge ratio to all viruses was challenged by more recent studies (26,27). Genome packaging involves other mechanisms such as Donnan effects (27) and specific biological interactions including the genome structure (28,29) that may lead to nonlinearity or different charge ratios in complexation between macromolecules of opposite charges.

The universality of a fixed charge ratio between the capsids and nucleic acids is not applicable to HBV. The conserved $N/Q$ value would predict the RNA size of ~4500 bases in WT HBV, much too large compared with the size of pgRNA (3400 bases). In addition, a fixed charge ratio is inconsistent with the aforementioned mutagenesis studies that led to the charge-balance hypothesis (7–9). To describe electrostatic interactions between capsid proteins and nucleic acids, we need an improved model that takes into account the uneven contribution of the charged residues at the C-terminal tail as well as confinement effects due to the partially charged capsid shell. Besides, small salt ions play an important role in genome packaging and must be explicitly considered. For HBV, genome packaging is also related to the partial exposure of the CTD tails (30,31). CTD exposure changes not only the charge balance but also the surface properties of the viral particle important for biological activities such as nuclear entry and export (32,33).

Developing new therapeutic antiviral strategies for HBV can be helped with a good understanding of the molecular details of viral packaging. One of the critical questions about the HBV replication is how the RNA/DNA chains are associated with the flexible domains of the capsid proteins leading to the nucleocapsid formation. Previously, we studied the thermodynamic basis of RNA encapsidation using an ion-explicit coarse-grained model for HBV capsids (34). The theoretical model accounts for important factors in regulating genome packaging including electrostatic correlations, molecular excluded-volume contributions, and chain connectivity effects. It predicts the optimal size of a packaged genome in fair agreement with existing experimental data for both WT and mutant HBV nucleocapsids. In this study, we examine the regulatory role of electrostatic interactions by considering the exposable feature of the CTD tails. In addition, we validate the thermodynamic model with more extensive in vitro data for partial RNA packaging in various HBV mutants that became available only recently. Although the universal ratio ($N/Q$) of the genome size to the charge of flexible tails is not applicable to HBV (22), both theoretical and experimental results indicate near-stoichiometric dependency of the RNA size to the net charge of the CTD tails under a specific solution condition. The discrepancies among previous theoretical studies may be reconciled by accounting for the molecular excluded-volume effects and other nonspecific interactions among all viral components including salt ions. The thermodynamic model allows us to evaluate the free energy of the capsid formation and the radial distributions of key viral components and subsequently, quantify the relative stability of nucleocapsids at different levels of CTD truncation, phosphorylation, and RNA encapsidation.

## MATERIALS AND METHODS

### Molecular model and theory

#### *Coarse-grained molecular model*

We investigate the thermodynamic properties of HBV nucleocapsids by representing the key viral components with a coarse-grained molecular model. While our model does not intend to capture the atomic details or the secondary structure of pgRNA that may play a key role in the kinetics of viral genome encapsidation (28,35), it keeps essential ingredients of nonspecific interactions among macromolecules underlying the thermodynamics of nucleocapsid formation. From in vitro assembly of HBV nucleocapsids in *Escherichia coli*, it has been noted that the RNA packaging is insensitive to the origin of the genome or the RNA sequence (5,7). The nonspecific nature of RNA encapsidation suggests that the essential features of HBV capsids can be faithfully described with a coarse-grained molecular model (34,36,37).

As in our previous work (34), the encapsidated RNA and the peptide tails of capsid proteins are modeled as tangential chains of charged hard spheres. Approximately, each segment corresponds to a nucleic acid or an amino-acid residue. The diameters of hard spheres are set as $\sigma_R = 0.75$ nm (38) and $\sigma_C = 0.5$ nm (39) for nucleic acids and amino-acid residues, respectively. We assume that each RNA segment bears a negative unit charge. At physiological conditions, a WT CTD tail contains 16 positively charged segments (all from arginine residues), four negative residues (one glutamate and three phosphorylated serine residues), and 14 neutral segments. The valence of $Z = -1$ or 1 is assigned for each nucleic acid ($Z_R$) or a charged amino-acid residue ($Z_C$). It has been demonstrated that the coarse-grained model is able to capture the essential features of macromolecular interactions (38,40–43).

For fully flexible chains of tangentially connected hard spheres, the bonding potential, designated as $V(\mathbf{R})$, satisfies

$$\exp\left[-\beta V_{\mathrm{RNA/CTD}}(\mathbf{R})\right] = \prod_{i=1}^{M-1} \frac{\delta\left(|r_{i+1} - r_i| - \sigma_{R/C}\right)}{4\pi\sigma_{R/C}^2}, \quad (1)$$

where $\beta = 1/(k_B T)$, $T$ is the absolute temperature, $k_B$ is the Boltzmann constant, and $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_M)$ represents the configuration of a coarse-grained RNA/CTD chain with the positions of $M$ spherical segments specified by $\{\mathbf{r}_{i=1,2,\ldots,M}\}$. For a WT HBV virus, the total number of segments is $M = 3400$ for pgRNA, and 34 for each CTD tail. In addition to the bonding potential, we consider the nonbonded Coulomb interactions and

the excluded volume effects among the RNA/CTD segments using the pairwise additive potential $u_{ij}(r)$,

$$\beta u_{ij}(r) = \begin{cases} \infty, & r < \sigma_{ij}, \\ \dfrac{Z_i Z_j l_B}{r}, & r \geq \sigma_{ij}, \end{cases} \quad (2)$$

where $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, and $l_B = 0.714$ nm is the Bjerrum length for an aqueous solution at room temperature. The Coulomb and hard-sphere potentials are also used to describe the interaction of macromolecules with salt ions. In this work, the intracellular fluid is represented by an aqueous solution of NaCl at the physiological concentration $C_s = 0.14$ M. It has been well documented that thermodynamic properties of the aqueous electrolyte solution can be accurately reproduced with the primitive model, assuming that the cations and the anions are charged hard spheres with diameters $\sigma_{Na^+} = 0.39$ nm and $\sigma_{Cl^-} = 0.36$ nm, respectively (44).

As in our previous publications (36,37), we use a spherical shell model for the HBV T4 capsid with an inner radius $R_C = 13$ nm and a shell thickness $d = 2$ nm. The symmetric pores on the capsid shell are modeled as a semipermeable membrane that allows transport of monomeric nucleic acids and small ions. Because the flexible CTD tails are able to extrude outside the capsid surface while the packaged RNA chain is fully confined, we use an effective external potential for biomacromolecules. The confining potential for each RNA segment is given by

$$\phi_{\mathrm{RNA}}(r) = \begin{cases} \infty, & r \geq R_C - 0.5\sigma_R, \\ 0, & \text{otherwise}. \end{cases} \quad (3)$$

Each T4 HBV capsid is composed of 240 capsid proteins. Accordingly, we assume that 240 CTD tails are randomly tethered at the inner surface of the spherical shell. While the first segment of each tail is tangentially attached at the inner capsid surface, the rest of the CTD segments are subject to an effective external potential that allows partial exposure (~10%, the approximated ratio of total pore area to inner capsid area) of the CTD chains due to the presence of the capsid pores:

$$\beta\phi_{\mathrm{tail}}(r) = \begin{cases} \ln(0.1), & R_C - 0.5\sigma_t < r < R_C + d + 0.5\sigma_t, \\ 0, & \text{otherwise}. \end{cases} \quad (4)$$

In addition to confinement effects, the HBV capsid exerts an electrostatic energy on all charged components,

$$\Psi_i^C(r) = Z_i e \varphi_s(r), \quad (5)$$

where $\varphi_s$ denotes the electrostatic potential corresponding to the capsid surface charge, and $e$ represents the unit charge. For a HBV capsid, the charge density has been estimated to be $Q_c = 0.7e/\mathrm{nm}^2$, assuming a uniform charge distribution (45). The surface charge density is used to calculate the electrostatic capsid external potential $\Psi^C$.

### *Thermodynamics of RNA encapsidation*

The thermodynamics of RNA encapsidation may be considered as an equilibrium partition of a long RNA molecule between the inner and outer space of the capsid. Fig. 1 depicts schematically a packaged chain at the early stage of assembly. The exposed portion of the RNA chain is eventually degraded by the cellular nuclease (7,46), and the inside portion defines the optimal size of the genome.

The thermodynamic potential for the partitioning of the RNA fragments both inside and outside the capsid can be written as

$$F = F_{\mathrm{cap}} + F_{\mathrm{out}}, \quad (6)$$

where $F_{\mathrm{cap}}$ accounts for the free energy of the entire nucleocapsids, and $F_{\mathrm{out}}$ represents the energy of the exposed RNA in the local cellular environment

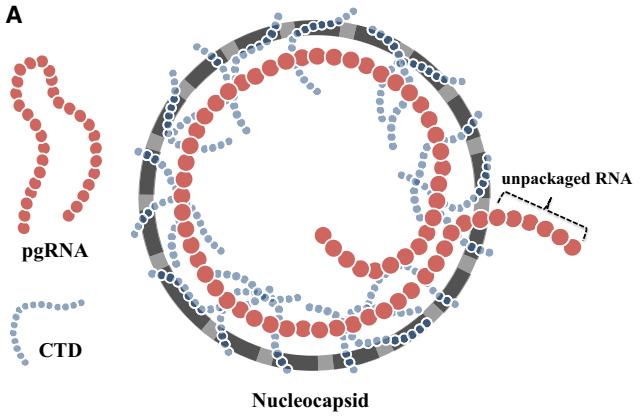


FIGURE 1 (*A*) Schematic of key components in an HBV nucleocapsid. The optimal genome size can be determined from a partial encapsidation of the RNA chain that minimizes the thermodynamic potential of the entire nucleocapsid in an aqueous electrolyte solution. The unpackaged RNA fragments are subjected to degradation by cellular nuclease. (*B*) The sequences of amino-acid residues in the CTD of the WT and mutant capsid proteins (from residue 150 to the end). In this and other figures, the WT and mutants of the core protein are distinguished with the number of residues (e.g., 183 179, 173, 171, and 164). The electrostatic status of each CTD residue is labeled with different colors available on line (*red*, *black*, and *blue* stand for positive (+), neutral, and negative (−), respectively). To see this figure in color, go online.

outside the capsid. At equilibrium, the optimal genome size can be found by minimizing the overall potential

$$\frac{\partial F}{\partial N_{\mathrm{RNA}}} = 0, \quad (7)$$

where $N_{\mathrm{RNA}}$ stands for the number of nucleotides of the packaged RNA chain.

We may estimate the free energy of RNA outside the capsid from the reversible work to insert the unpackaged RNA segments into the cellular environment. While an atomistic characterization of the cellular milieu is not attainable, the free energy of insertion may be approximated by the volumetric work to create a cavity in a uniform electrolyte solution with the intracellular osmotic pressure ($\Pi \approx 5$ bar (47,48)),

$$F_{\mathrm{out}} = \Pi v_R \left(N_{pg\mathrm{RNA}} - N_{\mathrm{RNA}}\right), \quad (8)$$

where $v_R$ stands for the excluded volume of each nucleotide, and $N_{pg\mathrm{RNA}}$ is the number of segments of the entire pgRNA chain. Because the free energy for the outside part of RNA is a linear function of $N_{\mathrm{RNA}}$, its derivative with respect to $N_{\mathrm{RNA}}$, i.e., the thermodynamic potential, is independent of the size of the unpackaged RNA. It has been shown that, under the cellular osmotic pressure, the contribution to the total thermodynamic potential due to the exposed RNA is relatively minor (34).

The thermodynamic potential within the nucleocapsid accounts for the genome-capsid interactions with the participation of salt ions. The corresponding semigrand potential ($F_{\mathrm{cap}}$) is estimated from the molecular

compositions of RNA and CTD tails, the chemical potentials of small ions, and the interaction potentials between those components and the capsid shell,

$$F_{\text{cap}} = \sum_{\alpha=\pm\text{ ions}} \int d\mathbf{r}\rho_\alpha(\mathbf{r})[\ln \rho_\alpha(\mathbf{r}) - 1] + \sum_{\alpha=\pm\text{ ions}} \int d\mathbf{r}\rho_\alpha(\mathbf{r})[\Psi_\alpha(\mathbf{r}) - \mu_\alpha] + \sum_{M=R,T} \int d\mathbf{R}\rho_M(\mathbf{R})\Psi_M(\mathbf{R}) + F^{ex}, \quad (9)$$

where subscripts $\alpha$ and $M$ denote salt ions and macromolecules ($R$ = RNA, $T$ = CTD tail); $d\mathbf{R} = d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_M$ stands for a set of differentials; $\mu_\alpha$ is the chemical potentials of salt ions; $\Psi_M(\mathbf{R}) = \sum_{i=1}^{M} \Psi_p(\mathbf{r}_i)$ represents the external potential for the polyion, while $\Psi_p(\mathbf{r}_i)$ is that of each segment; $\Psi_\alpha(\mathbf{r})$ is the external potential of small salt ions; $\rho_\alpha(\mathbf{r})$ and $\rho_M(\mathbf{R})$ are the density profiles of small ions and polymers, respectively; and $F^{ex}$ is the excess free energy due to intermolecular interactions. The excess free energy is calculated from the classical density functional theory (DFT). We have demonstrated in our earlier works the theoretical performance of the DFT for inhomogeneous polyelectrolyte systems (49,50) and its successful applications to HBV nucleocapsids (34,36,37). The Supporting Material details a self-contained description of the relevant DFT equations used in this work.

## RESULTS AND DISCUSSION

### The optimal genome size of HBV nucleocapsid

A newly formed WT HBV nucleocapsid contains a single copy of pgRNA, which consists of ~3400 nucleotides. It has been shown that the HBV capsid is able to package a similar amount of bacterial ssRNA by expression of the full-length capsid proteins in *E. coli* (51). When the CTD tails are truncated, however, the capsid contains a smaller RNA chain. Fig. 2 shows our theoretical predictions for the packaged RNA size in WT and mutant nucleocapsids with different degrees of CTD truncation. The number of amino-acid (aa) residues in the core protein (183 ~162) is used to denote the WT and C-terminally truncated mutants. The theoretical data for the optimal genome size is compared with mutagenesis studies in *E. coli* systems (7,52).

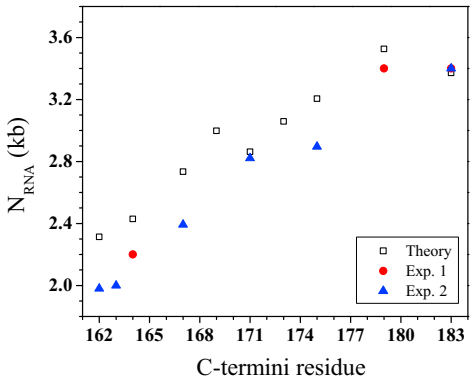


FIGURE 2 (*Black open square*) Theoretical predictions and experimental data (*Exp. 1* (7), *red circle*; *Exp. 2* (52), *blue triangle*) for the optimal RNA length (kilo-base, *kb*) in WT and mutant HBV nucleocapsids. To see this figure in color, go online.

In general, the theoretical predictions correspond well with those from experiments. The predicted values are slightly higher than the experimental data probably because our model determines the optimal genome size without considering the polymerase and host factors, such as hsp or hsc chaperones, kinases, and phosphatase that are packaged in the capsid (53). Although these single-copy proteins make little contribution to the overall charge of the HBV capsid in comparison to the core proteins, they occupy a certain capsid volume influencing the total amount of packaged RNA. Nevertheless, for WT and mutants with low-degree CTD truncations, the theoretical predictions are excellent. By contrast, the calculated RNA size for mutants with a shorter CTD tail (e.g., with C-termini residue of 162 or 164) exhibits more significant deviation (10~20%) from the experimental data. Such a corroborated result confirms that RNA packaging in HBV is mostly regulated by nonspecific interactions between RNA and the CTD tails.

### Charge balance between packaged RNA and CTD tails

It has been speculated that, for ssDNA/RNA viruses in general and HBV in particular, there might be a direct correlation between the net charge of flexible tails of capsid core proteins and the size of a packaged genome (7–9). For many ssRNA/ssDNA viruses, the genome packaging is controlled by the flexible peptide arms of the capsid proteins, and the genome size is determined by a fixed charge ratio between those macromolecules (22). While a positive correlation between the genome size and capsid charge has been widely recognized, a theoretical investigation from 2011 suggests that different charge ratios are possible for the interaction between nucleic acids and capsid proteins (27).

Fig. 3 presents the packaged RNA size ($N_{\text{RNA}}$) as a function of the net charge of CTD tails ($Q_{\text{tail}}$) of the HBV core proteins. Both the theoretical results and experiment data for the optimal genome size (shown in Fig. 2) are referred to $Q_{\text{tail}}$ for WT and mutants. Although there is an overall trend that $N_{\text{RNA}}$ increases with $Q_{\text{tail}}$, neither experimental data nor theoretical predictions yield a fixed charge ratio between RNA and CTD tails ($N_{\text{RNA}}/Q_{\text{tail}}$). For example, two mutants have same $Q_{\text{tail}}$ but noticeably different $N_{\text{RNA}}$ values (e.g., 167 and 169).

A number of thermodynamic models have been proposed to explain the optimal charge ratio between nucleic acids and capsid proteins. For example, Hu et al. (23) argued that the optimal RNA size is dependent on the contour length of peptide tails, and obtained a fixed charge ratio of $N/Q = 2$. Different charge ratios (<1, 1, and 2) have also been predicted with the assumption of a uniform capsid

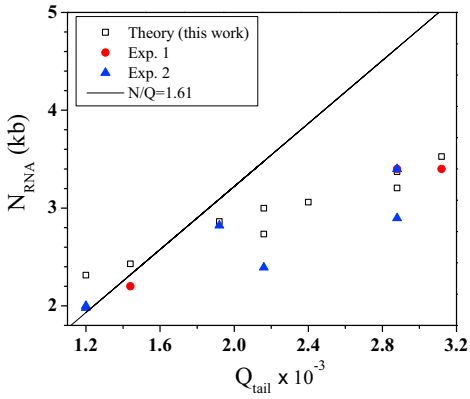


FIGURE 3 Correlation between the packaged RNA size ($N_{RNA}$) and the net charge of C-terminal tails ($Q_{tail}$). (*Solid line*) $N/Q = 1.61$ was proposed by Belyi and Muthukumar (22). To see this figure in color, go online.

charge placed at the inner capsid surface (21,24,25). Most notably, Belyi and Muthukumar (22) determined $N/Q \sim 1.6$ based on the counterion-condensation theory for RNA and peptide tails. The linear charge relation is compared in Fig. 3 together with our model and experimental data. Clearly, neither the WT nor mutant HBV nucleocapsids follow the universal charge ratio. Instead, the DFT calculations show that the charge ratio varies from 1.1 to 1.8 depending on specific mutants. A 2012 experiment also revealed the lack of a universality for the charge ratio on RNA packaging in brome mosaic virus (26). It was found that the number of charges in capsid protein arms is not a sole factor in controlling the RNA packaging. Our DFT calculations on HBV packaging draw a similar conclusion that the additional controlling factors beyond the net charge effect should be considered. As indicated before (34), the excluded volume and correlation effects of macromolecules and salt ions contribute to the RNA packaging process. A proper consideration of such effects allows us to quantify the equilibrium thermodynamics and obtain the optimal size of the packaged genome.

In obtaining the results shown in Fig. 3, we assumed that three serine residues (155, 162, and 170th aa) are phosphorylated for the WT virus and each phosphorylated serine carries a negative charge. Although those positions are most relevant to serine phosphorylation in vivo, it has been reported that serine residues at 176 and 178th aa are also possibly phosphorylated (54). If that is the case, we find the optimal genome size would be reduced by ~170 bases in comparison to that for the WT virus with only three sites of serine phosphorylation. When all five serine sites are phosphorylated, the RNA size is longer (by ~70 bases) than that of the C-terminal truncated mutant 169 with regular phosphorylation, although these mutants have the same net charge ($Q_{tail}$) on the CTD tails. Such a difference is not accounted for by the universal charge ratio or charge-balance hypothesis, but somewhat expected because, as discussed below, the additional phosphorylation sites are located around the C-terminal end region and mostly exposed at the capsid surface.

## Nucleocapsid stability

Earlier experimental work suggests that encapsidation of pgRNA stabilizes HBV nucleocapsids in comparison to empty capsids (51). In addition, the nucleocapsid stability changes during the reverse transcription from RNA to DNA, indicating its potential correlation with viral maturation (55). Moreover, mutagenesis experiments reveal that partial truncation of CTD influences the RNA packaging and DNA synthesis. For example, Le Pogam et al. (7) found that HBV mutants with a truncation of no more than 10 of C-termini (e.g., 173 CP residues) were able to produce a comparable amount of DNA as WT, while further deletion of CTD residues critically impeded the DNA production. Nucleocapsids of those mutants (i.e., >10 residues are truncated from CTD) maintain a reduced stability, and their packaged genome is easily digested by a cellular nuclease.

The stability of HBV nucleocapsids depends on the free energy of complex formation between the RNA/DNA genome and oppositely charged capsid proteins. Approximately, the free energy can be divided into contributions from capsid formation from the protein subunits ($F_{em}$) and that due to the interaction of nucleic acids with the flexible domains of the capsid proteins ($F_{\rm plex}$) (56):

$$F_{NC} = F_{em} + F_{\rm plex}. \tag{10}$$

In Eq. 10, the first term on the right side represents the reversible work to transfer diluted core protein subunits to the assembled empty capsid, and the other is the complexation energy of RNA/DNA genome and CTD segments. The complex formation energy is determined from the free energy differences among the confined macromolecules, the isolated genome, and CTDs chains at infinite dilution. Previously, we proposed an approximate complexation free energy to account for the changes in the chain elastic energy, the molecular excluded volume effects, the electrostatic interactions among macromolecules and small ions, and the entropy of mixing (56). Among various HBV mutants, the free energy of empty capsid is same for all cases, and thus we only compare the free energy of chain complexation ($F_{\rm plex}$) for mutants with different lengths of the CTD tail.

Fig. 4 shows the complex formation free energy versus the CTD chain length for those HBV mutations considered above. Here a negative value means that encapsidation of macromolecules is energetically favorable, i.e., stabilizes the nucleocapsid. As expected, $F_{\rm plex}$ is negative for WT nucleocapsids (CP183) and for mutants with a small-degree truncation (CP179 and CP173). In other words, the corresponding nucleocapsids are more stable than the empty capsid. However, the complexation free energy is positive for those with a large degree of CTD truncation. We note

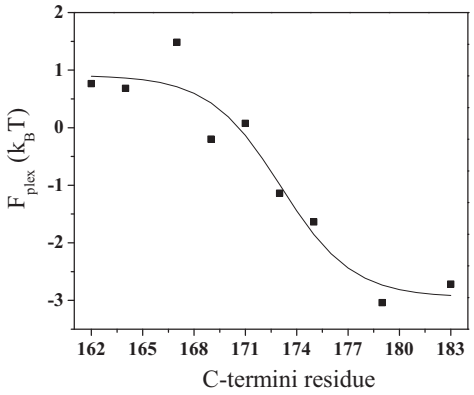


FIGURE 4 The free energy of RNA encapsidation depends on the CTD length. The complexation free energy ($F_{plex}$) of RNA and CTD chains is determined for WT and mutants. (*Sigmoidal line*) Guide to the eye.

a sigmoidal change of the complexation energy with regard to the size of CTD tails as shown by the solid line in Fig. 4. The transition occurs at approximately CP173 to CP171, where $F_{plex}$ changes from a negative to a positive value.

As mentioned before, earlier experiments indicate that nucleocapsids with a truncation up to 10 C-termini residues (CP173) are relatively stable. We may rationalize such a stability variation according to our coarse-grained model. Our theoretical results indicate that encapsidation of RNA-CTD chains entails a greater amount of unfavorable positive free energy for higher-order truncations than for the case of CP173 (e.g., CP171). Accordingly, the population of those mutants with unfavorable RNA encapsidation should be relatively small, and the resultant DNA synthesis should be substantially constrained.

## Structure of HBV nucleocapsids

Whereas cryo-EM images are available for both WT and mutant HBV nucleocapsids (34), the molecular structures for the flexible components of the virus remain poorly understood. With CTD tails fully confined inside the capsid, our previous DFT calculation predicted a thin layer of RNA chain and a brushlike distribution of the CTD segments. Here we conduct similar structural analyses by expanding the computation range from inside the capsid to the outside area. In addition, more detailed capsid features are implemented such that CTD segments are set to be exposable into the outside region. It has been shown that such semipermeable characters are relevant to build the HBV capsid model (36).

Fig. 5 presents the radial distributions of RNA and CTD segments at different levels of CTD truncation. Here the RNA content is selected from the optimal genome size calculated from DFT. As indicated in a previous report (34), RNA segments exhibit extremely high contact density at the inner capsid surface. Immediately following the RNA layer are tethered CTD segments, with the rest forming a brushlike

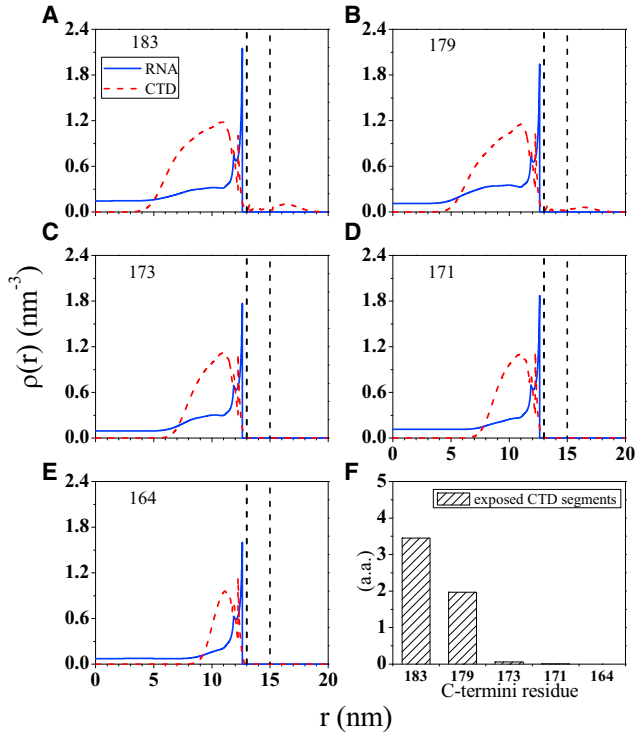


FIGURE 5 (*A–E*) Radial distributions of RNA (*blue solid line*) and CTD chains (*red dashed line*) in WT and mutant HBV nucleocapsids (57). (*r*) Radial distance from the capsid center; (*perpendicular dashed lines*) position of the capsid shell. (*F*) The average number of amino-acid (aa) residues from the CTD tails located outside the capsid wall for WT and mutants. To see this figure in color, go online.

structure. For mutants with truncation over five residues, we find the radial distributions of capsid components are virtually identical to our previous predictions from a confined capsid model (34). Noticeably, however, some CTD segments approach outside the capsid region for both WT and the CP179 mutant. Fig. 5, right side, compares the average number of exposed CTD residues for different mutants. For WT nucleocapsids, 3–4 CTD residues per capsid protein are exposed in the outside region. Because these residues do not participate in the complex formation with encapsidated RNA, we expect ~10% of CTD segments are not essential for the genome packaging. The exposure of CTD tails in WT HBV virus is consistent with previous biochemical analysis showing their accessibility to proteases and antibodies (58,59) and recent cryo-electron microscopy structures (31,60). Because our theoretical calculations assume a spherical symmetry for CTD distributions, we are not able to distinguish different holes of the capsid shell responsible for the CTD exposure.

Our theoretical results for the multiregional distributions of CTD segments inside and outside the capsid may help to explain how the CTD tails fulfill multiple roles in the HBV lifecycle. As discussed above, CTDs are indispensable for genome encapsidation and replication. Other roles

have also been identified, such as mediating the nuclear localization and maturation signal (30–33,61–63). Specifically, it has been noted that the arginine-rich domains of each CTD tail play a critical role in interaction with the capsid exterior, mediating nuclear import and export signals (32,33,63). The location of CTD tails could regulate the posttranslational process. Our model suggests that the WT HBV capsids are optimized to fulfill pgRNA packaging while maintaining a portion of exposed CTD segments to conduct a potent signaling role. Mutant CP179 has $<2$ residues per CTD tail exposed, and such exposure is not observed for mutants with shorter CTDs. Consequently, those mutants may lose the capability to deliver the maturation signal.

## CONCLUSIONS

We have investigated the thermodynamics of HBV nucleocapsid formation based on a coarse-grained molecular model for the key viral components and the intracellular electrolyte solution. The genome contents in stable nucleocapsids are determined following an equilibrium criterion that minimizes the thermodynamic potential for RNA encapsidation. The theoretical results are found in good agreement with existing experiments on the nucleocapsid formation for both WT and mutant HBV capsids with truncated C-termini of the core proteins. The thermodynamic model confirms that pgRNA encapsidation is mainly dominated by nonspecific electrostatic and excluded-volume interactions among oppositely charged macromolecules. In contrast to previous investigations, both experimental and theoretical results indicate that the RNA content is not linearly correlated with the net charge of the CTD tails.

This model is an extension of our previous work on electrostatic regulation of HBV replication. One key difference is that this model allows for the partial exposure of CTD tails outside the capsid surface, while the earlier work assumes a full confinement of CTD tails (34). The exposure of the CTD tails had been speculated upon for decades, but only recently confirmed by experimental and theoretical studies. We find that only ~10% of CTD segments are distributed outside the capsid in the WT HBV virus, and the CTD extension is insignificant for most mutants with truncated CTD tails. In a WT capsid, the majority of CTD segments are located near the RNA segments around the inner capsid shell, while 3–4 residues from each CTD tail are exposed outside the HBV capsid. To a certain degree, this work provides a support of the capsid confinement effect hypothesized in our earlier work. Indeed, the optimal size of packaged RNA predicted from this work is similar to that from our previous model.

In addition to RNA-capsid charge correlations, this work examines phosphorylation effects and capsid stability. Furthermore, we have investigated the spatial distributions of RNA and CTD segments in both WT and mutant capsids and how the location of CTD tails is optimized to fulfill the multiple roles in assembly, maturation, and envelopment of HBV capsids. The thermodynamic stability of a nucleocapsid is affected not only by the overall charge balance between oppositely charged macromolecules, but also by the molecular excluded volume effects and the charge distribution within the CTD tails. The RNA encapsidation free energy indicates that nucleocapsids with shortened C-termini are less stable in comparison to the WT capsid. The reduction in stability explains why mutants with higher degrees of truncation become easily degradable and fail to maintain its capability for DNA synthesis. These theoretical predictions help us to better understand the molecular details of the HBV replication cycle and develop more effective antiviral strategies.

## SUPPORTING MATERIAL

Supporting Materials and Methods are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00856-5.

## AUTHOR CONTRIBUTIONS

J.W. designed research; J.K. performed the DFT calculations; and J.K. and J. W. analyzed data and wrote the article.

## ACKNOWLEDGMENTS

The authors are grateful to an anonymous reviewer for extensive comments on the biological implications of our theoretical results.

The authors are grateful to the National Science Foundation for financial support (under grant No. NSF-CBET-1404046).

## SUPPORTING CITATIONS

References (64–73) appear in the Supporting Material.

## REFERENCES

1. Ganem, D., and A. M. Prince. 2004. Mechanisms of disease: hepatitis B virus infection—natural history and clinical consequences. *N. Engl. J. Med.* 350:1118–1129.
2. Seeger, C., and W. S. Mason. 2000. Hepatitis B virus biology. *Microbiol. Mol. Biol. Rev.* 64:51–68.
3. Locarnini, S., and W. S. Mason. 2006. Cellular and virological mechanisms of HBV drug resistance. *J. Hepatol.* 44:422–431.
4. 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.
5. Köck, J., M. Nassal, …, F. von Weizsäcker. 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.
6. Zlotnick, A., N. Cheng, …, P. T. Wingfield. 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.
7. Le Pogam, S., P. K. Chua, …, 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.* 79:1871–1887.

8. Chua, P. K., F.-M. Tang, …, C. Shih. 2010. Testing the balanced electrostatic interaction hypothesis of hepatitis B virus DNA synthesis by using an in vivo charge rebalance approach. *J. Virol.* 84:2340–2351.
9. Newman, M., P. K. Chua, …, C. Shih. 2009. Testing an electrostatic interaction hypothesis of hepatitis B virus capsid stability by using an in vitro capsid disassembly/reassembly system. *J. Virol.* 83:10616–10626.
10. Liu, S., J. He, …, J. Zhang. 2010. Structural comparisons of hepatitis B core antigen particles with different C-terminal lengths. *Virus Res.* 149:241–244.
11. Bringas, R. 1997. Folding and assembly of hepatitis B virus core protein: a new model proposal. *J. Struct. Biol.* 118:189–196.
12. Dong, X. F., P. Natarajan, …, A. Schneemann. 1998. Particle polymorphism caused by deletion of a peptide molecular switch in a quasiequivalent icosahedral virus. *J. Virol.* 72:6024–6033.
13. Kaplan, I. B., L. Zhang, and P. Palukaitis. 1998. Characterization of cucumber mosaic virus. V. Cell-to-cell movement requires capsid protein but not virions. *Virology.* 246:221–231.
14. Marshall, D., and A. Schneemann. 2001. Specific packaging of nodaviral RNA2 requires the N-terminus of the capsid protein. *Virology.* 285:165–175.
15. Schmitz, I., and A. L. N. Rao. 1998. Deletions in the conserved amino-terminal basic arm of cucumber mosaic virus coat protein disrupt virion assembly but do not abolish infectivity and cell-to-cell movement. *Virology.* 248:323–331.
16. Schneemann, A., and D. Marshall. 1998. Specific encapsidation of nodavirus RNAs is mediated through the C-terminus of capsid precursor protein $\alpha$. *J. Virol.* 72:8738–8746.
17. Scott, K. P., M. J. Farmer, …, A. F. Murant. 1996. Comparison of the coat protein of groundnut rosette assistor virus with those of other luteoviruses. *Ann. Appl. Biol.* 128:77–83.
18. Smith, T. J., E. Chase, …, K. L. Perry. 2000. The structure of cucumber mosaic virus and comparison to cowpea chlorotic mottle virus. *J. Virol.* 74:7578–7586.
19. Tihova, M., K. A. Dryden, …, A. Schneemann. 2004. Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. *J. Virol.* 78:2897–2905.
20. Hagan, M. F. 2014. Modeling viral capsid assembly. *Adv. Chem. Phys.* 155:1–68.
21. Angelescu, D. G., R. Bruinsma, and P. Linse. 2006. Monte Carlo simulations of polyelectrolytes inside viral capsids. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 73:041921.
22. Belyi, V. A., and M. Muthukumar. 2006. Electrostatic origin of the genome packing in viruses. *Proc. Natl. Acad. Sci. USA.* 103:17174–17178.
23. Hu, T., R. Zhang, and B. I. Shkovskii. 2008. Electrostatic theory of viral self-assembly. *Phys. A. Stat. Mech. Appl.* 387:3059–3064.
24. Siber, A., and R. Podgornik. 2008. Nonspecific interactions in spontaneous assembly of empty versus functional single-stranded RNA viruses. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 78:051915.
25. 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.
26. Ni, P., Z. Wang, …, C. C. Kao. 2012. An examination of the electrostatic interactions between the N-terminal tail of the brome mosaic virus coat protein and encapsidated RNAs. *J. Mol. Biol.* 419:284–300.
27. Ting, C. L., J. Wu, and Z.-G. Wang. 2011. Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation. *Proc. Natl. Acad. Sci. USA.* 108:16986–16991.
28. Perlmutter, J. D., C. Qiao, and M. F. Hagan. 2013. Viral genome structures are optimal for capsid assembly. *eLife.* 2:e00632.
29. Yoffe, A. M., P. Prinsen, …, A. Ben-Shaul. 2008. Predicting the sizes of large RNA molecules. *Proc. Natl. Acad. Sci. USA.* 105:16153–16158.
30. Rabe, B., A. Vlachou, …, M. Kann. 2003. Nuclear import of hepatitis B virus capsids and release of the viral genome. *Proc. Natl. Acad. Sci. USA.* 100:9849–9854.
31. Yu, X., L. Jin, …, Z. H. Zhou. 2013. 3.5 Å cryoEM structure of hepatitis B virus core assembled from full-length core protein. *PLoS One.* 8:e69729.
32. Li, H.-C., E.-Y. Huang, …, C. Shih. 2010. Nuclear export and import of human hepatitis B virus capsid protein and particles. *PLoS Pathog.* 6:e1001162.
33. Yang, C.-C., E.-Y. Huang, …, C. Shih. 2014. Nuclear export of human hepatitis B virus core protein and pregenomic RNA depends on the cellular NXF1-p15 machinery. *PLoS One.* 9:e106683.
34. Jiang, T., Z.-G. Wang, and J. Wu. 2009. Electrostatic regulation of genome packaging in human hepatitis B virus. *Biophys. J.* 96:3065–3073.
35. van der Schoot, P., and R. Zandi. 2013. Impact of the topology of viral RNAs on their encapsulation by virus coat proteins. *J. Biol. Phys.* 39:289–299.
36. Kim, J., and J. Wu. 2014. A theoretical study of SRPK interaction with the flexible domains of hepatitis B capsids. *Biophys. J.* 107:1453–1461.
37. Meng, D., R. P. Hjelm, …, J. Wu. 2011. A theoretical model for the dynamic structure of hepatitis B nucleocapsid. *Biophys. J.* 101:2476–2484.
38. Zhang, D., R. Konecny, …, 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.
39. Kolinski, A., A. Godzik, and J. Skolnick. 1993. A general method for the prediction of the three dimensional structure and folding pathway of globular proteins: application to designed helical proteins. *J. Chem. Phys.* 98:7420–7433.
40. Angelescu, D. G., and P. Linse. 2008. Modelling of icosahedral viruses. *Curr. Opin. Colloid Interface Sci.* 13:389–394.
41. Elrad, O. M., and M. F. Hagan. 2010. Encapsulation of a polymer by an icosahedral virus. *Phys. Biol.* 7:045003.
42. Hagan, M. F., and D. Chandler. 2006. Dynamic pathways for viral capsid assembly. *Biophys. J.* 91:42–54.
43. Petrov, A. S., and S. C. Harvey. 2007. Structural and thermodynamic principles of viral packaging. *Structure.* 15:21–27.
44. 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.
45. 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.
46. 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.
47. Jiang, H., and S. X. Sun. 2013. Cellular pressure and volume regulation and implications for cell mechanics. *Biophys. J.* 105:609–619.
48. Tinevez, J.-Y., U. Schulze, …, E. Paluch. 2009. Role of cortical tension in bleb growth. *Proc. Natl. Acad. Sci. USA.* 106:18581–18586.
49. Jiang, T., Z. Li, and J. Wu. 2007. Structure and swelling of grafted polyelectrolytes: predictions from a nonlocal density functional theory. *Macromolecules.* 40:334–343.
50. Li, Z., and J. Wu. 2006. Density functional theory for polyelectrolytes near oppositely charged surfaces. *Phys. Rev. Lett.* 96:048302.
51. Porterfield, J. Z., M. S. Dhason, …, A. Zlotnick. 2010. Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. *J. Virol.* 84:7174–7184.
52. Sominskaya, I., D. Skrastina, …, P. Pumpens. 2013. A VLP library of C-terminally truncated hepatitis B core proteins: correlation of RNA encapsidation with a Th1/Th2 switch in the immune responses of mice. *PLoS One.* 8:e75938.

53. Bartenschlager, R., and H. Schaller. 1992. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J.* 11:3413–3420.

54. Daub, H., S. Blencke, …, M. Cotten. 2002. Identification of SRPK1 and SRPK2 as the major cellular protein kinases phosphorylating hepatitis B virus core protein. *J. Virol.* 76:8124–8137.

55. Cui, X., L. Ludgate, …, J. Hu. 2013. Maturation-associated destabilization of hepatitis B virus nucleocapsid. *J. Virol.* 87:11494–11503.

56. Kim, J., and J. Wu. 2014. A molecular thermodynamic model for the stability of hepatitis B capsids. *J. Chem. Phys.* 140:235101.

57. Huang, Y.-M., M. Kang, and C. E. Chang. 2012. Mechanistic insights into phosphopeptide—BRCT domain association: preorganization, flexibility, and phosphate recognition. *J. Phys. Chem. B.* 116:10247–10258.

58. Gallina, A., F. Bonelli, …, G. Milanesi. 1989. A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. *J. Virol.* 63:4645–4652.

59. Seifer, M., and D. N. Standring. 1994. A protease-sensitive hinge linking the two domains of the hepatitis B virus core protein is exposed on the viral capsid surface. *J. Virol.* 68:5548–5555.

60. Wang, J. C. Y., M. S. Dhason, and A. Zlotnick. 2012. Structural organization of pregenomic RNA and the carboxy-terminal domain of the capsid protein of hepatitis B virus. *PLoS Pathog.* 8:e1002919.

61. Nassal, M. 1992. The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *J. Virol.* 66:4107–4116.

62. Ning, X., D. Nguyen, …, J. Hu. 2011. Secretion of genome-free hepatitis B virus—single strand blocking model for virion morphogenesis of para-retrovirus. *PLoS Pathog.* 7:e1002255.

63. Schmitz, A., A. Schwarz, …, M. Kann. 2010. Nucleoporin 153 arrests the nuclear import of hepatitis B virus capsids in the nuclear basket. *PLoS Pathog.* 6:e1000741.

64. Roth, R., R. Evans, …, G. Kahl. 2002. Fundamental measure theory for hard-sphere mixtures revisited: the White Bear version. *J. Phys. Cond. Matter.* 14:12063–12078.

65. Yu, Y. X., and J. Z. Wu. 2002. Structures of hard-sphere fluids from a modified fundamental-measure theory. *J. Chem. Phys.* 117:10156–10164.

66. 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.

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

68. Jiang, J. W., L. Blum, …, J. M. Prausnitz. 2001. Thermodynamic properties and phase equilibria of charged hard sphere chain model for polyelectrolyte solutions. *Mol. Phys.* 99:1121–1128.

69. Jiang, J. W., H. L. Liu, …, J. M. Prausnitz. 1998. A molecular-thermodynamic model for polyelectrolyte solutions. *J. Chem. Phys.* 108:780–784.

70. Li, Z., and J. 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.

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

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

73. Hoye, J. S., and L. Blum. 1978. The mean spherical model for asymmetric electrolytes: thermodynamics and the pair correlation function. *Mol. Phys.* 35:299–300.