UNIVERSITY OF CALIFORNIA
RIVERSIDE

Molecular Models for the Assembly and Replication of Hepatitis B Virus

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemical and Environmental Engineering

by

Jehoon Kim

March 2015

Dissertation Committee:
Dr. Jianzhong Wu, Chairperson
Dr. Xin Ge
Dr. Chia-en A. Chang

The Dissertation of Jehoon Kim is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

I wish to express my utmost gratitude to many people for their great support and assistance over the years. First and foremost, I would like to thank my advisor Prof. Jianzhong Wu. I appreciate his sincere support in my PhD research with superb and patient guidance. From him, I have learned the ways of designing theoretical research where I can observe and analyze the human biology. He has planted in my mind a strong motivation to learn about molecular modeling in biophysics. He also has been tremendous mentor on my career. His strong enthusiasm for the research was motivational for me. I am thankful for the excellent example he has provided as a successful chemical engineer and professor.

I sincerely appreciate my PhD committee, Prof. Xin Ge, Prof. Chia-en A. Chang and Prof. Roya Zandi for their great advice and helping me in qualifying exam and defense. I also wish to express my special thanks to Dr. Shuangliang Zhao, Dr. Zhehui Jin, Dr. Ke Wang, Dr. Yu Liu and Dr. Jia Fu for their practical and valuable support during my PhD study. I also thank the good friends that I have made in UC Riverside for their encouragement throughout my time in graduate school.

I am always grateful to my parents and brother for their endless support and heartfelt care. I would also like to express my gratitude to my family in-law for their love and encouragement. Finally, I want to thank my son Matthew Danwoo Kim and my wife Chung Hee Moon, who supported and encouraged me more than anyone else. Thank you for everything.

In this dissertation, some parts are the reprint of the materials appeared in Journal of Chemical Physics (Volume 140, Issue 23, Page 235101, June 2014), and Biophysical Journal (Volume 107, Issue 6, Page 1453-1461, September 2014). This work is financially sponsored by the US Department of Energy (DOE) (DE-FG02-06ER46296), the National Science Foundation (NSF-CBET-0852353, NSF-CBET-1404046), and the National Institute of Health (R21-AI077532). This work utilizes supercomputer from the National Energy Research Scientific Computing Center (NERSC).

ABSTRACT OF THE DISSERTATION

Molecular Models for the Assembly and Replication of Hepatitis B Virus

by

Jehoon Kim

Doctor of Philosophy
Graduate Program in Chemical and Environmental Engineering
University of California, Riverside, March 2015
Professor Jianzhong Wu, Chairperson

Hepatitis B virus (HBV) infects more than 2 billion people alive today and is responsible for over 1 million deaths caused by acute and chronic hepatitis and hepatocellular carcinoma every year. Although remarkable progress has been made in recent years, effective treatment and eradication of chronic HBV infection remain as a tremendous therapeutic challenge. Experimental investigations over the past few decades have resulted in vast information on the characteristic features of viral replication and molecular structures of the HBV genome as well as the viral proteins. However, it has long been recognized that the viral replication process is regulated by a balanced interaction of the geometric material and its viral proteins in a crowded cellular milieu that is hardly detectable with direct experimental means. Molecular modeling that provides a quantitative analysis of important intermolecular interactions will be invaluable for unraveling the microscopic details of physicochemical processes

underlying viral formation and replication cycle and for designing new experiments to develop novel anti-viral strategies.

This Ph.D. research focuses on molecular modeling of capsid formation, genome packaging and maturation of HBV particles under various mutagenesis and physiological conditions. A suite of theoretical models has been developed to facilitate better understanding of the fundamentals of viral replication cycles and enabling new therapeutic breakthrough for future treatment of HBV infection. Specifically, I applied an effective coarse-grained model for the key viral components and quantified the stability of nucleocapsids based on statistical mechanics. The theoretical model yields faithful results for describing the effects of temperature and ion concentration on viral particle formation, supporting an experimentally derived hypothesis that suggests the balanced electrostatic interaction for the capsid stability and genome encapsidation. In addition, I analyzed the thermodynamic basis for the viral genome packaging, the dynamic structures of nucleocapsids at different stages of the replication cycle, and the correlation between the viral structure to the maturing signals of HBV nucleocapsids by using the classical density functional theory (DFT). The DFT analysis provides a quantitative description of the microscopic structure of the protein-RNA/DNA complex underlying nucleocapsid formation and the related thermodynamic properties. The theoretical predictions on the optimal genome length and nucleocapsid structure are in good agreement with available experiments for the wild type HBV and mutants with truncated C-terminal domain of the capsid proteins. To establish concrete connection between the capsid structure and HBV maturing signals, I further investigated molecular recognition

between viral envelope and capsid proteins. Specific protein-protein interactions were identified through molecular docking and molecular dynamic (MD) simulations. It is expected accomplishments from this work will contribute to broadening a fundamental understanding of the HBV replication cycle.

Table of Contents

Title Page

Copyright Page

Approval Page

Acknowledgements iv

Abstract of the dissertation vi

Chapter 1. Introduction 1

1. 1 Background of HBV 1

1. 2 Dissertation organization 6

Chapter 2. Basic Concepts of Classical Density Functional Theory 10

2. 1 Density Profile 10

2. 2 Density Functional Theory 12

2. 3 Intrinsic Helmholtz Free Energy 16

2. 4 Euler-Lagrange Equation 16

2. 5 Excess Helmholtz Free Energy 18

2. 6 Numerical method 23

Chapter 3. A molecular thermodynamic model for the stability of hepatitis B capsids ... 26

Chapter 4. A theoretical study on SRPK interaction with the flexible domains of Hepatitis B capsids 69

Chapter 5. A thermodynamic model for the genome packaging in hepatitis B virus 92

Chapter 6. Molecular recognition in envelopment of hepatitis B virus 115

Chapter 7. Thermodynamic and structural evidences for reduced hydrogen bonding among water molecules near small hydrophobic solutes 132

Chapter 8. Conclusions 156

References 160

List of Figures

Figure 1 9

Figure 3-1(A),(B) 58

Figure 3-1(C) 59

Figure 3-2 60

Figure 3-3 61

Figure 3-4 62

Figure 3-5 63

Figure 3-6 64

Figure 3-7 65

Figure 3-8 66

Figure 3-9 67

Figure 4-1 86

Figure 4-2 87

Figure 4-3 88

Figure 4-4 89

Figure 4-5 90

Figure 4-6 91

Figure 5-1(A) 109

Figure 5-1(B) 110

Figure 5-2 111

Figure 5-3 112

Figure 5-4 113

Figure 5-5 114

Figure 6-1 126

Figure 6-2 127

Figure 6-3 128

Figure 6-4 129

Figure 6-5 130

Figure 7-1 149

Figure 7-2 150

Figure 7-3 151

Figure 7-4 152

Figure 7-5 153

Figure 7-6 154

List of Tables

Table 3-1 68

Table 6-1 131

Table 7-1 155

Chapter 1. Introduction

1. 1 Background of HBV

Hepatitis B virus (HBV) causes acute and chronic liver diseases such as hepatocellular carcinoma[1, 2]. The estimated number of HBV carriers is about 350 million worldwide. Although HBV vaccine was developed decades ago and there have been tremendous progress in understanding the biological basis of the viral replication cycle, current medical treatments are unable to eliminate the HBV infection[3].

As shown in Figure 1, a HBV virion (i.e. Dane particle) has a double layered structure with an inner icosahedral capsid core and an outer proteolipid envelope. The capsid is formed by 180 or 240 copies of the core protein (HBc) arranged in a icosahedral structure with a triangulation (T) number 3 or 4. Virions with the T4 capsid are the major form under the physiological condition. A T4 virion has an inner capsid shell of 28 nm in diameter and an outer envelope shell of 44 nm in diameter[4-6].

A wild-type core protein (HBc) consists of 183 amino-acid (aa) residues, with an assembly domain (1-140 aa), a linker (141-149), and the C-terminal domain (CTD, 150-183 aa). The assembly domain has a rigid structure and contributes to the construction of the capsid scaffold. Both the linker and the CTD are flexible; the latter is often termed as a 'protamine tail' because it riches in arginine residues[7, 8]. Because of the strong electrostatic interactions between the CTD tails and oppositely-charged genomic materials, HBV nucleocapsids are thermodynamically more stable than empty capsids[9, 10]. If the CTD is truncated by mutation, the CPs are unable to encapsidate the full content

of the pregenomic (pg) RNA, producing capsids with only a smaller amount of DNA or empty capsids[7].

CTD has three phosphorylation sites (S155, S162 and S170), which are important for regulating the HBV life cycle[7, 11]. It has been shown that the CTD chains are phosphorylated at the early stage of viral replication (with pgRNA or intermediate DNA)[12], but are dephosphorylated at the matured stage (with dsDNA)[13, 14]. Replacing those serine residues with charged amino acids to mimic the transient phosphorylation state affects pgRNA encapsidation and DNA synthesis[15-18].

The CTD chains are also know to play an important role in viral replication because only phosphorylated capsids were allowed to transport into cellular nucleus[19]. Whereas the assembly domain of HBV capsids has a rigid structure, the distribution of CTD chains is highly dynamic and changing in response to the evolving genome[19-24].

The capsid shell is fenestrated with symmetrical pores that allow mass transport and the extrusion of the CTD tails[25-27]. Apparently, the distribution of the CTD tails plays a key role in genome packaging, nuclear entry, capsid envelopment and egress. In addition to pgRNA packaging and reverse DNA polymerization[7, 8], it has been recognized that the CTD tails are involved in signaling capsid maturation[22, 28, 29].

The viral envelope is composed of the surface proteins (HBs) imbedded in a lipid bilayer membrane. The glycoprotein expresses in three forms (small, middle and large), differing in terms of the extended N-terminal domains (Pre-S1 or Pre-S2) after a common S domain. The small form (S-HBs) has 226 amino-acid (aa) residues, which are associated with lipid having four trans-membrane domains. Meanwhile, an additional

Pre-S2 (55 aa) exists in the middle form (M-HBs), and Pre-S1 (108 aa) plus Pre-S2 extension in the large surface protein (L-HBs). Such Pre-S domains are flexible polypeptide chains, not imbedded in the lipid membrane. The S and L form of HBs are required for envelopment and infectivity but the M-HBs is not essential for virion formation[27, 30].

In addition to the virion, HBV replication results in protein particles secreted from the infected cells. The so-called subviral particles (SVPs) have either a spherical or filamentous shape of variable lengths with 22 nm in diameter. SVP mainly consists of lipid and envelope proteins devoid of the capsid core[31]. In the serum of a HBV patient, SVP outnumbers the virion particles. However, the biological function of SVP remains little known. It has been suggested that SVP contributes to persistent viral infection by enhancing the immune tolerance[32]. A spherical SVP is predominately composed of S-HBs, but a tubular SVP is rich in L-HBs[33]. SVPs may induce immunological responses the same as the virion, and thus they can be used as effective immunogens for HBV treatments. The first human vaccine for HBV was developed from expression of S-HBs particles in yeast[34].

A critical step in HBV replication entails nucleocapsid binding with a pre-Golgi membrane before undergoing the secretory pathway[32, 35]. It was shown that L-HBs play a critical role in envelope formation[36], as well as in signaling subsequent interactions with the receptors of hepatocytes[37]. The Pre-S domain located either at inside or outside of the viral membrane. It has been suggested that the internal Pre-S domain mediates the interaction of L-HBs with cytosolic nucleocapsids[38, 39]. Mutagenesis studies also

indicated candidate residues in HBc that interact with the envelope. Those residues are located at the tip of the HBc spike[40], or around the base of the 4-helix spike where the C-terminus residues are distributed close to the capsid pores[41]. Many questions remain on how the envelope and the capsid are associated in a HBV virion. One of the key questions is whether there are specific protein-protein interactions between HBs and HBc. Although electron microscopy analyses on the HBV virion point to a non-icosahedral arrangement of HBs in the envelope[42], Seitz et al revealed distinct packaging rule of HBs, implying specific interactions of the envelope with the capsid[43].

The pregenomic(pg) RNA is packaged into the lumen of capsid at the early stage of HBV nucleocapsid formation. It is within the capsid where the pg-RNA is converted into first a single-stranded (ss) DNA and then into a partially double-stranded (ds) DNA. Capsids containing the evolving genome at any stage can be found in hepatocyte cells. Interestingly, the secreted virion particles contain only matured nucleocapsids with a partially dsDNA or empty capsids[44, 45]. It has been hypothesized that immature capsids, i.e., those capsids with RNA or ssDNA chains, are not able to develop into a virion because the viral genome maturation induces a change in the capsid structure that signals envelopment[46]. This hypothesis corroborates at least in part previous mutagenesis experiments. It has been shown that the DNA synthesis and capsid envelopment were blocked by C-terminal truncation of HBV core proteins[47], and that an inhibition of the reverse-transcriptase prohibits the maturation of nucleocapsids and their secretion[48, 49]. It has been suggested that the maturation signal arises from a structural difference in capsids with RNA and dsDNA[50]. HBV maturation may also be affiliated with the

phosphorylation state of capsid proteins[13, 51, 52]. For example, a point mutation of isoleucine to leucine at 97th amino-acid (aa) residue in the core protein (I97L) allows the envelopment of the immature capsids[53]. The I97L mutation changes the capsid conformation, exposing the envelopment signal at the early evolving stage[50]. It is also known that the additional mutation of proline to threonine at 130th aa (P130T) of the capsid proteins inhibits the premature signaling by I97L, and restores into wide-type maturation signal of HBV capsids[54].

From a therapeutic perspective, the HBV-polymerase has been well-recognized target for inhibiting the replication process, and several nucleoside-analogs have been utilized for treatment of HBV infection[55-58]. On the one hand, tremendous on-going viral studies aim to discover new polymerase inhibitors with increased the drug efficacy and resistance (e.g.,[59, 60]). On the other hand, there are numerous theoretical approaches to developing alternative anti-viral strategies, such as improving the vaccine treatment and immunogenic therapy(references?).

The viral capsid assembly has been investigated extensively in the literature, mostly through atomistic simulations (e.g.,[61, 62]). The theoretical analysis is mostly limited into a certain range of parameters due to the requirement of large number of atoms and a long simulation time for representing the assembly pathway. One way to deal with such a limitation is by adoption of coarse-grained approaches and continuum modeling methods[63-65]. The semi-empirical methods can be used to rationalize the assembly thermodynamics in accordance with experimental measurements[66-69]. In addition, there have been a number of modeling efforts focused on genome packaging[70-

[74], or the packaged genome conformation[75, 76]. In addition to thermodynamic properties, the coarse-grained models are applicable to study interesting dynamic properties of viral capsids[62, 64, 65, 77] and viral maturation[78-80]. In stark contrast to the experimental literature, theoretical investigations are often concerned with the generic nature of viral self-assembly rather than a specific virus like HBV. Regrettably, a general perspective provides little HBV specific information.

Several theoretical studies have contributed to understanding of the assembly thermodynamics and kinetics for the formation of empty capsids for HBV, [63, 81, 82]. In particular, the empty capsid models provides a good starting point to investigate the viral assembly process. Besides, it provides useful information to find inhibitors of viral assembly[83-85].

1. 2 Dissertation organization

This dissertation is focused on the theoretical investigation of thermodynamic properties of HBV nucleocapsids such as the stability, basis of genome packaging, structural dynamics in replication cycles, the capsid-envelope interactions and the correlation to the maturation transition. The eventual goal is not only on the formation of empty capsids, but also on the formulation of effective nucleocapsid models that allows to evaluate the entire genome mediated viral processes.

Chapter 2 covers the basic formulism of the classical density functional theory (DFT). Special attention is given to the construction of excess Helmholtz energy functional for the polyelectrolyte system considered in this work, including those related

to the hard-sphere repulsion, the chain connectivity, the direct Coulomb interaction, and the electrostatic correlations.

Chapter 3 proposes a molecular thermodynamic model for predicting the stability of Hepatitis B virus (HBV) capsids either with or without nucleic acids. Based on coarse-grained models for the key viral components, the stability of the viral particles is quantified in response to physiological cellular conditions with an explicit consideration of the hydrophobic association of capsid subunits, the electrostatic interactions, the conformational entropy and the excluded volume effects of all biomolecular species.

Chapter 4 presents the microscopic model for the dynamic structure of HBV nucleocapsids. The DFT calculations provide useful information on the radial distributions of individual amino-acid residues of the CTD tails for both the empty and RNA-containing capsids . The theoretical data correspond well with recent experimental studies on the transient CTD distribution.

Chapter 5 presents a thermodynamic basis of the pregenomic (pg) RNA encapsidation in HBV. The electrostatic regulation process in genome packaging is examined with detailed consideration of capsid properties including the accessibility of C-terminal domains (CTD) of the core proteins into outside capsid. The optimal RNA amount in nucleocapsids with variant CTD length is predicted with explicit consideration of all nonspecific interactions among key viral components,

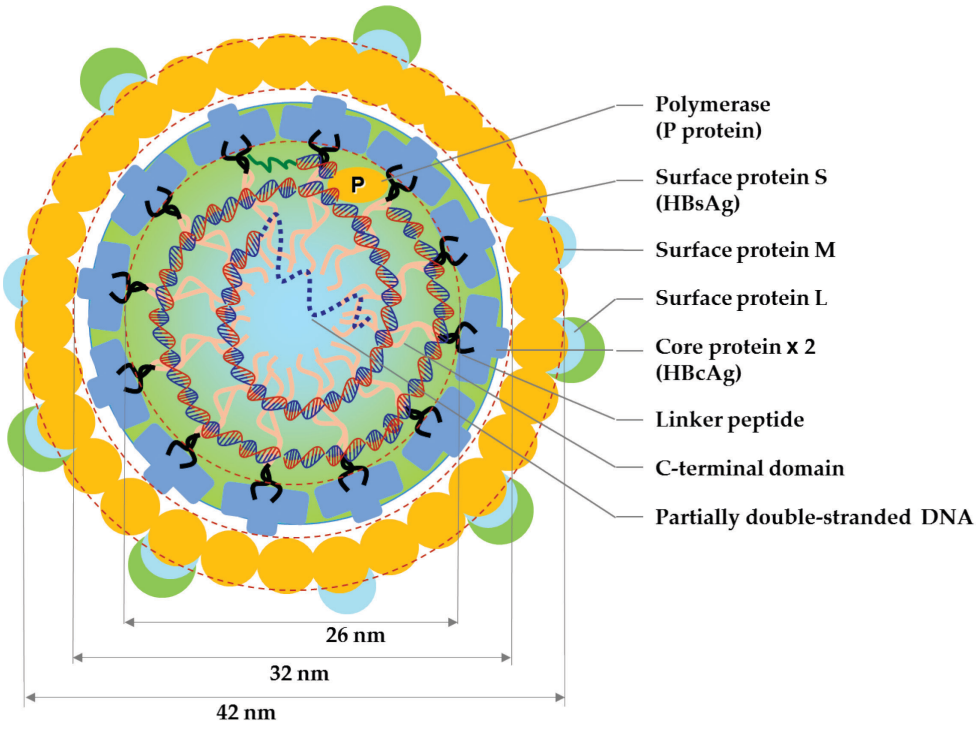
Chapter 6 investigates the capsid core - envelope association. The specific interaction between Pre-S domain of large surface protein (L-HBs) and the core protein (HBc) of HBV is determined by MD simulation and molecular docking analysis. The

binding site is identified for the interaction between HBc and a peptide fragment of L-HBs related with the molecular recognition, and the binding free energy is calculated. The theoretical results lead to a new model for the selective HBV envelopment mechanism.

Chapter 7 presents a statistical-mechanical investigation of hydrophobic effect, which constitutes an essential component of biomolecular interactions in general. Aiming at a more precise description of hydrophobicity from a molecular perspective, this study examines the structure of water molecules near a hydrophobic solute by a combination of thermodynamic analysis for Henry's constants of several nonpolar gases over a broad range of temperatures and molecular dynamic simulations for the hydration energies using two popular molecular models of liquid water.

Finally Chapter 8 summarizes key conclusions from this dissertation and perspectives for future work.

Figure 1. Schematic of a HBV virion with T4 capsid



Chapter 2. Basic Concepts of Classical Density Functional Theory

This chapter outlines the basic formulism of classical density functional theory (DFT). In comparison to conventional methods in molecular simulations, DFT is significantly more effective in quantifying the thermodynamic and microscopic structural properties of complex fluid systems from a molecular perspective. However, it requires various approximations for the free-energy functionals in terms of the one-body density profiles underlying the microscopic structure of molecular systems. In essence, DFT provides quantitative relationships between thermodynamic properties and microscopic structure. While molecular simulation is extremely time consuming for free-energy calculations, numerical implementation of DFT relies on variational computation for the equilibrium density profiles and the corresponding thermodynamic properties, which makes it computationally very efficient. The mathematical procedure is naturally applicable to complex molecular systems with intermolecular interactions spanning multiple time and length scales.

2. 1 Density Profile

Density functional theory expresses the properties of a multi-body system in terms of the density profiles, i.e., the ensemble-averaged spatial-temporal distributions of individual elements (particles). Intuitively, the molecular density profiles represent the average spatial distributions of molecules. To illustrate this mathematical concept, consider a monatomic system with N identical spherical particles. At a given

configuration of the multi-body system, the instantaneous particle density at position $\mathbf{r}$ is a summation of the Dirac delta-functions:

$$\hat{\rho}(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i), \tag{2.1}$$

where $\delta(\mathbf{r})$ stands for the 3-dimensional Dirac delta function

$$\delta(\mathbf{r} - \mathbf{r}') = \begin{cases} 0, & \mathbf{r} \neq \mathbf{r}' \\ \infty, & \mathbf{r} = \mathbf{r}' \end{cases}. \tag{2.2}$$

The Dirac function can be understood as a probability density to find the particle at position $\mathbf{r}$; it satisfies the normalization condition

$$\int d\mathbf{r}\delta(\mathbf{r} - \mathbf{r}') = 1. \tag{2.3}$$

The density profile is defined as an ensemble average of the instantaneous local density [86]

$$\rho(\mathbf{r}) = \hat{\rho}(\mathbf{r}) = \left\langle \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i) \right\rangle. \tag{2.4}$$

The density profile of a multi-body system is intimately related to its thermodynamic properties. For a one-component system with N identical spherical particles considered above, we may write the grand partition function Ξ as

$$\Xi = \sum_{N} \frac{1}{N!\Lambda^{3N}} \int d\mathbf{r}^N \exp\left\{ -\beta \left[\Gamma(\mathbf{r}^N) + N\mu + \sum_{i=1}^{N} \varphi(\mathbf{r}_i) \right] \right\} \tag{2.5}$$

where Λ stands for the thermal wavelength; $\mathbf{r}^N$ is a short notation for $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, \cdots \mathbf{r}_N)$, *i.e.,* the configuration of all particles in the system; $\Gamma(\mathbf{r}^N)$ represents the total interaction potential among these particles; μ is the chemical potential; $\varphi(\mathbf{r})$ is the one-body

external potential, $\beta = 1/(k_B T)$, and k_B is the Boltzmann constant. The one-body profile $\rho(\mathbf{r})$ is given by

$$\rho(\mathbf{r}) = \frac{1}{\Xi}\sum_N \frac{1}{N!\Lambda^{3N}} \int d\mathbf{r}^N \sum_{i=1}^{N} \delta(\mathbf{r}_i - \mathbf{r}) \exp\left\{-\beta\left[\Gamma(\mathbf{r}^N) + N\mu + \sum_{i=1}^{N}\varphi(\mathbf{r}_i)\right]\right\}. \tag{2.6}$$

We can rewrite Eq. (2.6) in terms of a functional derivative of Ξ with respect to the one-body external potential

$$\rho(\mathbf{r}) = -\frac{1}{\beta\Xi}\frac{\delta\Xi}{\delta\varphi(\mathbf{r})} = -\frac{\delta \ln\Xi}{\beta\delta\varphi(\mathbf{r})}. \tag{2.7}$$

Because the grand potential Ω is defined as

$$\beta\Omega = -\ln\Xi. \tag{2.8}$$

Substituting Eq. (2.8) into Eq. (2.7) yields a simple relation between the one-body density profile and the grand potential

$$\rho(\mathbf{r}) = \frac{\delta\Omega}{\delta\varphi(\mathbf{r})}. \tag{2.9}$$

2. 2 Density Functional Theory

The basic ideas of DFT can be traced back to van der Waals' pioneering work published in the late 19th century that concerned with the surface tension of a vapor-liquid interface. Functional minimization of the free energy was utilized as a criterion of equilibrium for inhomogeneous systems[87]. Modern developments in classical DFT is, however, based on the Hohenberg-Kohn (HK) theorem[88], which states that, in an equilibrium system at a given temperature and the chemical potentials of individual species, the one-body external potential for each species can be uniquely determined by

the corresponding one-body density profile. The HK theorem enables the definition of intrinsic Helmholtz free energy, a thermodynamic quantity for inhomogeneous systems that is independent of the one-body external potentials[89]. For a one-component system, we can write the intrinsic Helmholtz energy F as

$$F \equiv A - \int d\mathbf{r}\varphi(\mathbf{r})\rho(\mathbf{r}), \tag{2.10}$$

where A denotes the conventional Helmholtz energy. The conventional and intrinsic Helmholtz energies are related to the grand potential

$$\begin{aligned}\Omega = A - N\mu = F + \int d\mathbf{r}\varphi(\mathbf{r})\rho(\mathbf{r}) - N\mu \\ = F + \int d\mathbf{r}\left[\varphi(\mathbf{r}) - \mu\right]\rho(\mathbf{r})\end{aligned}. \tag{2.11}$$

Because the one-body external potential for each species can be uniquely determined by the corresponding one-body density profile, the intrinsic Helmholtz free energy and subsequently the grand potential are functionals of the one-body density profiles. According to the HK theorem, the equilibrium density profiles minimize the grand potential, i.e., for a one-component system,

$$\frac{\delta\Omega\left[\rho(\mathbf{r})\right]}{\delta\rho(\mathbf{r})} = 0. \tag{2.12}$$

Substituting Eq. (2.11) into Eq. (2.12) yields

$$\frac{\delta F\left[\rho(\mathbf{r})\right]}{\delta\rho(\mathbf{r})} = \mu - \varphi(\mathbf{r}). \tag{2.13}$$

Given an analytical expression for the intrinsic Helmholtz energy, the HK theorem thus provides a mathematical framework for solving the microscopic structure and

corresponding thermodynamic properties, *viz.*, from the functional derivative of the grand potential with respect to the one-body density profiles.

The mathematic procedure can be readily extended to mixtures including polymeric systems of interest in this thesis. In the latter case, we follow a coarse-grained model to describe polymeric species, i.e., polymers are represented as freely-joined chains of spherical particles. Approximately, the number of segments correspond to the degree of polymerization M and segment diameter σ_p is about the same as that of a monomer. In consistent with the polymer model, we assume that monomeric species, such as cations and anions in an aqueous solution, can be represented by spherical particles of diameter σ_a and the solvent is represented by a dielectric continuum. Clearly, the coarse-gained model does not provide atomic details. Nevertheless, we expect that it should be adequate for describing the properties of macromolecular systems significantly larger than atomic length scale.

For most polymeric systems considered in this thesis, we assume further that, for simplicity, the bending potential or the effect of bond angles plays little role in both structure and thermodynamic properties. In that case, the polymer bond potential $V_b(\mathbf{R})$ satisfies

$$\exp\left[-\beta V_b(\mathbf{R})\right] \propto \prod_{i=1}^{M} \delta\left(\left|\mathbf{r}_{i+1}-\mathbf{r}_i\right|-\sigma_p\right), \tag{2.14}$$

where $\mathbf{R} \equiv \left(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_M\right)$ stands for the polymer configuration, i.e., a set of coordinates defining the segmental position $\mathbf{r}_i$. The proportionality constant in Eq. (2.14) is determined from the normalization condition

$$\frac{1}{V}\int \exp\left[-\beta V_b\left(\mathbf{R}\right)\right] d\mathbf{R} = 1, \tag{2.15}$$

to yield,

$$\exp\left[-\beta V_b\left(\mathbf{R}\right)\right] = \prod_{i=1}^{M} \frac{\delta\left(\left|\mathbf{r}_{i+1} - \mathbf{r}_i\right| - \sigma_p\right)}{4\pi\sigma_p^2}. \tag{2.16}$$

For a multi-component system containing polymers and monomers, the grand potential can be expressed in terms of intrinsic Helmholtz free energy functional and one-body external potentials,

$$\begin{aligned}\Omega\left[\rho_M\left(\mathbf{R}\right), \left\{\rho_a\left(\mathbf{r}\right)\right\}\right] = F\left[\rho_M\left(\mathbf{R}\right), \left\{\rho_a\left(\mathbf{r}\right)\right\}\right] + \\ \int\left[\Psi_M\left(\mathbf{R}\right) - \mu_M\right]\rho_M\left(\mathbf{R}\right) d\mathbf{R} + \sum_a \int\left[\Psi_a\left(\mathbf{r}\right) - \mu_a\right]\rho_a\left(\mathbf{r}\right) d\mathbf{r}\end{aligned}, \tag{2.17}$$

where $d\mathbf{R} = d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_M$ is a set of differential volumes; $\rho_M\left(\mathbf{R}\right)$ is a multi-dimensional polymer density profile as a function of the polymer configuration $\mathbf{R}$; $\rho_a\left(\mathbf{r}\right)$ is the segmental distribution of monomers; $\Psi_a\left(\mathbf{r}\right)$ is for the external potential of monomers, and $\Psi_M\left(\mathbf{R}\right)$ is the summation of the external potential of polymer segments, i.e. $\Psi_M\left(\mathbf{R}\right) = \sum_{i=1}^{M} \varphi_i\left(\mathbf{r}_i\right)$. As for a one-component system, the density profiles can be obtained from minimization of the grand potential and based on which thermodynamic properties can be calculated from standard thermodynamic relations.

2. 3 Intrinsic Helmholtz Energy Functional

For model systems considered in this thesis, the intrinsic Helmholtz energy can be decomposed into two parts: the ideal-gas term F^{id} and the excess term F^{ex}

$$F\left[\rho_M(\mathbf{R}),\{\rho_a(\mathbf{r})\}\right]=F^{id}\left[\rho_M(\mathbf{R}),\{\rho_a(\mathbf{r})\}\right]+F^{ex}\left[\rho_M(\mathbf{R}),\{\rho_a(\mathbf{r})\}\right]. \tag{2.18}$$

The ideal-gas term can be obtained exactly; for a system of polymers and monomeric species, it is given by

$$\begin{aligned}\beta F^{id}=&\int\left[\ln\rho_M(\mathbf{R})-1\right]\rho_M(\mathbf{R})d\mathbf{R}+\beta\int V_b(\mathbf{R})\rho_M(\mathbf{R})d\mathbf{R}\\&+\sum_a\int\left[\ln\rho_a(\mathbf{r})-1\right]\rho_a(\mathbf{r})d\mathbf{r}\end{aligned}. \tag{2.19}$$

The excess intrinsic Helmholtz energy accounts for the thermodynamic non-ideality due to inter- and intra-molecular interactions. Unlike its ideal counterpart, the exact representation for the excess free energy is unknown. Indeed, attaining an accurate formulation of the excess intrinsic Helmholtz energy is critical to all DFT calculations.

2. 4 Euler-Lagrange Equation

For a polymeric system at equilibrium, the polymer and monomer density profiles can be determined from minimization of the grand potential (Eq. (2.12))

$$\frac{\delta\Omega}{\delta\rho_M(\mathbf{R})}=0, \tag{2.20}$$

$$\frac{\delta\Omega}{\delta\rho_a(\mathbf{r})}=0. \tag{2.21}$$

The functional derivatives lead to a set of Euler-Lagrange equations:

$$\rho_M(\mathbf{R})=\exp\left[\beta\mu_M-\beta V_b(\mathbf{R})-\beta\Psi_M(\mathbf{R})-\beta\varphi(\mathbf{R})\right], \tag{2.22}$$

$$\rho_a(\mathbf{r}) = \exp\left[\beta\mu_a - \beta\Psi_a(\mathbf{r}) - \frac{\delta\beta F^{ex}}{\delta\rho_a(\mathbf{r})}\right]. \tag{2.23}$$

In the above equations, $\varphi(\mathbf{R})$ represents an effective potential field due to intra- and inter-molecular interactions[90]; $\beta\varphi(\mathbf{R}) = \partial\beta F^{ex} / \partial\rho_M(\mathbf{R})$. The functional derivative of the excess Helmholtz energy with respect to the density profile of a monomeric species gives the local excess chemical potential:

$$\mu_a^{ex}(\mathbf{r}) = \frac{\delta\beta F^{ex}}{\delta\rho_a(\mathbf{r})}. \tag{2.24}$$

For a homopolymer consisting of identical segments, the number density $\rho_p(\mathbf{r})$ is defined as

$$\rho_p(\mathbf{r}) = \sum_{i=1}^{M}\rho_{si}(\mathbf{r}_i) = \sum_{i=1}^{M}\int d\mathbf{R}\delta(\mathbf{r}-\mathbf{r}_i)\rho_M(\mathbf{R}), \tag{2.25}$$

where $\rho_{si}(\mathbf{r})$ is the local density of *i*-th segment. From Eq. (2.25), $\varphi(\mathbf{R})$ is simplified to:

$$\varphi(\mathbf{R}) = \frac{\delta F^{ex}}{\delta\rho_M(\mathbf{R})} = \sum_{i=1}^{M}\frac{\delta F^{ex}}{\delta\rho_p(\mathbf{r}_i)}. \tag{2.26}$$

Substitution of Eq. (2.26) in Eq. (2.22) yields

$$\rho_M(\mathbf{R}) = \exp\left[\beta\mu_M - \beta V_b(\mathbf{R}) - \beta\sum_{i=1}^{M}\lambda_i(\mathbf{r}_i)\right] \tag{2.27}$$

where $\lambda_i(\mathbf{r}_i)$ represents an effective one-body potential for *i*-th segment

$$\lambda_i(\mathbf{r}_i) = \frac{\delta F^{ex}}{\delta\rho_p(\mathbf{r}_i)} + \Psi_i(\mathbf{r}_i), \tag{2.28}$$

with $\Psi_i(\mathbf{r}_i)$ represents the contribution from the external potential. The segment density $\rho_{si}(\mathbf{r})$ is obtained from

$$\rho_{si}(\mathbf{r}) = \int d\mathbf{R}\delta(\mathbf{r}-\mathbf{r}_i)\exp\left[\beta\mu_M - \beta V_b(\mathbf{R}) - \beta\sum_{i=1}^{M}\lambda_i(\mathbf{r}_i)\right]. \tag{2.29}$$

Summation over all segments in the polymer chain gives

$$\rho_p(\mathbf{r}) = \exp(\beta\mu_M)\int d\mathbf{R}\sum_{i=1}^{M}\delta(\mathbf{r}-\mathbf{r}_i)\exp\left[-\beta V_b(\mathbf{R}) - \beta\sum_{j=1}^{M}\lambda_j(\mathbf{r}_j)\right]. \tag{2.30}$$

2. 5 Excess Helmholtz Free Energy

Now we move on to define the one-body effective potential $\lambda_i(\mathbf{r}_i)$ or, equivalently, excess Helmholtz free energy F^{ex}. Formulation of these quantities are key in DFT calculations.

For polymeric systems considered in this dissertation, the excess Helmholtz free energy includes several contributions arising from different molecular interactions and correlation effects. According to the coarse-grained model, the excess Helmholtz free energy can be decomposed into four inter-related contributions:

$$F^{ex} = F_{hs}^{ex} + F_{ch}^{ex} + F_{C}^{ex} + F_{el}^{ex}. \tag{2.31}$$

In Eq.(2.31), the subscripts indicate contributions to the excess Helmholtz energy due to the hard-sphere repulsion (F_{hs}^{ex}), the chain connectivity (F_{ch}^{ex}), the direct Coulomb energy (F_{C}^{ex}), and the electrostatic correlations (F_{el}^{ex}), respectively. Each contribution to the thermodynamic potential can be formulated based on a combination of physical insights and mathematic analysis. In the following, we give the final expressions for the excess

Helmholtz free energy functional. The mathematical details for derivations of these equations can be found from the literature.

The excess Helmholtz energy due to the hard-sphere repulsion F_{hs}^{ex} accounts for molecular excluded-volume effect. It is determined from the modified fundamental measure theory (MFMT)[91, 92]:

$$\beta F_{hs}^{ex} = \int d\mathbf{r}\{-n_0 \ln(1-n_3) + \frac{n_1 n_2 - \mathbf{n}_{V1}\mathbf{n}_{V2}}{1-n_3} + \frac{1}{36\pi}[n_3 \ln(1-n_3) + \frac{n_3^2}{(1-n_3)^2}]\frac{(n_2^3 - 3n_2\mathbf{n}_{V2}\mathbf{n}_{V2})}{n_3^3}\}, \quad (2.32)$$

where $\{n_\alpha, \alpha = 1,2,3,V1,V2\}$ stand scalar and vector weighted densities as defined in the original version of the FMT[93]:

$$n_\alpha(\mathbf{r}) = \sum_i n_{\alpha i}(\mathbf{r}) = \sum_i \int \rho_i(\mathbf{r}')\omega_i^{(\alpha)}(|\mathbf{r}-\mathbf{r}'|)d\mathbf{r}'. \quad (2.33)$$

In comparison to the original FMT, MFMT utilizes the Boublik-Mansoori-Carnahan-Starling-Leland (BMCSL) equation of state for hard-sphere mixtures,[94] which is more accurate than the original counterpart as described by the Percus-Yevck (PY) equation for bulk systems. As a result, MFMT provides a more accurate description of the density profiles of inhomogeneous hard-sphere fluids and the pair distribution functions of bulk systems. In general, MFMT is applicable to inhomogeneous hard-sphere fluids at all densities.

The excess Helmholtz energy due to chain connectivity describes thermodynamic properties related with correlation effects arising from the polymer bond potential. This term is represented by the generalized thermodynamic perturbation theory (TPT)

proposed by Yu and Wu in 2002. As in original TPT for bulk systems, we use the corresponding monomeric system as a reference. For uniform systems, TPT was first proposed by Wertheim[95] to investigate association in a monomeric hard-sphere fluid. The first-order thermodynamic perturbation theory (TPT1) can be empirically generalized for inhomogeneous systems[96]:

$$\beta F_{ch}^{ex} = \frac{1-M}{M}\int n_{0p}\zeta_p \ln y(\sigma_p, n_\omega)\, d\mathbf{r}, \tag{2.34}$$

where $\zeta_p = 1 - \mathbf{n_{V2p}}\mathbf{n_{V2p}}/n_{2p}^2$ describes the effect of local density variation, and $y(\sigma_p, n_\omega)$ presents the contact value of the cavity correlation function (CCF) of segments.

For an inhomogeneous system of monomeric charged particles, we use the mean-spherical approximation (MSA) [97] for $y(\sigma_p, n_\omega)$:

$$y(\sigma_p, n_\omega) = \left[\frac{1}{1-n_3} + \frac{n_2\sigma_p(1-\mathbf{n_{V2}}\mathbf{n_{V2}}/n_2^2)}{4(1-n_3)^2}\right]\exp\left(-\frac{\Gamma^2 a_p^2}{4\pi^2 l_B \sigma_p}\right)\exp\left(\frac{l_B Z_p^2}{\sigma_p}\right), \tag{2.35}$$

where Γ and a_p are described as

$$\Gamma = \sqrt{\pi l_B \sum_{k=p,+,-} n_{0k}\left(\frac{1}{1+\Gamma\sigma_k}\right)^2\left(Z_k - \frac{\pi P_n \sigma_k^2}{2(1-n_3)}\right)^2}, \tag{2.36}$$

$$a_p = \frac{2\pi l_B\left(Z_p - \frac{\pi P_n \sigma_p^2}{2(1-n_3)}\right)}{\Gamma(1+\Gamma\sigma_p)} \tag{2.37}$$

with

$$P_n = \sum_{k=p,+,-} \frac{2n_{1k}Z_k}{1+\Gamma\sigma_k} \Bigg/ \left(1+\frac{3}{(1-n_3)}\sum_{k=p,+,-}\frac{n_{3k}}{1+\Gamma\sigma_k}\right). \tag{2.38}$$

In above equations, we adopted the weighted densities from FMT. The DFT equations can be applied to systems containing both homogeneous and heterogeneous polymers (composed of segments with different charges and sizes).

There are two contributions to the non-ideality due to electrostatic interactions, one from the direct Coulomb interaction and the other from charge correlations. The contribution to excess Helmholtz energy from the direct Coulomb interaction is the same as that from the conventional Poisson-Boltzmann (PB) equation[98]:

$$\beta F_C^{ex} = \frac{l_B}{2}\sum_{i,j=p,+,-}\int\int d\mathbf{r}d\mathbf{r}'\frac{Z_iZ_j\rho_i(\mathbf{r})\rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{2.39}$$

Because the integral on the right side of Eq. (2.40) diverges for any pair potential owing to the long-range nature of the Coulomb potential, the electrostatic energy is usually calculated by defining the local mean electrostatic potential $\psi(\mathbf{r})$:

$$\psi(\mathbf{r}) = \int d\mathbf{r}' \sum_{j=p,+,-}\frac{Z_j e\rho_j(\mathbf{r}')}{4\pi\varepsilon|\mathbf{r}-\mathbf{r}'|}. \tag{2.41}$$

This local mean electrostatic potential satisfies the Poisson equation

$$\nabla^2\psi(\mathbf{r}) = -\frac{4\pi e}{\varepsilon}\rho_c(\mathbf{r}), \tag{2.42}$$

where $\rho_c(\mathbf{r}) = \sum_{i=p,+,-} Z_i\rho_i(\mathbf{r})$.

The Poisson equation can be numerically solved with appropriate boundary conditions. As an example, we consider the ionic density profiles inside a spherical

cavity, Because of the spherical symmetry, the ionic density distributions depend only on the radial r direction. In that case, the Poisson equation is reduced to:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi(r)}{dr}\right)=-\frac{4\pi e}{\varepsilon}\rho_c(r). \tag{2.43}$$

With boundary conditions

$$\left.\frac{d\psi(r)}{dr}\right|_{r=r_c}=0, \tag{2.44}$$

$$\psi(r)\big|_{r=r_c}=0, \tag{2.45}$$

where r_c is the calculation range at which the electrolyte solution reaches to the bulk properties. An integration of Eq. (2.46) with the boundary conditions gives

$$\begin{aligned}\psi(r)&=-\frac{4\pi e}{\varepsilon}\int_r^{r_c}\frac{1}{r_2^2}dr_2\int_{r_2}^{r_c}r_1^2\rho_c(r_1)dr_1\\&=\frac{4\pi e}{\varepsilon}\int_r^{r_c}\left(r_1-\frac{r_1^2}{r}\right)\rho_c(r_1)dr_1\end{aligned}. \tag{2.47}$$

We calculate the excess Helmholtz free energy due to electrostatic correlations from the hypernetted chain approximation (HNC). It has been shown that HNC provides a faithful description of the long-ranged electrostatic correlations[99, 100]. According to HNC, the excess Helmholtz energy is represented by a quadratic functional expansion with respect to the property for a bulk fluid with uniform densities $\{\rho_i^b\}$:

$$\begin{aligned}\beta F_{el}^{ex}\left[\{\rho_i(\mathbf{r})\}\right]=\beta F_{el}^{ex}\left[\{\rho_i^b\}\right]&+\sum_{i=p,+,-}\beta\mu_i^{el}\int d\mathbf{r}\Delta\rho_i(\mathbf{r})\\&-\frac{1}{2}\sum_{i,j=p,+,-}\iint d\mathbf{r}_1d\mathbf{r}_2\Delta\rho_i(\mathbf{r}_1)\Delta\rho_j(\mathbf{r}_2)c_{ij}^{el}(|\mathbf{r}_1-\mathbf{r}_2|)\end{aligned}, \tag{2.48}$$

where μ_i^{el} denotes the excess chemical potential due to electrostatic correlations for specie i in the bulk; c_{ij}^{el} represents the corresponding second-order direct correlation functions (DCF). We assume that the chain connectivity effect on that electrostatic part of the DCF can be neglected[101]. In that case, the direct correlation function can be obtained from MSA [102]:

$$c_{ij}^{el}(r) = c_{ij}(r) - c_{ij}^{C}(r) - c_{ij}^{hs}(r). \tag{2.49}$$

The DCF due to direct Coulomb interaction is $c_{ij}^{C}(r) = -l_B Z_i Z_j / r$; that due to the hard-sphere repulsion is calculated from the PY equation, $c_{ij}^{hs}(r)$[103].

It has been demonstrated in our previous work that the DFT provides an accurate description of the thermodynamic properties of free and tethered polyelectrolytes in comparison with experiment and simulation[104-109]. In addition, the DFT has been used to describe the equilibrium properties of HBV nucleocapsids and provides proper accounts of all important factors in genome packaging[110, 111] as well as the structural change during HBV replication[112].

2. 6 Numerical method

Here we review briefly the numerical procedure for application of the DFT to tethered polyelectrolytes. For a one-dimensional system with spherical symmetry, the density distribution varies only in r direction. As a result, Eqs. (2.23) and (2.30) can be simplified to

$$\rho_a(r) = \exp\left[\beta\mu_a - \beta\Psi_a(r) - \frac{\partial\beta F^{ex}}{\partial\rho_a(r)}\right], \tag{2.50}$$

$$\rho_p(r) = \exp(\beta\mu_M) \int d\mathbf{R} \sum_{i=1}^{M} \delta(r - r_i) \exp\left[-\beta V_b(\mathbf{R}) - \beta \sum_{j=1}^{M} \lambda_j(\mathbf{r}_j) \right]. \quad (2.51)$$

Substituting Eq. (2.16) into Eq. (2.51), we have

$$\rho_p(r) = \exp(\beta\mu_M) \sum_{j=1}^{M} \exp\left[-\beta\lambda_j(r_j) \right] G^i(r) G^{M+1-i}(r), \quad (2.52)$$

where $G^i(r)$ is a chain propagator function (i.e., Green function), which is described by the recursive relation:

$$G^i(r) = \frac{1}{2\sigma_p r} \int_{|\sigma_p - r|}^{r+\sigma_p} dr' \exp[-\beta\lambda_p(r')] G^{i-1}(r') r' \quad (2.53)$$

for $i = 2, \cdots, M$, while $G^1(r) = 1$ with σ_p denoting the segment diameter.

The tethered polyelectrolytes possess unique properties in comparison to free polyelectrolytes because of the tethering effect and end segments distribution. Such properties are reflected in the oscillation of the polyion density near the tethering point and in the distinguished behaviors of the end segments distribution. If the chain length approaches infinity, however, the effect of end segments becomes less significant and all segments in polymer become indistinguishable. In that case, the Green function approaches to a limiting function $G(r)$, which can be solved self-consistently from the relation [113],

$$G(r) = \int dr' \exp\left[-\beta\lambda(r') \right] \frac{\theta\left(\sigma_p - |r - r'|\right)}{2\sigma_p} G(r'). \quad (2.54)$$

Subsequently, Eq. (2.52) is simplified to

$$\rho_p(r) = M \exp\left(\beta\mu_M - \beta\lambda(r)\right) \left[G(r) \right]^2. \quad (2.55)$$

In this work, the density distributions of polyelectrolyte components are predicted through above equations for tethered and free polyelectrolytes, respectively, using the Picard iteration method. From an initial input for the density profiles, we obtain relevant quantities such as the reduced electrostatic potential ($\psi^*(r)$), the effective one-body potential for particles ($\lambda_k(r)$), and the Green function ($G^i(r)$). The new density profiles are then calculated and selected as an updated input for the density profiles. If the change between input and output profiles is beyond a certain threshold, the iteration cycle is continued. For DFT calculations in this work, the iteration is considered converged when the variance on the density distributions of monomeric ions and polyions at all positions is smaller than 0.01%.

Chapter 3. A Molecular Thermodynamic Model for the Stability of Hepatitis B Capsids

Abstract

Self-assembly of capsid proteins and genome encapsidation are two critical steps in the life cycle of most plant and animal viruses. A theoretical description of such processes from a physiochemical perspective may help better understand viral replication and morphogenesis thus provide fresh insights into the experimental studies of antiviral strategies. In this work, we propose a molecular thermodynamic model for predicting the stability of Hepatitis B virus (HBV) capsids either with or without loading nucleic materials. With the key components represented by coarse-grained thermodynamic models, the theoretical predictions are in excellent agreement with experimental data for the formation free energies of empty T4 capsids over a broad range of temperature and ion concentrations. The theoretical model predicts T3/T4 dimorphism in good agreement with the capsid formation at *in vivo* and *in vitro* conditions. We have also studied the stability of the viral particles in response to physiological cellular conditions with the explicit consideration of the hydrophobic association of capsid subunits, electrostatic interactions, molecular excluded volume effects, entropy of mixing, and conformational changes of the biomolecular species. The course-grained model captures the essential features of the HBV nucleocapsid stability revealed by recent experiments.

I. Introduction

Understanding the stability of viral capsids under various intracellular environments is vitally important for studying the molecular basis of viral replication cycles and for developing efficient therapeutic strategies against viral infection[114, 115]. Controlling capsid stability is also imperative for the development of robust viral vectors for gene delivery. Whereas the thermodynamic basis of capsid formation and stability is not much different from that for a wide variety of cooperative processes in nature including micellization and polymerization, viral assembly is unique for its precise coordination of a large number of capsid proteins and the generic materials in exceedingly complicated intracellular surroundings of the host cell. The environmental complexity and atomic-level precision of viral replication make a quantitative description of the self-assembly processes extremely difficult[116-119]. Fundamental studies have greatly advanced our knowledge on the generic mechanisms of viral assembly, stability and dynamics, as shown in recent theoretical investigations[120-122]. However, relatively few investigations have been dedicated to their applications to specific viruses of practical concern. The purpose of this work is to present a theoretical framework for predicting the stability of hepatitis B virus (HBV) capsids. By using coarse-grained models for the key components, *i.e.*, capsid subunits, nucleic acids, and the protamine tails of the capsid proteins, we demonstrate that the molecular thermodynamic model is able to predict the free energies of capsid formation in good agreement with experimental data for both empty capsids and nucleocapsids over a broad range of solution conditions.

HBV is a para-retrovirus that replicates through an RNA intermediate[123]. Although the viral infection is now preventable through vaccination, HBV affects more than 2 billion people alive today and is responsible for over half a million deaths every year. Despite remarkable progress in understanding the natural history and pathogenesis of HBV, effectual treatment and eradication of chronic infection continue to be a clinic challenge. The HBV virion consists of a double-shelled structure with an icosahedral nucleocapsids (NC) and a spherical shell of lipoproteins known as HBV surface antigen or HBsAg[2, 27, 124]. The wild-type virus contains a relaxed circular, partially double-stranded (ds) DNA of about 3200 base pairs (bp) in length. The native NCs exist in two different forms, with the majority (>90%) having T = 4 icosahedral symmetry and a small fraction with T = 3 symmetry[125]. The dimorphism composition of HBV capsids produced *in vitro* is quite different from that of native NCs, sensitive to the capsid charge and the genome content[5, 126]. The wild-type capsid protein, referred to as HBV core antigen or HBcAg, consists of 183 amino acid (aa) residues, which can be divided into three domains: the N-terminal assembly domain (1-140), the nonapeptide linker (141-149), and the highly basic, protamine-like C-terminal domain (CTD) (150-183). A T4 capsid is composed of 120 dimers of the capsid protein and, for a T3 capsid, 90 dimers. As shown schematically in Figure 3-1, the capsid dimers are arranged at the triangulation lattices of the icosahedral surfaces. The T4 capsid is fenestrated with hydrophilic pores 1.4 nm in diameter around the 3-fold axes, hydrophobic oval pores about 2.4 nm long and 1.8 nm wide at the 2-fold axes, and smaller pores 0.3 nm in diameter at the positions of the 5-fold icosahedral axes[6]. The T3 capsids have similar pores but with slightly different

shape[5]. These pores make nucleotides and small chemical species directly accessible to the capsid lumen[127, 128].

HBV replication *in vivo* undergoes sequential events that require spatially and temporally coordinated interactions of the key components of the infectious virion[129]. The replication cycle begins with the binding of surface proteins to hepatocytes and subsequent internalization of the entire virion by membrane fusion[124]. After uncoating the surface proteins by lipid molecules, the viral capsid migrates into the cytosol of the infected cell and is destabilized during its transport to the cellular nucleus due to the change in the local environment. Within the nucleoplasm, the partially dsDNA is first repaired to complete the double strand and then converted into a covalently closed circular (ccc) DNA, which serves as the transcriptional template to generate multiple copies of pregenomic (pg) and subgenomic (sg) RNAs[22]. The self-assembly of progeny NCs starts with the specific binding of pgRNA with the reverse transcriptase and capsid proteins. Within the newly formed NC (immature), pgRNA is reversely transcribed into the partially completed dsDNA genome[130, 131]. A remarkable feature of HBV replication is that only the dsDNA-containing NCs (mature) can be transported back to the nucleoplasm for intracellular amplification or translocated to the endoplasmic reticulum (ER) of the host cell for envelopment and egress[27]. It has been shown that NCs containing single stranded (ss) DNA or RNA are blocked from importing to the nucleoplasm or binding with the surface proteins despite their striking external similarity[132]. More interestingly, recent *in vivo* experiments indicate that HBcAg form empty capsids that exhibit envelopment and egression properties similar to the matured capsids[132]. It has

been speculated that the dynamic structure of CTD (C-terminal domain with 34 residues) plays an essential role in HBV replication[133-135].

In vitro assembly of HBV capsids devoid of CTD and nucleic acids has been analyzed in detail[5, 126]. By describing the process in terms of a cascade of low-order association reactions, Zlotnick and coworkers found that the intermediate concentrations were extremely low and thus the kinetic of self-assembly followed essentially a one-step process[81]. The single-step model allows for the quantification of the thermodynamic stability of empty HBV capsids on the basis of the concentrations of the capsomers and the fully-assembled capsids at various solution conditions[136]. It has been shown that the thermodynamic stability increases at higher temperature or salt concentration and such effects can be rationalized with hydrophobic and electrostatic interactions between capsid proteins[118, 136, 137]. Numerical analysis confirmed the dependence of protein-protein interactions on temperature and ionic strength[119, 121, 138]. It is worthwhile noting that other effects may also influence the capsid satiability. For example, it has been proposed that the increase in ionic strength might lead to the conformation change of the capsid protein thereby enhancing the capsid stability[136]. It is also known that chemicals such as zinc ions or heteroaryldihydropyrimidines alter the conformation of the capsid protein preventing their proper assembly[83-85, 139]. Besides, recent experiments indicate that mutation on certain residues of the capsid protein modulates the tendency of capsid assembly[140-143].

Whereas existing experimental analyses of capsid stability were mostly performed on empty capsids devoid any nucleic material or CTD tails, there have been a number of investigations on the role of CTD in HBV replication, such as its relations with the

genome size[134] as well as the unknown functions associated with phosphorylation[135]. Besides, several experimental studies have also been reported on the *in vitro* assembly of empty or RNA-containing HBV capsids with the full-length CTD[18, 144, 145]. In comparing to that for empty capsids without CTD, the assembly of the full-length capsids occurs only at high ionic strength (≥0.25 M) because of the strong electrostatic repulsion among the basic tails. Capsid destabilization due to the increased charge repulsion is evident by digesting the encapsidated RNA in nucleocapsids (NC)[144]. On the other hand, favorable electrostatic attractions between CTD and RNA chains are expected to make nucleocapsids (NC) more stable than corresponding empty capsids[18, 126, 144, 145]. It has been suggested that NC particles undergo significant stability changes during the replication cycle, and NCs with dsDNA are relatively unstable in comparison to those containing ssRNA or ssDNA[146, 147]. Further evaluation of the NC stability is much desirable in particular with respect to different genome contents and characteristics.

II. Theory and Methods

A. Thermodynamics of capsid formation

The stability of HBV capsids has been analyzed by *in vitro* experiments either with or without the genomic materials. In both cases, the intermediate structures are insignificant such that the thermodynamics of capsid formation can be approximately described in terms of a one-step process similar to that commonly used for micellization. Unlike micelles that show broad size and shape polydispersities, however, viral capsids typically adopt a uniform size with the capsid proteins arranged in a lattice-like structure. In particular, many viral capsids take an icosahedral structure with the capsid proteins

forming a near spherical shell. The intrinsic symmetry of 20 equilateral triangular facets requires that an icosahedral capsid must consist of multiples of 60 identical proteins. The integer multiplication is known as triangulation number or T number. Natural HBV capsids may exist in either T4 or T3 icosahedral structures, and the dimorphism persists *in vitro* synthesis but with significantly different composition.

a) Stability of CTD-free empty capsids

In vitro formation of empty HBV capsids free of CTD can be described by a quasi-chemical reaction for the core protein dimers (*viz.* capsomers) [148]

$$\mathrm{N} \times \mathrm{Dimer} \leftrightarrow \mathrm{Capsid}, \tag{3.1}$$

where $N = 90$ for a T3 capsid and 120 for a T4 capsid [149]. In a dilute solution, the equilibrium dimer concentration (D_0) and that of the fully assembled capsids (C) satisfy the mass action law

$$K_C = [\mathrm{C}]/[\mathrm{D}_0]^N, \tag{3.2}$$

where K_C denotes the apparent equilibrium constant. In terms of the mole fractions of dimers (X_{D_0}) and capsids (X_C), the equilibrium constant is dimensionless

$$K_0 = \frac{\mathrm{X}_C}{X_{D_0}^N}, \tag{3.3}$$

One can easily find the connection between the two equilibrium constants by converting the molar concentrations into the corresponding mole fractions:

$$\begin{aligned} [\mathrm{D}_0] &= [\mathrm{W}] \cdot \mathrm{X}_{\mathrm{D0}} \\ [\mathrm{C}] &= [\mathrm{W}] \cdot \mathrm{X}_{\mathrm{C}} \end{aligned}, \tag{3.4}$$

where [W] is the molarity of pure water. Substituting Eq. (3.4) into Eq. (3.3) gives

$$\ln K_0 = \ln K_c + (N-1)\cdot \ln[\mathrm{W}], \tag{3.5}$$

The free energy of capsid formation is related to the dimensionless equilibrium constant

$$\Delta F_0 = -k_B T \ln K_0, \tag{3.6}$$

where k_B is the Boltzmann constant and T is the absolute temperature. According to Eq. (3.6), the free energy of capsid formation is defined as the difference between the chemical potential of the empty capsid and that of the CTD-truncated protein dimers at the standard conditions, *i.e.*, in their corresponding hypothetical pure states[118]. With the concentrations of the core-protein dimers and the fully assembled capsids measured at various equilibrium conditions, one can quantify the equilibrium constant and subsequently the free energy of capsid formation[115].

It has been shown that the self-assembly of empty HBV capsids from CTD-truncated core proteins is entropy driven, primarily due to inter-dimer hydrophobic interactions[136]. If one divides the free energy of capsid formation into the contact values between nearest neighboring dimers, the equilibrium constant provides an indirect measure of the pair association constant[115]

$$K_{\text{contact}} = (NK_0 / j^{N-1})^{2/cN}, \tag{3.7}$$

where j represents the degeneracy of the capsid with N dimers, and c stands for the number of contacts per dimer. The association free energy per contact (for each dime pair) is given by

$$\Delta F_{\text{contact}} = -k_B T \ln K_{\text{contact}}. \tag{3.8}$$

Experiments for *in vitro* assembly of empty HBV capsids indicate that the average value for the contact free energy is between -3 and -4 kcal/mol, which is "*surprisingly*" small in comparison to alternative estimations based on the dimer structure (~ -10 kcal/mol)[136]. The discrepancy had been attributed to multiple causes including possible changes in the protein structure due to the self-assembly, the pairwise additivity assumption for the binding free energy, or the imprecise calculation of the hydrophobic interactions.

b) Stability of nucleocapsids

The above thermodynamic analysis can be similarly applied to the assembly of HBV nucleocapsids with native core protein dimers in the presence of either RNA or DNA. In this case, the quasi-chemical reaction becomes

$$\mathrm{N} \times \text{Dimer}\,(\mathrm{D}) + \text{Nucleic Acids}\,(\mathrm{NA}) \leftrightarrow \text{Nucleocapsid}\,(\mathrm{NC}), \tag{3.9}$$

and the corresponding equation for the mass action law is

$$K_{NC} = \frac{[N\mathrm{C}]}{[\mathrm{D}]^{N}[NA]}, \tag{3.10}$$

In a dilute solution, the apparent equilibrium constant is related to the free energy of capsid formation

$$\beta \Delta F = -\ln K_{NC} - N \cdot \ln[\mathrm{W}], \tag{3.11}$$

where $\beta = 1/(k_B T)$.

Assuming that NC capsids maintain the structure integrity for both the core-protein assembly domain as well as the capsid shell, we may divide the free energy of nucleocapsid formation in terms of that corresponding to the formation of a CTD-free

empty capsid, designated as ΔF_0, and a contribution due to CTD-RNA/DNA interactions and biomacromolecules encapsidations:

$$\Delta F = \Delta F_0 + F_{\text{plex}}, \tag{3.12}$$

where F_{plex} denotes the free energy of polyelectrolyte complexation between CTD and nucleic acids. Interestingly, complex formation between oppositely charged polyelectrolytes is also entropy driven, primarily due to the release of a large number of counterions associated with the electrostatic binding, here between the positively charged CTD chains and negatively charged nucleic acids[150]. Whereas *in vitro* assemblies of both empty and RNA-containing nucleocapsids with intact CTD into have been recently demonstrated[132, 144, 145, 151], we are unaware of explicit quantification of the formation free energy for HBV nucleocapsid at different solution conditions. Because the native HBV core proteins are able to assemble into empty capsids both *in vitro* and *in vivo*, the competition between formations of empty and nucleocapsids makes the experimental procedure for studying nucleocapsid formation conceivably much more difficult than that for empty capsids.

B. Energetics of viral assembly

The experimental investigations on the formation of viral capsids and polyelectrolyte complexes render an excellent starting point for interrogating the molecular driving forces of the capsid assembly processes. As mentioned above, capsid formation is entropy driven associated with the reorganization of water molecules for hydrophobic interactions and the release of massive counterions due to the binding of the nucleic acids with oppositely charged polypeptides. Whereas a first-principles approach

for quantifying either cause has been actively pursued for decades and remains of interest in physical chemistry, a number of semi-empirical models are readily available to capture the entropic forces leading to virus assembly. In this work, we use such models to describe hydrophobic and polyelectrolyte interactions underlying the formation of empty and RNA-loaded capsids. While the discussion here is mainly focused on HBV, we expect similar procedures applicable to capsid formation for other viruses in general.

a) Rigid capsids

It was proposed first by Crick and Watson that the capsids of most small viruses exhibit an icosahedral structure with near perfect arrangement of the protein subunits[152]. The lattice structure ensures regular packing of the repeating subunits in identical environment and allows a good estimation of the formation free energy simply based on the contact value of the protein-protein interactions[136]. In this work, we propose an icosahedral lattice model to estimate the association free energy of empty capsids consisting of the assembly domain of the HBV core proteins (*viz*., without CTD). While a similar thermodynamic model has been proposed before for empty HBV capsids[133], our estimation of the binding free energy is based on explicit coarse-grained representations of the core-protein subunits and free of any adjustable parameters.

In our coarse-grained model of HBV capsids, we assume that each capsomer has a rigid structure that can be represented by three tangentially connected spheres mimicking the topology of HBcAg determined from experiments. Figure 3-1 shows schematically the coarse-grained model for capsomers and the arrangements of the protein subunits at the triangular facets of T3 and T4 capsids. The diameters of these coarse-grained

segments are selected such that they approximately reproduce the protein volume as determined from the crystal structure[153]. The large sphere of the capsomer, 3.2 nm in diameter, corresponds to the 4-helix bundle of each subunit; it is flanged between two small spheres, each 1.6 nm in diameter, representing the base part of the HBcAg assembly domain. According to the crystalline structure of the HBV capsids and the icosahedral symmetry[149], each subunit has 4 nearest neighbors in both T3 and T4 capsids, joined together through hydrophobic contacts. Whereas the coarse-grained model ignore atomistic details, it allows us to quantify the hydrophobic association leading to the self-assembly of capsomers to both T3 and T4 capsids. The simplified representation of hydrophobic interactions is useful to interpret the T3/T4 isomorphism and the mutation effects of HBV capsid formation.

The free energy of capsid formation corresponds to the reversible work to bring the subunits from infinitely apart to the T3 or T4 icosahedral lattice. Approximately, it can be decomposed into contributions due to hydrophobic (F_{hyd}), van der Waals (F_{vdw}), and electrostatic (F_{el}) interactions:

$$\Delta F_0 = F_{hyd} + F_{vdw} + F_{ele}. \tag{3.13}$$

As well documented[154], the hydrophobic interaction is short-ranged and can be related to the solvent accessible surface area of the protein subunits. While both van der Waals and electrostatic interactions are longer-ranged, depending on the protein-protein separations and the salt concentration. The long-range potentials are relatively insignificant for the formation of empty HBV capsids because of the low charge density of the assembly

domain of the core proteins and the small Hamaker constant for proteins in an aqueous solution[115].

A conventional way to estimate the hydrophobic free energy is based on the reduction of the surface area of the hydrophobic domain of the subunits due to capsid formation[118]. For a HBV capsid with the T4 structure, bringing together each pair of the capsid subunits reduces the hydrophobic area on the order of $1100 \sim 1520$ Å^2. Applying this model to fit the experimental values for the free energy of capsid formation would lead to an interfacial tension of $\gamma \sim 5.5\, mN/m$[136], which is much too small in comparison with the common value of the interfacial tension for a hydrophobic surface ($\gamma \sim 50\, mN/m$). The coarse-grained model for the subunits shown in Figure 3-1 allows us to estimate the hydrophobic free energy without explicitly evoking the protein surface area. Because the subunits are in contact only through the small spheres, the total hydrophobic free energy of association corresponds to a summation of their contact hydrophobic potentials. According to a recent work on hydrophobic interactions[155], the potential of mean force between two hydrophobic spheres of diameter $\sigma = 1.6$ nm at contact can be estimated from the overlapping volume of the hydration layers:

$$\beta w_{hyd} = -A\sigma_W^2 (2\sigma_W + 3\sigma), \tag{3.14}$$

where $\sigma_W = 2.8$ Å denotes the diameter of a water molecule, and A is a constant reflecting the difference in the entropy of water molecules within the solvation shell and that in the bulk[155]. Based on the solubility of small hydrocarbons in water over the temperature range of interest in this work, we find that parameter A (in units of nm^{-3}) can be linearly correlated with the absolute temperature

$$A = 0.067T + 4.481, \tag{3.15}$$

Substituting the expression for A into Eq.(3.14) gives the hydrophobic free energy for each inter-dimer contact

$$\beta w_{hyd} = -0.028T - 1.883. \tag{3.16}$$

According to Eq.(3.19), the hydrophobic free energy between the neighboring dimers is around -6 kcal/mol at room temperature (T=298 K). The value is close to that estimated by Ceres and Zlotnick[136] and comparable to the binding free energy used in molecular simulation of capsid assembly[156]. In a T4 HBV capsid, our coarse-grained model predicts 120×4/2=240 direct contacts between small hydrophobic spheres, which yields a hydrophobic free energy of $F_{hyd} = 240 w_{hyd}$. For a T3 capsid, the hydrophobic free energy is $F_{hyd} = 180 w_{hyd}$.

We can estimate the van der Waals interaction between the subunits based on the standard equations from the colloids literature[157]. Because the solvent-mediated van der Waals force between small spheres is insignificant in comparison with the hydrophobic potential, we consider the van der Waals interactions only between the large spheres. For spherical particles much larger than water molecules, the van der Waals free energy is given by

$$w_{vdw} = -\frac{H\sigma_L}{24d}. \tag{3.17}$$

In Eq.(3.17), $\sigma_L = 3.2$ nm is the diameter of the large sphere in our coarse-grained model of the capsomers, H denotes the Hamaker constant, and $d = 0.8$ nm is the surface-to-surface distance between the large spheres. For proteins in water, the Hamaker constant is

not subject to large variations[158]. In this work, we assume $H \simeq 4. \times 10^{-21} J$, independent of temperature. Using this Hamaker constant, we can calculate the van der Waals attraction between two neighboring dimers, which is about -0.1 kcal/mol. As expected, the contribution of van der Waals energy to the free energy of capsid formation is much smaller than that of the hydrophobic interaction. Nevertheless, each big sphere interacts with four nearest neighbors and, in a T4 HBV capsid, the total van der Waals free energy is $F_{vdw} \approx -40H$. For a T3 capsid, the total van der Waals free energy is $F_{vdw} \approx -30H$.

We follow Kegel and van der Schoot[118] to estimate the electrostatic repulsion among the subunits. In this model, all electrostatic charges are assumed to distribute uniformly at the surface of a spherical capsid shell and the free energy of surface charging is given by

$$F_{ele} = A_c \cdot \sigma^2 \cdot l_B \cdot \kappa^{-1} \cdot k_B T , \tag{3.18}$$

where A_c represents the total charged surface area, σ is the charge density, l_B stands for the Bjerrum length, $\kappa^{-1} = 1/\sqrt{8\pi l_B \rho_S}$ is the Debye screening length, and ρ_s is the overall number density of ions in the electrolyte solution. The Bjerrum length depends only on temperature and the solvent dielectric constant:

$$l_B = e^2 / (4\pi\varepsilon_0 \varepsilon_r k_B T) , \tag{3.19}$$

where e is the unit charge, ε_0 is the vacuum permittivity, and ε_r is the solvent dielectric constant. Over the temperature range considered in this work (273K to 318K), the water dielectric constant can be accurately represented by a quadratic function of the absolute temperature[159]

$$\varepsilon_r(T) = 249.21 - 0.79069T + (0.72997 \times 10^{-3})T^2, \quad (3.20)$$

In Eq.(3.18), the effective charge density can be estimated from the number of charged residues in the region of α-helical hairpins (50-110 aa) of the capsid proteins. This region forms spikes at the capsid surface and is depicted as a large sphere in our dimer model. At the physiological condition, each native HBV core protein contains 8 charged residues in the α-helical domain[153]. The same number was used in an earlier work for the capsid charge without CTD[118]. We assume that these charges are uniformly distributed at a spherical surface of R = 15 nm in radius, approximately equal to the inner radius of the T4 capsid. From the spherical cavity model, we can estimate parameters σ and A_c.

b) Nucleocapsids

We now extend the coarse-grained model for the free energy of empty capsid formation discussed above to nucleocapsids by considering additional contributions due to the confinement of CTD tails and the nucleic acid. While both T3 and T4 HBV capsids may contain ssRNA, ssDNA or dsDNA, depending on the stage of viral replication, here we consider only T4 capsids with ssRNA for simplicity. Similar models can be applied to capsids containing ssDNA or dsDNA.

We assume that, at the physiological condition, every arginine residue from CTD carries one positive charge, every glutamate residue or phosphorylated serine (155, 162 and 170th residue) carries one negative charge, and all other residues from the CTD are electrostatically neutral. Accordingly, the CTD chain of the wild type capsid has 16 positive (Z_i= 1), 4 negative (Z_i= -1), and 14 neutral (Z_i=0) amino-acid residues. To

estimate the molecular excluded volume effects, we represent the CTD chains as tangentially connected hard spheres, each sphere standing for one aa residue. These coarse-grained segments have the same size (with diameter a_c = 0.5 nm) but different electrostatic valences. Similarly, the confined RNA chain is represented by a tangentially connected chain of charged spheres with identical diameter and valence, a_r = 0.75 nm and Z_r = -1, respectively. While recent theoretical studies suggest the importance of the secondary structure of RNA on formation of nucleocapsids,[160, 161] it is known that RNA encapsidations in HBV is not sensitive to the origin of the genome, and the nucleocapsid formation is driven by strong and non-specific RNA-CTD interactions[144]. As a result, we assume that the RNA secondary structure has no effects on capsid formation. In both wild type and E-coli regenerated T4 capsids, the number of RNA segments is about 3400, and the RNA content is slightly less in a T3 nucleocapsid[126].

In comparison to that for empty capsids, the additional free energy for nucleocapsid formation includes contributions due to the change in polymer conformations, molecular excluded volume effects, entropy of mixing, and electrostatic interactions. We may estimate these contributions using a simple model shown schematically in Figure 3-2. The confining effect of the capsid is now considered as a spherical cavity of radius *R*, with the inner space divided into two regions: region 1 is free of polymers and region 2 consists of the nucleic acids binding to the tethered CTD chains. The division between these two regions, *i.e.*, the thickness of the complex region *D*, can be determined by minimization of the complexation free energy

$$\frac{\partial F_{\text{plex}}(D)}{\partial D} = 0, \tag{3.21}$$

We assume that the electrolyte solution in Region 1 is fully equilibrated with the surroundings and makes negligible contribution to the free energy of capsid formation. The assumption is consistent with our coarse-grained model for the formation of empty capsids. For the formation of polyelectrolyte complex between RNA and CTD tails in Region 2, the concentrations of cations and anions, ρ_2^+ and ρ_2^-, are defined by the Donnan equilibrium:

$$\rho_2^+ \rho_2^- = \rho_o^2, \tag{3.22}$$

$$\rho_2^+ - \rho_2^- - \rho_p = 0, \tag{3.23}$$

where ρ_p stands for the total charge density of RNA and CTD chains, ρ_0 denotes the overall number densities of small ions in Region 1. The total charge density of the biomacromolecules is related to their molecular characteristics and the volume of Region 2

$$V_2 = 4\pi(R^2 D - RD^2 + D^3/3). \tag{3.24}$$

The free energy of complex formation reflects the difference between the free energy of the confined polymer chains and those corresponding to isolated RNA and CTD chains at infinite dilution. Approximately, this free energy includes contributions due to the changes in polymer elastic energies (*elas*), segment excluded volume effects (*exc*), electrostatic interactions and entropy of mixing (*els*)

$$F_{\text{plex}} = \Delta F_r^{elas} + \Delta F_c^{elas} + \Delta F^{exc} + \Delta F^{els}. \tag{3.25}$$

We can calculate the change in elastic energy for transferring the RNA from an isolated state to that in the spherical shell from the freely rotating chain model

$$\beta\Delta F_r^{elas} = \frac{\pi^2 n_r a_r^{\,2}}{6D^2}, \tag{3.26}$$

where $n_r \approx 3400$ is the number of nucleotide segments in the RNA chain, and a_r is the monomer diameter. In Eq.(3.26), the proportionality constant accords with the ground-state dominance approximation[162]. Similarly, the conformation free energy associated with the confinement of CTD tails is obtained from[163]

$$\beta\Delta F_c^{elas} = \frac{3}{2} N_t \left(\frac{\mathrm{R}_e^2}{D^2} + \frac{D^2}{\mathrm{R}_e^2} \right), \tag{3.27}$$

where $N_t = 240$ is the number of CTD tails in each T4 capsid, and R_e is the average end-to-end distance for a free CTD chain. Assuming that each CTD tail is a freely rotating chain with a bond angle of 116°, we have $\boldsymbol{R_e^2 \approx 2.56 n_c a_c^2}$ where $\boldsymbol{n_c = 34}$ is the number of amino acid residues and a_c is the segment diameter.

Because the excluded volume is negligible for CTD and RNA chains at infinite dilution, the excluded-volume free energy corresponding to the polyelectrolyte complex can be estimated from the modified Flory theory[133]

$$\beta\Delta F^{exc} = \frac{\nu_r N_r^2 + \nu_{rc} N_r N_c + \nu_c N_c^2}{4\pi R^2 D}, \tag{3.28}$$

where $\boldsymbol{\nu}$ stands for segmental excluded volume. While in principle ν depends on both the segment size and electrostatic interactions[161, 164, 165], we use the "bare" excluded volumes of the polymers on the right side of Eq.(3.28), i.e., $v_r = 4\pi a_r^2/3$, $v_c = 4\pi a_c^3/3$ and

$v_{rc} = 4\pi\{(a_r + a_c)/2\}^3/3$ for RNA, CTD and RNA-CTD interactions, respectively. As detailed in the following, we will take into account the charge effects along with their Coulomb interactions with the small ions.

As in conventional polyelectrolyte complexes, the excess free energy due to electrostatic interactions and entropy of mixing can be calculated from the Overbeek-Voorn (OV) theory [166]

$$\beta F^{els} = -\frac{V_2 \kappa_2^3}{12\pi} + \sum_i N_i \ln \phi_i, \tag{3.29}$$

where κ_2 represents the Debye screening parameter in Region 2, N_i stands for the number of RNA or CTD chains or salt ions, and $\phi_i = n_i v_i N_i / V_2$ represents the volume fraction. Such the excess property determines the deviation from the ideal behavior of elements. For simplicity, we assume that $n_i = 1$ for salt ions and that the ion molar volume is the same as that for water molecules, $v_i = 18$ cm^3/mol. For RNA and CTD chains, the volume of individual segments is related to the segment diameter, $v_i = \pi a_i^3 / 6$.

Because the OV theory includes the electrostatic energy and the entropy of mixing among small ions, we must subtract Eq.(3.30) with that corresponding to the background electrolyte occupying the same volume and the free energies of RNA and CTD chains at infinite dilute. In calculating the background free energy, we use Eq. (3.30) by setting the polymer concentrations to zero and the salt concentrations equal to their bulk values. To use the OV theory for RNA and CTD chains at infinite dilution, we

assume that the electrostatic energy and entropy of mixing are the same as those corresponding to a confined chain in a spherical cage with the cavity radius the same as the polymer gyration radius. According to the literature[167-170], the radius of gyration for a single RNA chain depends on its chain length and the surrounding salt concentration

$$R_{g,r} = 5.84 \cdot n_r^{1/3} C_s^{-1/8}, \tag{3.30}$$

where R_g has the units of Å and C_s in M. We calculate the radius of gyration for each CTD chain according to the freely rotating chain model with a bond angle of 116°

$$R_{g,c} = \left(0.43 n_c a_c^2/6\right)^{1/2}, \tag{3.31}$$

We used different models for RNA/CTD in energetic and geometric characterizations because the former is divided into several interrelated contributions (e.g., elastic, electrostatic, etc.) while the latter (e.g., the radius of gyration) represents the overall behavior.

III. Results and Discussion

In the following, we will first validate the theoretical predictions with experimental data for the formation of empty T4 capsids. Subsequently, we discuss the effects of HBV "modulators" on capsid stability, the thermodynamic basis of T3/T4 dimorphism, and the relative stability of empty and nucleocapsids. Because the equilibrium constants and the free energy of nucleocapsid formation have been rarely quantified, we will present the theoretical predictions in the context of recent experimental investigations. Despite the simplicity of molecular models from a microscopic perspective, we should emphasize that our theoretical analysis of capsid stability entails no adjustable parameters. As discussed above, the theoretical calculations

are all based on the basic characteristics of the pertinent biomacromolecules and protein-protein interactions.

A. Equilibrium constant for the formation empty T4 capsids

We first consider formation of CTD-free empty capsids and compare the theoretical results with *in vitro* experiments. Figure 3-3 shows the equilibrium constants of empty T4 capsids over a large range of temperatures and salt concentrations[136]. Here the symbols represent the experimental data from the *in vitro* formation of CTD-free empty T4 HBV capsids, all lines are from the theoretical predictions. Clearly the coarse-grained model captures both the effects of temperature and salinity on the capsid stability. In light of the simplicity of the theoretical model and noticeable scattering in the experimental data, the quantitative agreement between the theory and experiment is remarkable. Our theoretical results further support the conjecture that capsid formation is mainly driven by hydrophobic interactions[136] and that the contact hydrophobic free energy can be quantitatively described with the solvation shell model[155].

Although the hydrophobic attractions are mainly responsible for capsid formation, it is worthwhile mentioning that electrostatic interactions and van der Waals forces are indispensable for the quantitative performance of the theoretical model. For example, Figure 3-4 shows the equilibrium constant versus temperature at 0.15 M salt condition. Excluding van der Waals association free energy (F_{vdw}) would result in a systematic deviation from the experimental values. The sensitivity of the numerical results is related, in part, to the partial cancelation of the electrostatic and hydrophobic interactions and to the large number of proteins for the formation of capsid. Besides, the sensitivity also

reflects the exponential relationship between the formation free energy and the equilibrium constant.

B. Modulation of capsid assembly and stability

Recently several strategies have been suggested to design new drug molecules for HBV by controlling the kinetics of capsid assembly and stability. For example, heteroacryldihydropyrimidine (HAP) is well known as a potent HBV capsid modulator, which may misdirect the capsid formation[83-85]. Because of the potential therapeutic values of such molecules, it is worthwhile testing whether our molecular thermodynamic model is able to capture their effects on the capsid formation.

According to previous publications[83-85], HAP molecules are able to intercalate at the dimer-dimer contact regions of the capsid proteins and the HAP intercalation enhances dimer-dimer association. HAP disrupts capsid formation mainly by kinetic trapping of the assembly intermediates. HAP molecules strengthens hydrophobic association between capsid proteins by keeping the dimer-dimer distance unchanged while filling the void hydrophobic regions[84]. Such wedging effect is naturally accounted for within our hydrophobic model for the capsomers association. We expect that the modulation effect would be most significant if the size and shape of the hydrophobic molecule could match those corresponding to the hydrophobic pores of two-fold symmetry (Fig. 3-1B &C) or wedge between dimer-dimer contact regions.

We may also predict a situation such that hydrophilic molecules are able to bind the dimer contact regions of the capsid proteins. In this case, the association energy would be decreased due to the reduction of the dimer contact area. Figure 3-5 shows that

a reduction of hydrophobic patch area of the protein dimers by 10% would decrease the capsid stability by ~200 of *lnK* value. For comparison, we may consider the effect of electrostatic interactions on the capsid stability by increasing either the charged surface area or the surface charge density. Figure 3-5 shows that the influence is much less significant in comparison to the hydrophobic effect. According to our thermodynamic model, an efficient HBV "modulator" should bind specifically to the base part of the HBcAg assembly domain while enhance repulsion among the protein subunits, such as an amphiphilic molecule with a hydrophobic binding domain associated with a charged group or a hydrophilic chain.

The hydrophobic interactions, and subsequently the capsid assembly, can also be influenced by mutations on certain core protein residues. We may describe the mutation effect in a way similar to that for the changes in the hydrophobic potential. For example, a recent study by Alexander and coworkers [140] indicates that 3 critical mutations, F23A, L42A, and Y132A, significantly inhibit capsid formation *in vitro*. It shows that the equilibrium composition of the assembled capsids decreases to 1~2% in comparison to 80% for the wild type. In particular, F23 and Y132 mutations greatly hinder the inter-dimer interactions. It was speculated that L42 mutation induces the allosteric structural change on the capsid protein dimer, not directly altering the inter-dimer contact because L42 is close to the surface exposed region. According to our thermodynamic model, the inter-dimer hydrophobic interaction is directly related to the hydration shell of the hydrophobic domain. The change of residue F23 or Y132 at the inter-dimer contact region into alanine, a non-hydrophobic and also non-aromatic small residue, decreases

the size of the hydrophobic patch domain thereby reducing the hydrophobic attractions. Besides, L42 mutation may alter the hydrophobic interactions by allosteric structural changes. According to our thermodynamic model, the hydrophobic interaction depends on the proper arrangement of the hydrophobic domains of the capsid proteins. In principle, we may estimate the mutation effects by quantifying their modulation on hydrophobic interactions.

C. T3/T4 dimorphism

T3 and T4 capsids occur concurrently during both *in vivo* and *in vitro* syntheses of HBV capsids, and the dimorphism reflects the difference in the free energy of capsid formation. In a T3 capsid, each dimer maintains the same number of contact neighbors as that in a T4 capsid (see Figure 3-1). As the result, it appears that the interactions between small and large dimer units have little influence on T3 and T4 dimorphism except that the number of dimers being 90 in T3 instead of 120 in T4. However, the size of an empty T3 capsid is slightly larger than that would be expected according to the degree of association. The average radius of a T4 capsid is around 15 nm, and that of T3 capsid is 13.3 nm[5]. Accordingly, the ratio of spherical surface area is $13.3^2 : 15^2 = 0.786 : 1$, which is slightly larger than the ratio for the number of capsomers (0.75 : 1). Such a loosely assembled structure can be depicted in our capsid model by setting two small units at each dimer contact slightly farther apart, i.e., by setting the distance between small units to 1.5 nm for a T3 capsid instead of 1.4 nm as for the T4 capsid. Table 3-1 presents the hydrophobic contact energy and the electrostatic repulsion energy calculated from our thermodynamic model. Due to the reduced hydrophobic potential, the free energy for T3

capsid formation per each monomer (-4 k_BT) is significantly less negative that that of T4 form (-7.7 k_BT). The higher energy state for T3 capsids explains the prevalence of T4 capsids at both *in vivo* and *in vitro* conditions[5]. The higher energy for T3 capsids agrees with an earlier analysis of the capsid stability in vacuum[171].

It has been suggested that the capsid protein plays an important role in T4/T3 dimorphism [5, 126]. Without CTD, T4 capsids are more prevalent than T3 capsids. However, truncation of more than 34 C-terminal residues inverses the relative composition, i.e., T3 capsids may be preferred than the T4 form at the same assembly condition[5, 126]. We demonstrated above that the coarse-grain model provides a good description of empty capsid formation without CTD (CP149). For comparison, we have also quantified another truncation case that only 3-144 residue of the wild type capsid protein (denoted as 'HBeAg') are maintained. Such a system was studied many years ago by Wingfield et al.[126]. Following their data, the mean radius of HBeAg T3 capsid is 12.9 nm, while that for T4 is 15.1 nm. Because the ratio of spherical surface area is now slightly smaller than the dimer number ratio, the T3 HBeAg capsid is more tightly assembled in comparison to that for a conventional HBcAg capsid. We estimate that the contact distance between nearest dimers (r) is about 1.38 nm. Although the change of r is small (about 10% of the small unit overlap distance), the contact free energy per each monomer increases up to -8.6 k_BT, which makes prevalence of the T3 capsids (Table 3-1). As discussed in the following, the dimorphism of HBV capsid is also sensitive to the variation of C-terminal domain of the capsid proteins.

D. Stability of T4 nucleocapsids

Figure 3-6 shows the free energy for the encapsidation of RNA and CTD chains versus the thickness of the polyelectrolyte complex, D. As expected, the free energy shows a minimum at an optimal D, which varies with the solution conditions. In these calculations, we use R =13 nm for the inner radius of the HBV capsid, and calculate the ion concentrations from the Donnan equilibrium with polymer charge density ρ_p = $520/V_2$. For the range of bulk salt concentrations (0.1 ~ 0.7 M) considered in this work, the optimal complex thickness lies around 10.5 ~ 11 nm, much larger than that detected in cryo-EM measurements. As demonstrated in our previous work[172, 173], the complex structure can be captured with more sophisticated polymer theories that account for the CTD tethering effect. Nevertheless, it is shown that the suggested formulation on viral components in this work is effective enough to depict the swelling of the complex in response to salt concentration, which is a characteristic feature in the polyelectrolyte system. Moreover, this level of approximation is expected to be a relatively efficient way to calculate the strong electrostatic interaction between RNA and CTDs.

Figure 3-7 shows the effects of salt concentration on the total free energy of CTD/RNA encapsidation. For all conditions considered in this work, F_{plex} is negative, indicating that RNA encapsidation generally stabilize the T4 capsid. The nucleocapsids are thermodynamically more stable than the corresponding empty capsids due to the encapsidation of RNA and CTD chains. As expected, the complex formation free energy becomes less negative as the salt concentration increases. Experimental studies confirmed that the nucleocapsids are relatively more stable than the corresponding empty capsids[126,

[145], and such an enhanced stability has been implied by a previous theoretical investigation[173]. Figure 3-7 also shows that the encapsidation free energy approaches to zero at sufficiently high salt concentration (C_s), and the stabilizing effect of complex formation is more effective when C_s is relatively low. In Figure 3-8, we compare Debye screening length in the complex region with that in bulk salt condition. It shows that the deviation of the parameter from bulk state approaches to small as C_s increases, implying the stabilization by electrostatic interaction is to be reduced by the charge screening. Experiments on the effect of salt on the interaction between capsid proteins and the viral genome indicate that the complex binding energy decrease with higher salt concentration[145].

Figure 3-9 shows the thickness of RNA-CTD complex inside the nucleocapsid versus the salt concentration. The reduced complex formation free energy is responsible for the swelling of the RNA-CTD complex at high salt concentration. We may explain the overall salt effect on the nucleocapsid stability as a balance of protein-protein and RNA-CTD interactions. While the salt ions reduce the electrostatic repulsion between capsid proteins, they also reduce the electrostatic binding energy between CTD and RNA chains. As a result, nucleocapsid assembly would be most favorable at an intermediate salt condition. Indeed, an intermediate salt condition (0.25 ~ 0.35 M) must be applied for *in vitro* nucleocapsid assembly (e.g., ref[145, 174]). Moreover, our result provides key information in evaluation of experimental data which are conducted in a wide range of ionic condition. For several *in vitro* analyses on viral capsids, it is required to specify a composition of ionic elements according to the purpose of measurement. If the analyzing

result is supposed to be related with the capsid stability, then the salt effect is to be significant in overall assay. Electrophoresis or chromatography is one of key methods to purify the capsid and calibrate its thermodynamic property. Unless we properly consider the resultant effect of ionic strength, there is a possibility that a certain portion of less stable capsids might be selected out during a course of the analysis. In addition, a quantification of viral genome in a certain evolving stage of nucleocapsid may be varied due to the salt effect, not only by a direct role of ionic elements in genome replication process but also by a possibility that the amount of nucleocapsid product might depend on the salt concentration.

We have also evaluated the effects of RNA-CTD interaction on the dimorphism of HBV nucleocapsids. It is well known that T3 nucleocapsids are rare (< 5%) in both *in vivo* or *in vitro* assembly[175]. To check the dimorphic probability using our nucleocapsid model, we estimate the pgRNA and CTD encapsidation free energy in a way similar to that for T4 nucleocapsids. For T3 nucleocapsids, we set the radius of the inside shell to 11 nm, considering that the capsid thickness is about 2 nm. Whereas the encapsidation free energy on T4 capsid is to be obtained in a self-consistent way with respect to an optimal thickness D for RNA-CTD complex region, the complex region for a T3 capsid tends to expend up to all of inside capsid lumen without maintaining a certain point that retains a free energy minimum. The absence of a polymer-free Region 1 is probably because the excluded volume effect is too large to package pgRNA into a narrow T3 capsid. A recent study on *in vitro* encapsidation of nanoparticles also indicates that the volumetric restriction makes T3 capsids less stable than T4 capsids[176]. For pgRNA

encapsidation in wild-type T3 capsids, our model predicts that the free energy of CTD-RNA complex formation, F_{plex} = -1863 k_BT at 37°C with 0.15 M salt condition. At the same condition, F_{plex} = -4582 k_BT for a T4 nucleocapsid. Because the stabilizing effect by RNA-CTD complex formation is much smaller for T3 nucleocapsids, we ascertain that the energetics in genome encapsidation results in the prevalence of T4 nucleocapsids.

IV. Conclusions

We present a comprehensive thermodynamic model for describing the stability of various Hepatitis B virus (HBV) capsids that have been subjected to a number of recent experimental studies. The systems considered include empty T3 and T4 capsids without the C-terminal domain (CTD), T3 and T4 nucleocapsids with CTD and RNA encapsidation, and capsids with various site mutations. Our thermodynamic analysis is based on the quasi-chemical equilibrium for capsid formation from the core protein dimers with or without the presence of nucleic acids. The free energy is described in terms of an icosahedral lattice model for the arrangement of core proteins, which are represented by hard-sphere trimers mimicking the protein topology, and a coarse-grained model for CTD and RNA encapsidation. We developed analytical expressions to quantify various thermodynamic driving forces of the self-assembly process such as hydrophobic, electrostatic, van der Waals, biomolecular conformation, excluded volume and mixing effects. We find that capsid formation is dominated by entropy changes thereby involving little heat effects. Whereas the association between the core protein dimers is mainly driven by hydrophobic interactions, electrostatic attraction between CTD and RNA chains is accompanied by the release of a large number of counterions.

We have validated the numerical performance of the molecular thermodynamic model by extensive comparison with experimental data for *in vitro* assembly of T4 capsids over a broad range of temperatures and salt concentrations. While the capsid assembly is mainly controlled by the competing interactions of hydrophobic association and electrostatic repulsion, van der Waals attraction among the core proteins is indispensible for the quantitative theoretical performance. The van der Waals free energy was often omitted in previous investigations of viral assembly because of its relatively small strength in comparison to the hydrophobic attractions.

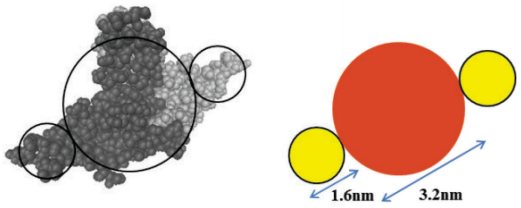
An important feature of our thermodynamic model is that all parameters can be independently estimated from the molecular characteristics of the key components of the viral capsid. In particular, the solvation shell model for hydrophobic interactions allows us to rationalize different chemicals for efficient modulation of the capsid stability. Our model predicts that a mutation on the hydrophobic domain may significantly alter the capsid stability and thus the process for the viral assembly. Although precise experimental data are not yet available for direct comparisons, it appears that our model captures the qualitative features of the experimental results for stability modulation by either capsid mutation or chemical binding. We expect that this model may be utilized to guide further experiments to optimize for the mutual sites or the design of efficient capsid modulators.

We have also studied the relationship between the HBV capsid stability and T3/T4 dimorphism for both empty capsids and nucleocapsids. Although the connectivity among the core protein dimers is rather similar, T3 and T4 capsids have slightly different

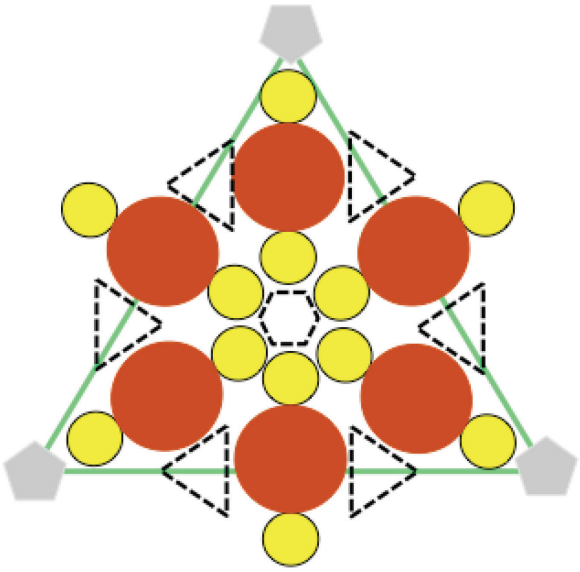
protein packing density thereby different assembly free energy per protein. The dimorphism of HBV capsids can be explained by analyzing the free energy difference on the basis of the structural data. Our theoretical model predicts that, in good agreement with experiments, T4 capsids are energetically more favorable than T3 capsids for *in vitro* assembly of CTD-free empty capsids and for all nucleocapsids. However, the T3 form becomes more favorable if the capsid retains only up to 144^{th} C-terminal residue.

Finally, we have studied the stability of nucleocapsids (NC) by taking into account all pertinent biomolecular components including capsid proteins with CTD, nucleic acids, and small ions in the electrolyte background. The interactive energies on those confined molecules resemble those appeared in a polyelectrolyte complex and can be described qualitatively with the Overbeek-Voorn (OV) theory. We ascertain that RNA encapsidation enhance the NC stability in both physiological and conventional *in vitro* assembly conditions. The NC is most stable at an intermediate salt concentration because the electrostatic interaction disfavors capsid formation at low concentration and RNA and CTD binding decreases at high salt condition. The effect of CTD-RNA complex formation is more predominant for T4 than T3 capsids, which explains the overwhelming population of T4 capsids at both *in vivo* and *in vitro* conditions.

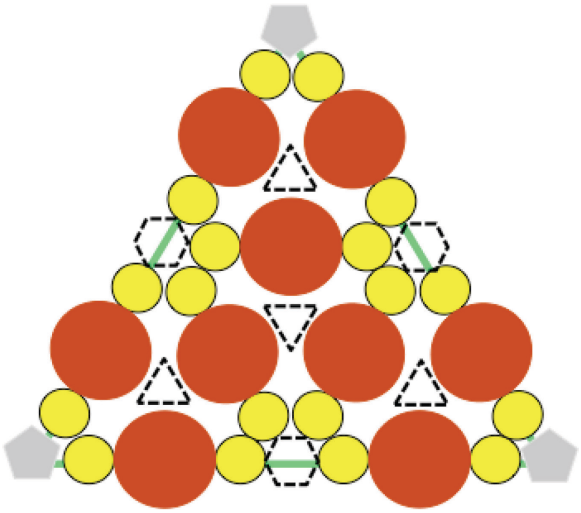
Figure 3-1 (A) A course-grained model for HBV capsomers. Each capsomer consists of two core proteins (HBcAg), depicted as 3 tangentially connected spheres with their radii determined from the protein crystal structure. (B) A lattice model for the arrangement of capsid proteins at a triangular facet of a T3 capsid. (C) The lattice model for a T4 capsid. In both cases, the major capsid holes are shown as hexagons and triangles.



(A)



(B)



(C)

Figure 3-2 A schematic representation of HBV nucleocapsid containing CTD and RNA chains. Here *D* denotes the thickness of Region 2 and *R* the capsid radius. The electro-microscopy image of HBV capsid is adapted from the literature[174].

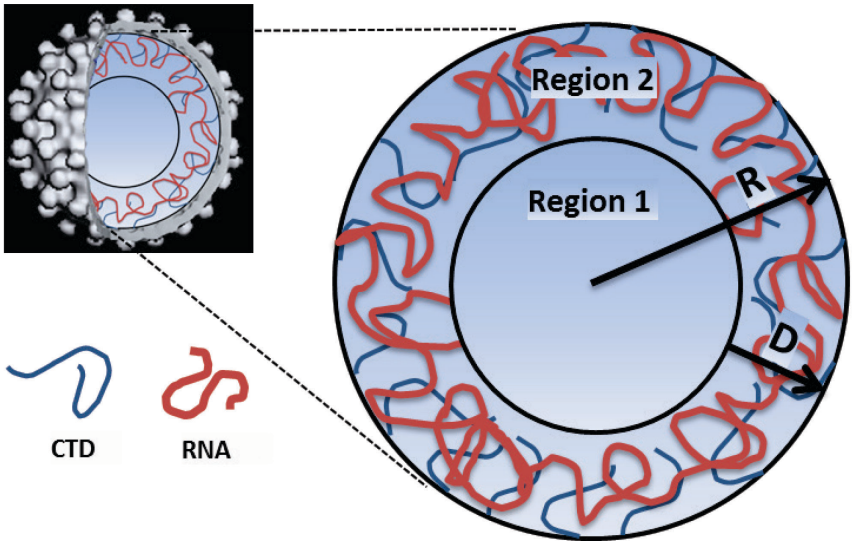


Figure 3-3 The equilibrium constant for the formation of CTD-free empty T4 capsids. Symbols are experimental data[136] and lines are theoretical predictions. The salt concentration (M) is denoted beside each line.

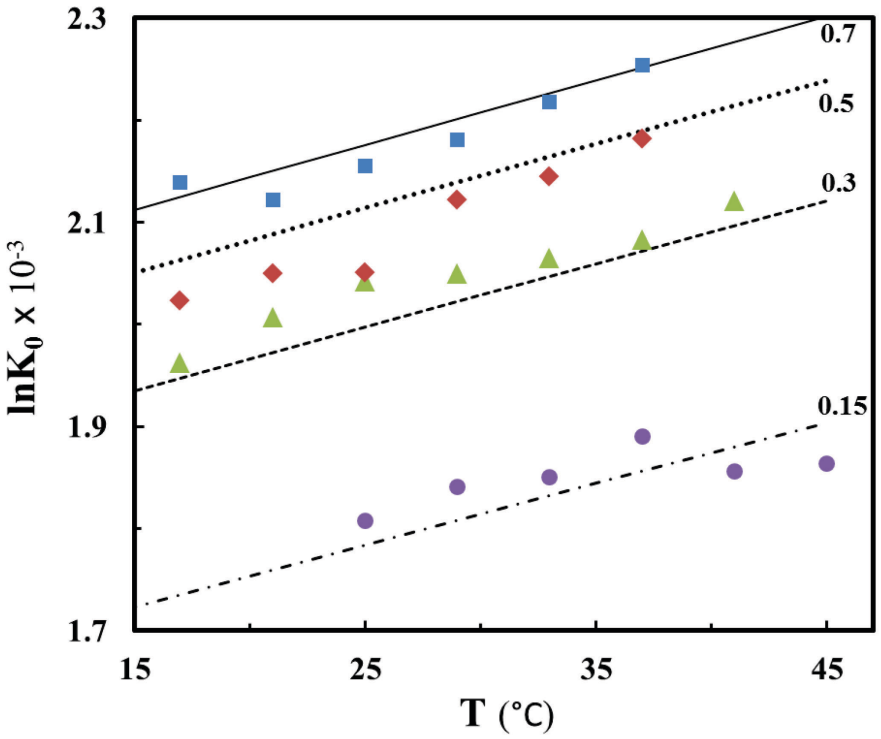


Figure 3-4 The effect of van der Waals force on capsid stability. The symbols and the dashed line are the same as those shown in Figure 3-3, and the solid line represents lnK_0 without considering van der Waals interactions.

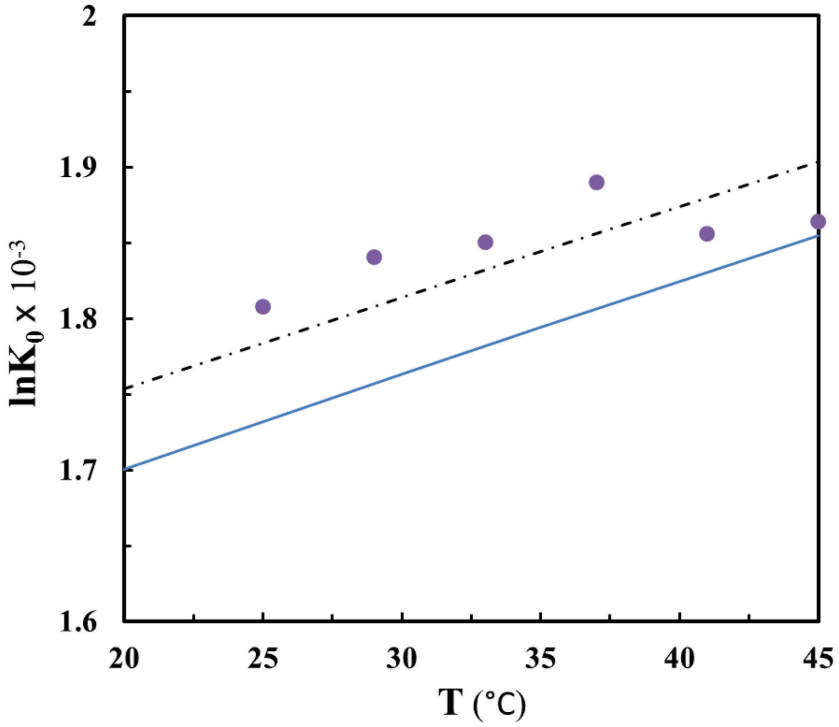


Figure 3-5 Hydrophobic and electrostatic effects on the equilibrium constant of empty HBV capsids. The solid line corresponds to the theoretical results if there is a 10% reduction of the buried hydrophobic area at each dimer contact; the dotted line is for a 10% increase of the capsid charged area; and the dashed line is for an increase of one unit charge for each capsid protein. For comparison, also shown in this figure are the equilibrium constants of wild-type capsids from theory (WT, dash-point line) and experiment (circle points) at 0.15 M.

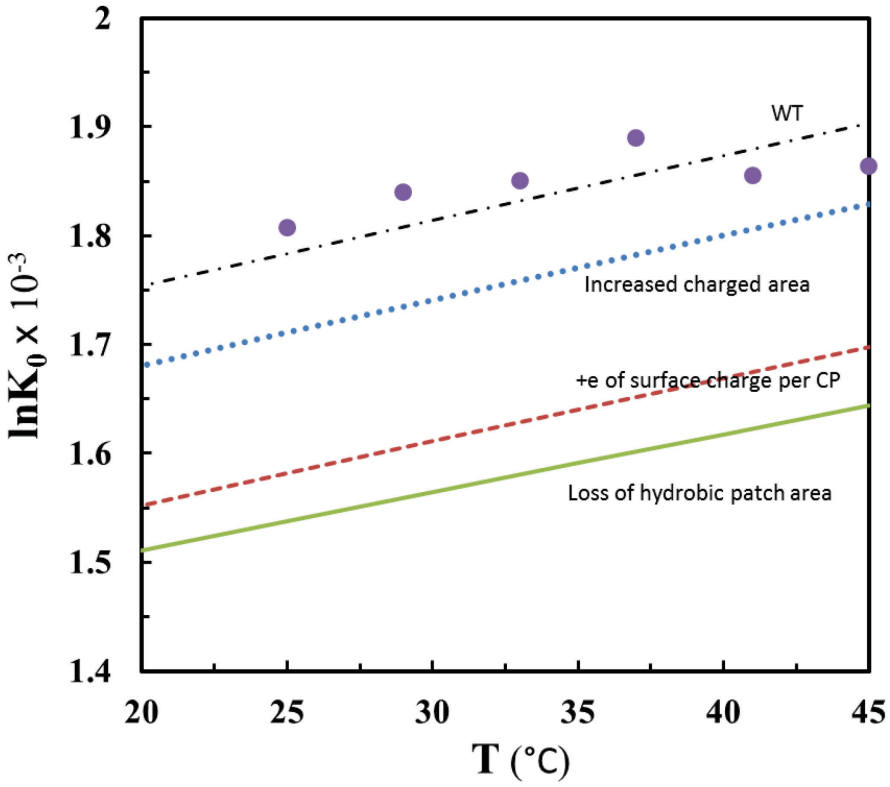


Figure 3-6 The free energy of encapsidation for CTD and RNA chains as a function of the thickness of the complex layer. The concentration (M) is denoted beside each line, and the minimum points are indicated by triangle symbol.

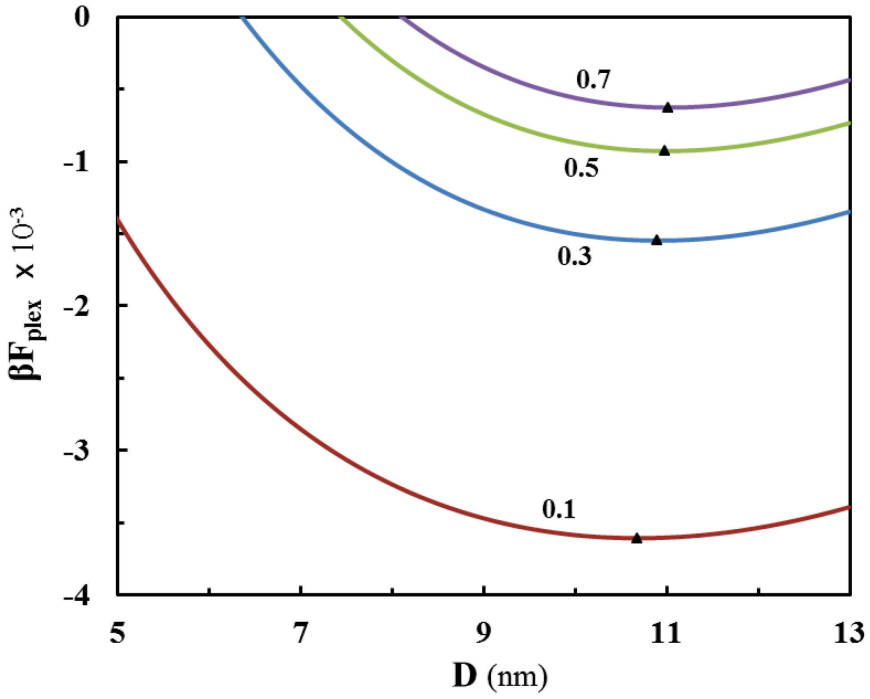


Figure 3-7 The encapsidation free energy for CTD and RNA chains in a T4 capsid (βF_{plex}) as a function of the salt concentration (C_S).

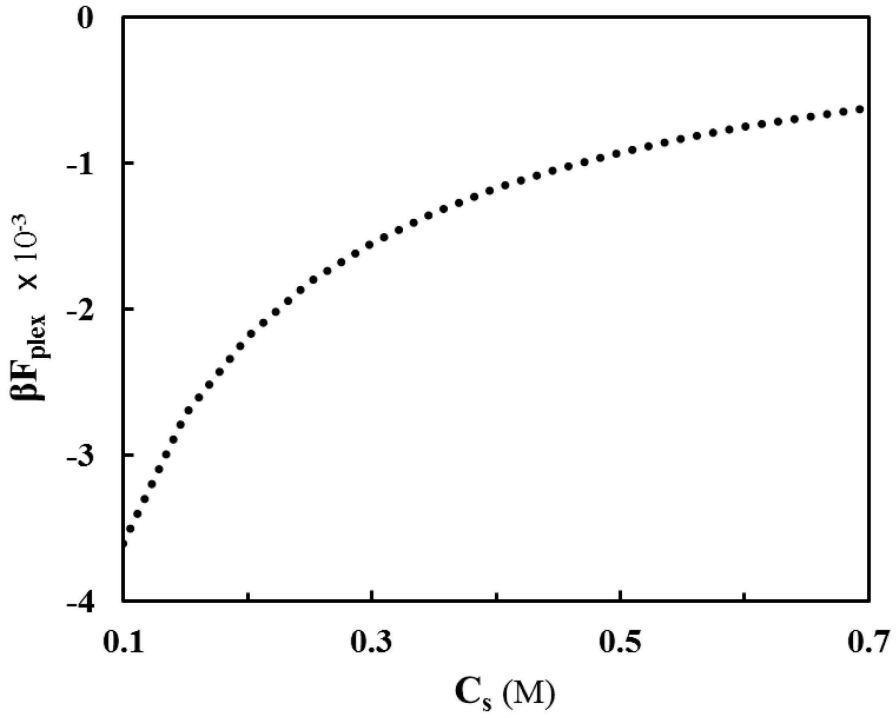


Figure 3-8 The Debye screening length (κ^{-1}) in the region with CTD-RNA complex (Region 2) and in the bulk as a function of the bulk salt concentration (C_S).

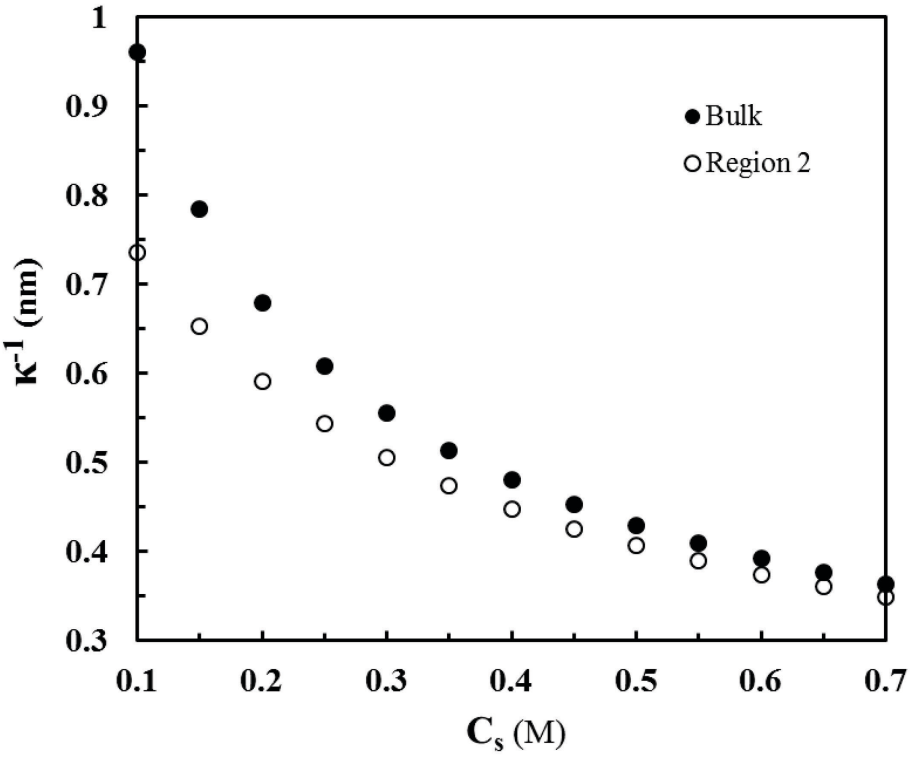


Figure 3-9 The thickness of CTD-RNA complex region (D) as a function of the salt concentration (C_S).

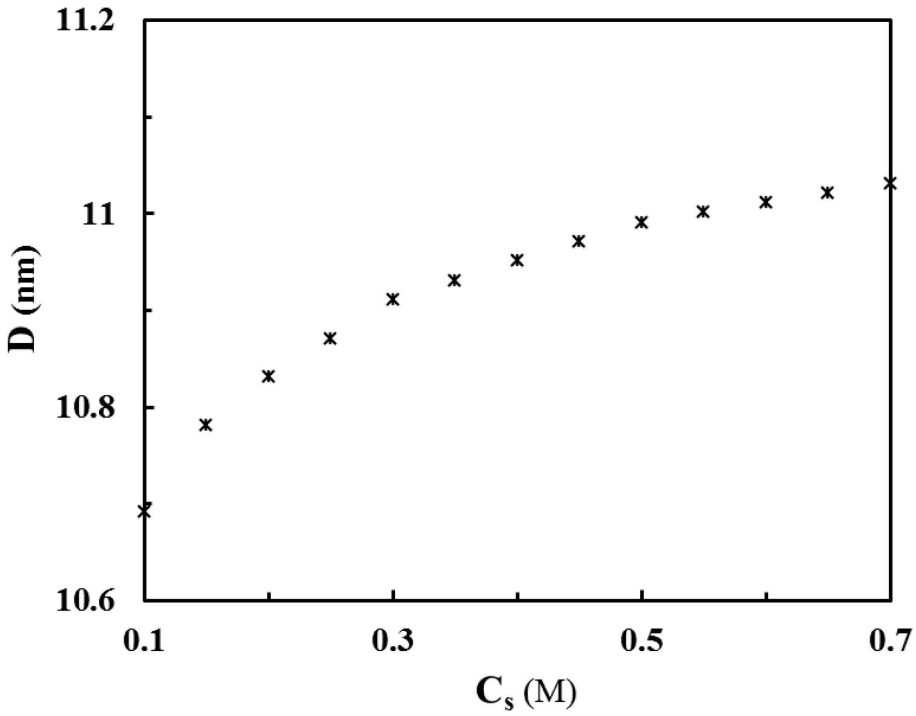


Table 3-1. Free energy of capsid formation

	T4 (Cp149[*], HBeAg[**])	T3 (Cp149[*])	T3 (HBeAg[**])
Number of monomers (n)	240	180	180
R_{ave}	15	13.3[*]	12.9[**]
F_{cap}	-1856.19	-725.99	-1540.13
F_{cap}/n	-7.73	-4.03	-8.55
$F_{hyd+vdw}$	-2592.44	-1252.77	-2100.08
$F_{hyd+vdw}$ /n	-10.80	-6.96	-11.67
F_{ele}	736.25	526.78	559.95
F_{ele} /n	3.07	2.93	3.11

The average radius (R_{ave}) is obtained from Zlotnick et al. for Cp149 capsid[*] [5], and from Wingfiled et al. for HBeAg capsid[**][126]. Unit of R is *nm*. The free energy of capsid formation, F_{cap} , and its division into hydrophobic and van der Waals interactions, $F_{hyd+vdw}$, and electrostatic interaction F_{ele} are compared with the corresponding values per monomer for both T4 and T3 capsids. Here the unit of energy is k_BT, and the solution condition is 37°C and 0.15M salt.

Chapter 4. A theoretical study on SRPK interaction with the flexible domains of Hepatitis B capsids

Abstract

Hepatitis B virus (HBV) regulates genome encapsidation and reverse transcription from a single-stranded RNA to a double-stranded DNA through the flexible C-terminal domain (CTD) of the capsid proteins. While the microscopic structure of the nucleocapsid (NC) plays a critical role in the life cycle of HBV, the location of CTD residues at different stages of viral replication remains poorly understood. In this work, we report the radial distributions of individual amino-acid residues of CTD tails for both empty and RNA-containing HBV capsids by using a coarse-grained model for the key biological components and the classical density functional theory (DFT). The DFT calculations reveal substantial exposure of the CTD residues outside the capsid shell, in particular when the capsid is devoid of any nucleic materials. The outermost layer of the capsid surface mainly consists of residues from ^{170}Ser-^{175}Arg of the CTD tails, *i.e.*, the serine-arginine protein kinase (SRPK) binding motif. The theoretical description corroborates with recent *in vitro* studies that show transient CTD distribution captured by SRPK binding. We have also investigated the NC structural changes due to phosphorylation of serine residues and demonstrated a correlation between the CTD location and the internal distribution of RNA.

Introduction

Hepatitis B virus (HBV) is a human pathogen infecting around 2 billion people worldwide. Acute and chronic HBV infection may cause serious dysfunctions including liver fibrosis, cirrhosis and cancer[177]. The HBV virions, also known as Dane particle, have a double-shelled structure consisting of surface proteins and nucleocapsid (NC) containing a circular partially double-stranded DNA (dsDNA). The HBV capsid is formed by the self-assembly of the capsid proteins with a pregenomic RNA (pgRNA) and the viral polymerase (Pol), a reverse transcriptase for transformation of pgRNA to dsDNA[2, 46, 178]. Each NC has an icosahedral structure with triangulation number T = 3 or 4. T4 is the major form (95%) *in vivo*, which is made of 240 copies of the core protein (CP)[5, 6, 179]. A wild-type CP has 183 amino-acid (aa) residues, including an assembly domain (1-140 aa), a nonapeptide linker (141-149 aa), and the C-terminal domain (CTD, 150-183 aa). The assembly domain has a rigid structure that contributes to the capsid scaffold. However, both the linker and the CTD chains are flexible; the latter is referred to as 'protamine tails' because they are rich in arginine residues[7, 8].

It has been long recognized that CTD plays an essential role in the life cycle of HBV[7]. The capsid shell is fenestrated with symmetrical pores that allow mass transport and the extrusion of the CTD tails[25-27]. The exposure of the CTD tails affects genome packaging, nuclear entry, as well as capsid envelopment and egress[7, 11, 12]. In addition, it has been recognized that the CTD tails are involved in signaling capsid maturation[22, 28, 29]. Because CTD exposure alters the surface properties of the NC particle, its location at different stages of replication is critical to differentiate the mature and immature NCs. If

the CTD chains were truncated by mutation, the CPs would be unable to encapsidate the full content of the pgRNA, generating capsids with a smaller amount of DNA or empty capsids[7]. Besides, each CTD chain has three phosphorylation sites (S155, S162 and S170) that are important for RNA encapsidations and reverse transcription[7, 11]. For wild-type viruses, the CTD chains are phosphorylated at the early stage of replication (with pgRNA or intermediate DNA)[12], but dephosphorylated for matured capsids (with dsDNA)[13, 14]. Replacing three serine residues with charged amino acids to mimic the transient phosphorylation state affects pgRNA encapsidation and DNA synthesis. A phosphorylated analog with replaced serine by glutamic acid succeeded to package the pregenomic RNA but failed to support viral DNA replication[11]. Meanwhile, a substantial reduction of RNA packaging was observed in a case of aspartic acid substitution to mimic phosphorserine[15, 16]. The CTD chains have also an important role in viral replication because only phosphorylated capsids were allowed to transport into cellular nucleus[19]. Despite such significance, the kinase responsible for the CTD phosphorylation remains little known. Because all phosphorylated serine residues at CTD are followed by repeated arginine residues, serine/arginine protein kinase (SRPK) was suggested as one of the candidates [180, 181]. SRPK is effective to phosphorylate capsid proteins *in vitro*[10].

In order to clarify how CTD chains regulate genome evolvement as well as intracellular trafficking of HBV capsids[19, 28], it is imperative to determine the spatial distribution during different stages of replication. Regrettably, experimental detection of the CTD structure has been very difficult. Whereas the assembly domain of HBV capsids maintains a relatively rigid structure, CTD chains are highly flexible and their structure is

changing in response to the evolving genome[19-24]. Recently, the enzymatic recognition of SRPK was utilized to analyze the location of CTD tails [182]. The *in vitro* experiments examined the binding affinity of SRPK with the CTD of HBV capsids. It was found that CTDs are exposed at the surface when the capsid does not include any nucleic materials. Conversely, the CTD tails are mainly distributed inside the capsid for the RNA-containing nucleocapsids[182]. However, the CTD chains were not visible through a conventional microscopy such as cryo-EM[6, 183]. Only a few recent studies capture a trace amount of CTD chains through reconstructed EM images[24, 184]. Besides, HBV capsids expressed in *E. coli* contain the bacterial RNA instead of pgRNA[9, 10 19, 24, 182]. The *in vitro* assembly of HBV capsids typically occurs at an electrolyte concentration much larger than that corresponding to the physiological condition. The high salinity is required not only to avoid capsid aggregation but also to maintain the thermodynamic stability of the reassembled capsids[10]. While DNA-containing viral particles are available from expression in mammalian cells at physiological conditions, a discrete separation of the capsids at each maturation state is currently impracticable. Detection of the capsid structure is also complicated by the intracellular environmental effects[185].

Previously we developed a theoretical platform for predicting the microscopic structure and thermodynamic properties of HBV nucleocapsids using a coarse-grained model for the key viral components and the classical density functional theory (DFT)[112, 186]. Our model predicted a quantitative relationship between the CTD length and the genome content in good agreement with experimental data. In this work, we apply the same model to locate the SRPK binding motif by analyzing its affinity with HBV capsids

at different distributions of CTD tails. Specifically, we consider CTD distributions in HBV capsids either with or without RNA as examined in recent experiments[24]. In addition, we examine the structural reorganization of RNA chains due to CTD phosphorylation.

Molecular model and methods

As in our previous work[112, 186], we use a coarse-grained model to account for non-specific interactions among the key ingredients of the HBV capsid, *i.e.,* RNA and CTD chains in an electrolyte background. Whereas the atomistic details, including the secondary structure of RNA, are unquestionably important to the biological functions and the kinetics of encapsidation, the essential features of the genome packaging and its interaction with the flexible domains of the capsid proteins are mainly affiliated with electrostatic interactions and molecular excluded volume effects. The nature of non-specific interactions can be justified by the fact that HBV capsids expressed in *Escherichia coli* have RNA content similar to that in the wild-type capsids[27].

In our theoretical calculations, the electrostatic interactions and molecular excluded-volume effects are described with tangentially connected chains of hard spheres that each corresponds to one amino-acid residue or nucleotide. We assume that all RNA segments have the same diameter, $\sigma_R = 0.75nm$, and electrostatic valence of $Z_R = -1$[187]. The diameter of CTD segments, $\sigma_T = 0.5nm$, reflects the average van der Waals diameter of amino acid residues[188]. The electrical charge corresponds to the characteristic value of each amino-acid residue under physiological condition, and is assigned with the

valence of +1, 0 and -1. Specifically, $Z=+1$ is assigned to arginine, -1 for glutamate, and all other residues carry no net charge.

The primitive model of electrolyte solutions is used to describe the salt ions and water molecules in the background. As in the experimental work[24], the surrounding environment for the HBV capsids is represented by a *NaCl* solution at concentration of 0.14 M. Approximately, Na^+ and Cl^- have the diameters of $\sigma_{Na}=0.39nm$ and $\sigma_{Cl}=0.36nm$, respectively, and the dielectric constant for water is 78.4[189]. The pair potential between ions includes a hard-sphere repulsion and a Coulomb energy:

$$\beta u_{ij}(r)=\begin{cases}\infty, & r<\sigma_{ij}\\ \dfrac{z_i z_j l_B}{r}, & r\geq\sigma_{ij}\end{cases} \tag{4.1}$$

where $\beta=1/k_BT$, k_B is Boltzmann constant, T=298K, $\sigma_{ij}=(\sigma_i+\sigma_j)/2$, and $l_B=0.714$nm is the Bjerrum length. Eq.(1) is also applied to the pair interaction between the polymeric segments. The bonding potential between the coarse-grained RNA segments is described by the Dirac δ-function:

$$\exp[-\beta V_{RNA}(R)]=\prod_{i=1}^{M-1}\frac{\delta(|r_{i+1}-r_r|-\sigma_R)}{4\pi\sigma_R^2} \tag{4.2}$$

where M is the degree of polymerization, and $R=(r_1,r_2,\cdots r_M)$ specifies the positions of monomeric segments. The CTD tails (and 9 aa linkers) are also modeled as tangentially connected spheres of the same size but different valences.

As mentioned early, the HBV capsids are fenestrated with pores about $\sim 1.5nm$ in diameter, allowing the permeation of salt ions and CTD tails but not the RNA genome[9,

[26, 190]. The inner and outer radii of the capsid are $R_{in} = 13nm$ and $R_{out} = 15nm$, respectively [191]. The confinement effect for the RNA segments can be described as a spherical cavity with a hard-wall potential

$$\phi_{RNA}(r) = \begin{cases} \infty, & r > R_{in} - \dfrac{\sigma_R}{2} \\ 0, & otherwise \end{cases}. \tag{4.3}$$

The linker-CTD tails are partially confined with the first linker segment tethered at the inner surface of the capsid. Whereas certain capsid pores are more favorable for tail permeation[9, 26, 190], we estimate the overall probability (10%) considering the approximated ratio of total pore area to inner capsid area. Accordingly, we set an external potential for the tail segments,

$$-\beta\phi_{Tail}(r) = \begin{cases} \ln(0.1), & R_{in} - 0.5\sigma_T < r < R_{out} + 0.5\sigma_T \\ 0, & otherwise \end{cases}. \tag{4.5}$$

We assume the charges on the capsid are uniformly distributed over a spherical surface located at $r = 15nm$. The small ions are distributed either inside or outside the HBV capsid. According to previous investigations[63, 192], the charge density of the capsid is $Q_c = 0.7e/nm^2$, where e denotes unit charge. The linker-CTD tails are tethered on the inner capsid surface.

Application of the DFT to electrolyte solutions and polymers has been well documented [193-196]. In our previous work[186, 189], we demonstrated that the DFT performs well for the electrostatic interactions and the excluded volume effects.

Results

Empty capsid versus RNA-containing NC

We consider first the radial distributions of the polymeric segments and small ions for empty and RNA-containing capsids. Figure 4-1 shows that, in the absence RNA, the CTD chains distribute like that of a typical polymer brush on both sides of the capsid. The oscillatory density profile within the protein shell reflects the molecular excluded volume effects amplified by an artificial wall potential. The distributions of small ions outside the capsid surface constitute an electric double layer (EDL), *i.e.,* a strong accumulation of counterions accompanied by depletion of coions. Different from a typical EDL, however, the counterions are not directly in contact with the surface charge. Instead, counterion accumulation arises mainly from the CTD brush distributed inside the capsid. Within the CTD brush, the counterion density is significantly lower than the density of polymer segments. As a result, it swells significantly more in comparison with a normal polyelectrolyte brush[104]. The concentrations of cations and anions are nearly identical at the capsid center, where the salt concentration is virtually the same as that in the bulk solution. Because the CTD chains are tethered near the capsid pores, they do not contribute to the Donnon equilibrium. As a result, there is no uneven distribution of free ions far from the capsid shell.

Figure 4-1(b) shows that the microscopic structure of the RNA-containing capsid. Interestingly, encapsidation of the RNA chain does not lead to significant changes on the density profiles of cations and anions within the EDL at the capsid out surface. However, the CTD brush outside the NC virtually disappears, indicating that the CTD tails are

mostly confined within the NC due to the strong electrostatic attraction between RNA and CTD chains. The microscopic structure of the NC inside is also quite different from that of the empty capsid. First of all, the counterions within the CTD brush are virtually replaced by the RNA segments. Indeed, the counterions are almost depleted as one may expect from the Donnon equilibrium. Because the CTD chains are mostly confined with the capsid, the brush in NC is slightly thicker than that in the empty capsid. Because the CTD chains are rather extended within the NC, the RNA segments distribute almost uniformly within the brush. Such structure inside the NC cannot be detected with conventional cryo-EM methods but probably measureable with neutron or x-ray scattering.

The location of CTD residues

A recent *in vitro* study of HBV structure by Chen *et al.* [182] indicates that the CTD chains are exposed outside the capsid shell for empty capsids, while they are confined inside the capsid for RNA containing-NCs. By analyzing CTD interaction with the surrounding molecules in the solution, Chen *et al* provided additional information for the specific sequence of exposed CTD segments. Motivated by these *in vitro* results for the HBV structure, we have examined the density profiles for the individual segments of the CTD tails.

The DFT allows us to determine the distributions of individual residues of the CTDs. The distribution data may be analyzed in several ways. Here we first consider the accumulated number of linker-CTD residues as a function of the radial distance r from the capsid center. The sum of all residues at each radial distance is normalized by the

total number of amino acid residues, *viz*., 240. Except for the first tail segment, 42 of linker-CTD segments are traced, thus the accumulated number approaches from 0 to 42 as r increases. For empty capsid case (Fig. 3(a)), we can see that the accumulated number of CTD residues is about 29 at inner capsid wall (r =13 nm) and 30 at outer surface (r = 15 nm). Accordingly, except those 30 of inside segments, 12 of rest residues are exposed outside the empty capsid. However, for RNA-containing NC (Fig. 3(b)), the counted number is almost 40 at outer surface (r = 15 nm), indicating only rest of 2 CTD residues are distributed outside ($r > 15nm$). At $r \sim 19nm$ from center of the NC the counted CTD segments number reaches to final value of 42. On the other hand, for empty capsid (Fig. 3(a)), CTD segments reside up to further distance ($r \sim 22nm$) from capsid center. These results suggest how surface characters are different between the empty and NC due to the extruded CTD residues. We have also considered the distribution data in terms of charged or neutral amino-acid groups but find not much variation between the empty and RNA-containing capsids.

Figure 4-3 compares the exposed ratio of individual CTD segments. A segment is '100% of exposure' if it is fully distributed in the outer region of the capsid, *i.e.,* $r \geq 17nm$. We assume that the residues in the outer region are able to fully interact with surrounding molecules like cellular enzymes. For convenience, we rand each residue according to its position at the CTD chain, *i.e*., 1 represents the amino acid tethered at the inner capsid surface, and 34 is the C-terminal residue. As expected, the empty capsid shows higher overall CTD exposure compared to the NC. Assuming that the RNA content is mainly responsible for the deviation between the empty capsid and the NC, we

find that the attraction between CTDs and RNA internalizes more than half of exposed CTD segments in the empty capsid. In the latter case, the exposed ratio peaks around 21-26th residues (^{170}Ser-^{171}Pro-^{172}Arg-^{173}Arg-^{174}Arg-^{175}Arg,), which overlap with a typical binding motif of serine/arginine protein kinase (SRPK)[197, 198].

Schematically Figure 4-4 depicts the distribution of CTD residues for the empty capsid according to our DFT calculations. Because the end residues have lower exposure rate than the 21-26th residues, a CTD tail must somehow bend in the middle, making it to have a hook like shape. The bended structure increases the possibility of the motif to contact with the SRPK. According to our DFT predictions, ~ 30% present of the residues in such motif are within the reach of the SRPK kinase.

The *in vitro* study by Chen *et al.* indicates the transient CTD location in HBV capsids [182]. It revealed the association of SRPK on the outer surface of the empty capsid, mediated by the enzyme-binding motif in the CTD. The same assay was conducted for the NC containing RNA genome but in that case, the enzyme binding was mostly inhibited. The SRPK binding demonstrated unequivocally that the surface characteristics of the capsid change with the CTD location. Corresponding to the experimental approach, the DFT results capture the transient feature of the capsid surface, rendering additional evidence on the interaction of CTD with SRPK.

The effect of phosphorylation

The CTD distribution is sensitive to the phosphorylation of serine (S) residues[24]. In wild-type capsids, three of the serine residues (S155, S162, S170) have been recognized as phosphorylated upon the capsid formation and packaging with pgRNA[7, 11].

In our DFT calculations, phosphorylation can be studied simply by setting the valence of those three serine residues from 0 to -1 for the phosphorylated case. The addition of negative charges affects both the RNA distribution and the exposure of CTD chains.

Figure 4-5 shows that the RNA structure inside the capsid varies significantly in respond to CTD phosphorylation. Compared to the unphosphorylated case, the RNA segments become more uniformly distributed and are positioned closer to the capsid surface. CTD phosphorylation makes the RNA distribution transduced to have relatively higher peak near the capsid surface. Phosphorylation reduces the extend CTD exposure outside the capsid. Because the addition of negative charge reduces intra-chain electrostatic repulsion, the CTD brush is slightly collapsed in comparison to the unphosphorylated brush.

Discussion

Implication of the CTD exposure

Several recent investigations presented the transient exposure of CTDs to the capsid surface[19, 21, 28, 191, 199, 200]. It has been postulated that the exposed CTDs regulate the post-translational process of HBV core, *i.e.,* the trafficking into nucleus and the enveloped secretion. Kann *et al.* determined the fraction of the exposed CTDs for NC in different maturation stages, and suggested the CTD-associated signal modulates the capsid delivery into cellular nucleus[201]. Ning *et al.* observed that the secreted HBV particles contained either empty capsids or NC with double stranded DNA, while the immature NCs, *i.e.,* those filled with pgRNA or single-stranded DNA were excluded from secretion[185]. Accordingly, a hypothesis was set such that the immature NCs

negatively regulate the trafficking process[24, 182, 185]. Zlotnick's group compared the structural characteristics of empty and RNA-NC, and suggested that the strong interaction of CTDs with RNA genome obstructs the CTD exposure.

Our theoretical model supports the mechanism of genome-regulated exposure of CTDs. Although we are not describing the whole process of HBV replication, a substantial structural change of CTD implies its functional correlation with the maturation signaling. Our model predicted that about ten residues for each CTD tail (34 residues) tend to be exposed outside the capsid when the tails are free from the genome contents. Thus, $\sim 30\%$ of CTD segments additionally extruding outside would modify the capsid surface characteristics, which trigger the cellular trafficking. For such the empty capsid, the CTD tails have been suggested to be outward into far space from the capsid center, so that the outermost reachable r to be $\sim 19nm$ for RNA-NC but $\sim 22nm$ for the empty capsid. Such a structural deviation between empty and RNA-filled capsid supports the hypothesis that the degree of CTD exposure may trigger selective selection upon the post-translational process [24, 182, 185]. The hypothesis, specifically, the rational on the transient CTD structure was also endorsed by experiments [24, 182]. In supporting those observations, our model gives the evidence on the CTD exposure and accessibility into outer capsid space.

Structural changes associated with CTD phosphorylation

It was postulated that HBV carries serine residues in different phosphorylation states during the process of the capsid assembly and reverse transcription of the genome. Specifically, in a duck hepatitis B virus (DHBV), the capsid proteins (CP) were in

phosphorylated form upon the capsid assembly. However, they were dephosphorylated for the mature NC filled with rcDNA [13, 202]. The CTD in phosphorylation state is believed to control the pgRNA encapsidation as well as the reverse transcription. Mutation of phosphorylation sites on CTD to mimic unphosphorylation hindered RNA packaging[15-17]. Similar mutation studies showed CTD phosphorylation affects DNA synthesis[12, 16, 17]. We have investigated the structural effect of phosphorylation. DFT calculation shows that CTD phosphorylation induces RNA segments to be more localized toward inner capsid surface (Fig. 4-5). Such a RNA structure alteration was observed by Wang *et al.* [24]. In the phosphorylation-mimic case, the pgRNA inside the capsid presented more ordered structure than that in *E-coli* derived nucleocapsid. Our theoretical analysis indicates that such a structural ordering of RNA is correlated with the CTD location. In phosphorylated state, more of CTDs are localized at region of 7 ~ 12 nm from capsid center, and the inner shell (< 7 nm) distribution of the phosphorylated CTDs is relatively depleted. It is expected that the RNA-CTD interaction would be reduced because of added negative charges to the CTD by the phosphorylation. Figure 4-5 shows such retarded complex formation between RNA and CTDs. At inside region (r < 7 nm), density profile of RNA for the phosphorylated case is higher than that for unphosphorylated one. However, corresponding densities of CTD segments for each case are inversed at the region, thus the unphosphorylated CTD chains show higher segmental density than phosphorylated CTD chains. In other words, CTD chains stay relatively apart from the RNA when they gain additional negative charges by the phosphorylation. Accordingly, phosphorylation results in higher RNA density close to inner surface of the capsid and it maintains

monotonic radial distribution except near the inner wall. By contrast, RNA in the unphosphorylated case shows more inhomogeneous distribution.

Exposure of SRPK binding motif in the CTD

Although the primary kinase responsible for HBV phosphorylation is still under active investigation, SRPK has been a well-speculated candidate [181, 203]. SRPK is able to fulfill the *in vitro* phosphorylation of HBV CPs[181]. Moreover, the binding of this enzyme to every 2-fold pores of HBV empty capsid has been reported [182].

Our DFT calculation provides the density distributions of individual CTD segments in HBV capsids. We find that a specific motif, 21th -26th CTD residues of the CTD, is located in the outer region ($r \geq 17nm$) with the highest probability. Such a serine/arginine repeats domain serves as an ideal substrate for SRPK binding and nuclear localization signals (NLS) [204, 205]. At the physiological condition ($C_s = 0.14M$), the exposure fraction of SRPK binding motif is expected to be $\sim 30\%$ of that corresponding to the empty capsid. A recent *in vitro* HBV study demonstrated the empty capsid was decorated by SRPKs at all of 2-fold capsid pores. However, RNA-NC failed in that binding assay indicating that CTDs are not substantially exposed to be accessible for outside SRPKs [182]. Regarding the exposure of kinase binding motif, the deviation between the empty capsid and the NC has been confirmed by our DFT calculation. In the presence of the RNA, the exposed fraction of SRPK binding motif of 21- 26th CTD segments is reduced to ~5%.

Chen *et al.* claimed that no interaction between RNA-NCs and SRPKs were observed in conducting a chromatography of capsids sample into the SRPK bound

column[182]. However, our theoretical model predicts that CTD chains are also partially exposed in RNA-containing NCs but in relatively much smaller portion. One explanation for this discrepancy is that the enzyme binding requires sufficient substrate contact as for the case of the empty capsids. For RNA-containing NCs, the target motif exposure is insufficient to retain capsids bound on the column. Also in experiment, the RNA-NCs were obtained from *E. coli* system, filled with the host RNA. Although the bacterial RNA contents have been regarded to mimic the authentic pgRNA regarding its amount[6], but their characteristics can be different[206]. Accordingly, we assume that the trapped bacterial RNA induces relatively stronger interaction with CTDs than pgRNA does. The calculation shows that if the size of RNA in the NC model is set to have larger value than that of the authentic pgRNA (3.5 kb), the CTD exposure ratio decreases (data not shown). It was observed by Rabe et al. that the structure of *E-coli*-derived capsids and authentic immature capsids were clearly different[22]. In their analysis, RNA-NCs were obtained *in vivo* with inhibiting the reverse-transcription to maintain the pgRNA. Certain portion of CTDs was exposed outside capsid, so that cleaved by outside digesting enzyme. However, NCs obtained by expressing the capsid protein in *E. coli* were not affected by outside enzyme, suggesting their CTDs were not exposed.

Even for *E-coli*-derived nucleocapsid with RNA, it was speculated that certain portion of CTDs are exposed. Vanlandschoot *el al.* observed the CTD-mediated attachment of nucleocapsids to glycosaminoglycans expressed on the plasma membranes of cells[207]. A recent structural analysis showed that a small portion of CTDs are stretched out through NC pores[184]. These results lead to hypothesize that RNA-filled NCs also

carry the exposed CTDs but its fraction is substantially small compared to the empty capsid. DFT calculation supports that hypothesis with quantified comparisons. The distinct structures of capsids suggest that the CTD location, specifically, the distribution of the kinase target motifs can be modulated in response to the inside genome. Thus, the phosphorylation process should be entangled with the genome contents by the CTD involvement.

Conclusions

To summarize, we have used a coarse-grained model to describe the radial distribution of the flexible domains of HBV capsids and encapsidated RNA segments. The molecular model captures the molecular excluded volume effects and electrostatic interactions among the key viral components. The predicted profiles of the CTD location correspond with experiments supporting the postulated CTD exposure outside capsids. The DFT predictions validate the role of CTD in signaling viral maturation. Furthermore, the theoretical model predicts the probable position of a SRPK binding motif in the CTD in good agreement with that determined by a recent experiment. In addition, we provided theoretical description on the reorganization of RNA and CTD distributions in response to phosphorylation.

Figure 4-1. The radial distributions of polymeric segments in empty and RNA-containing HBV capsids. The density profiles of salt ions, segments of linker-CTDs and RNA chains are shown for the empty capsid (A) and the nucleocapsid (B). In this and other figures, the perpendicular dotted lines indicate the inner and outer positions of the capsid surface.

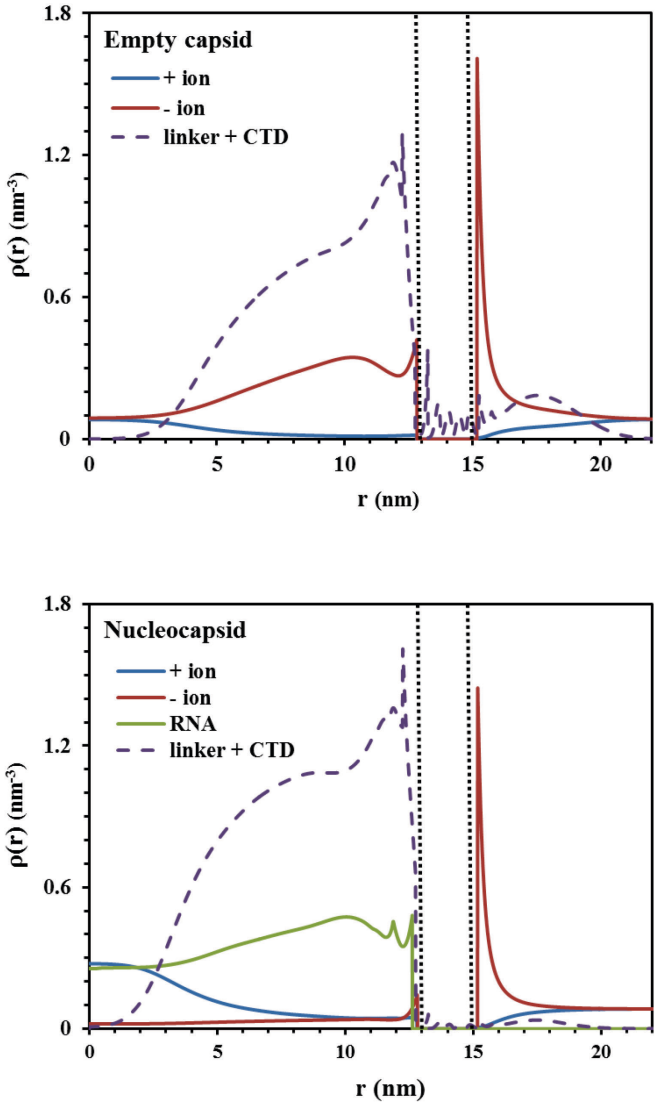


Figure 4-2. Schematic representation of HBV capsids. The structure of empty capsid and nucleocapsid with RNA (NC) are compared by schematic representation of spherical capsid shell (dashed lines) and coarse grained RNA, C-terminal domains (CTD) and salt ions. Relatively large portion of CTD tails on the empty capsid are exposed outside capsid, while RNA molecules in the NC associate with CTDs making a thin layered complex shell.

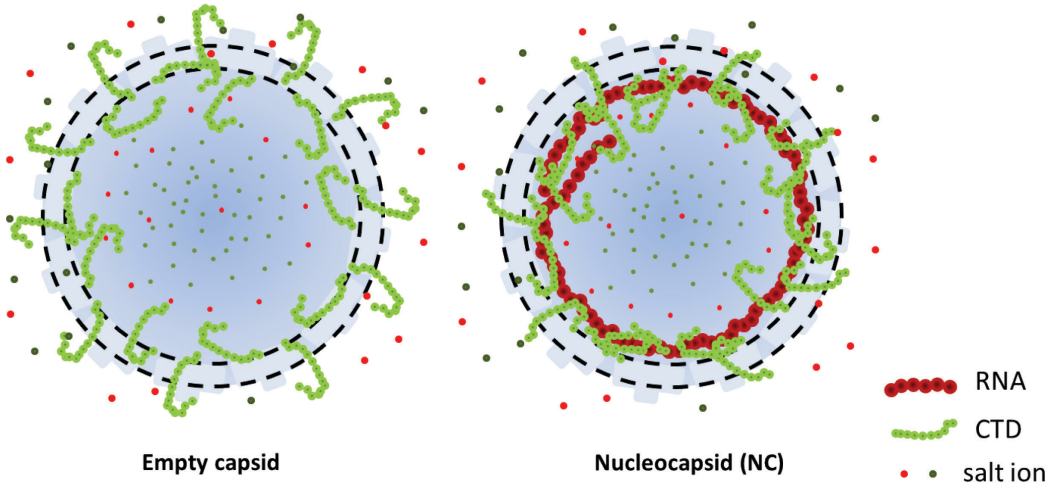


Figure 4-3. The accumulated number of CTD residues for the empty (A) and RNA-containing (B) HBV capsids. The accumulated number is radially integrated from the density profiles shown in Figure 4-1 for the linker-CTD residues and normalized by the number of CTD tails (N_t = 240). Thus y-axis value indicates summed number of residues (per each tail) up to the radial distance from capsid center (r). As r increases, the sum value approaches to 42, which is number of traced residues on each tail. The accumulated values for charged (+ and -), neutral (·) residue groups are shown separately.

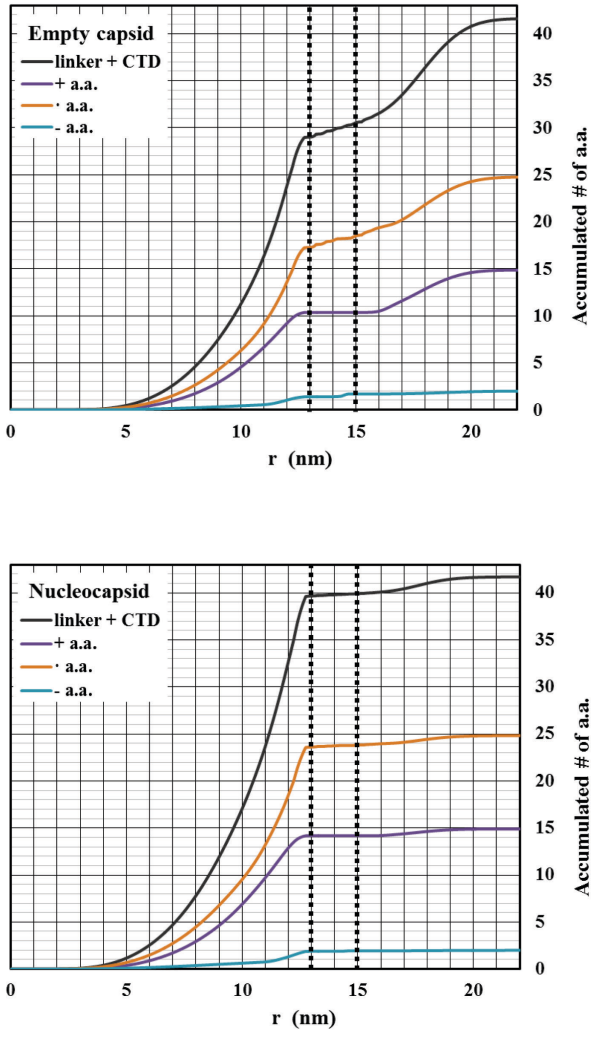


Figure 4-4. The ratio of the exposed CTD residues exposed to the outer region of HBV capsids. The CTD residues exposed in the region $r \geq 17nm$ are counted for both the empty capsid and nucleocapsid. Each residue is ranked from the tethering site (1) to the end segment (34), and the panel above denotes amino-acid residues.

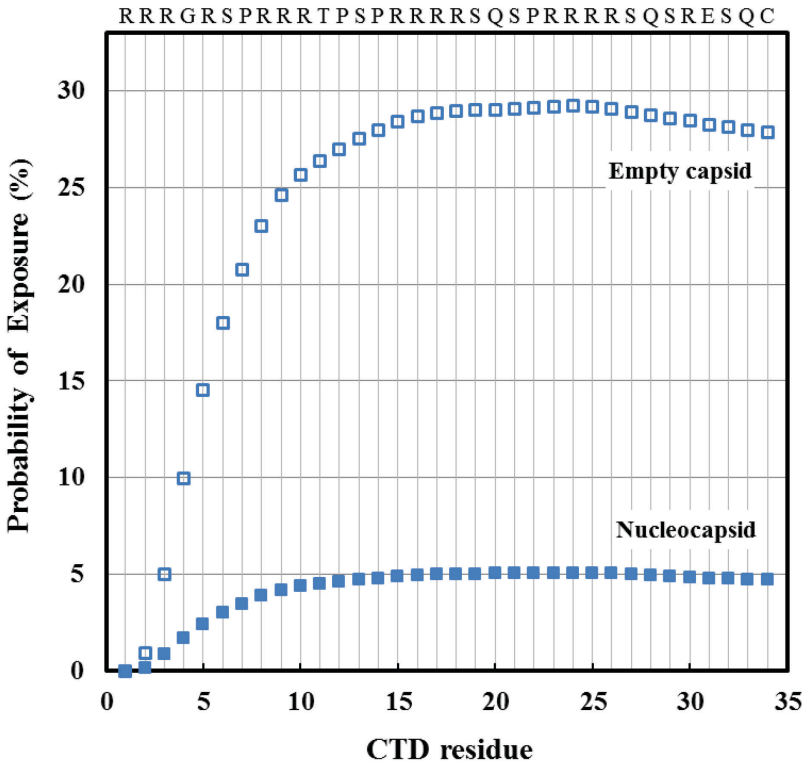


Figure 4-5. Schematic representation of the CTD location. (A) a 2-fold capsid pore and a dimer of the capsid core protein (CP). (B) 6 CTD tails in each 2-fold pore. (C) CTD tails of the empty capsid. Here the tails are distributed in both inside and outside the empty capsid through the 2-fold hole. Red segments indicate the SRPK binding motif.

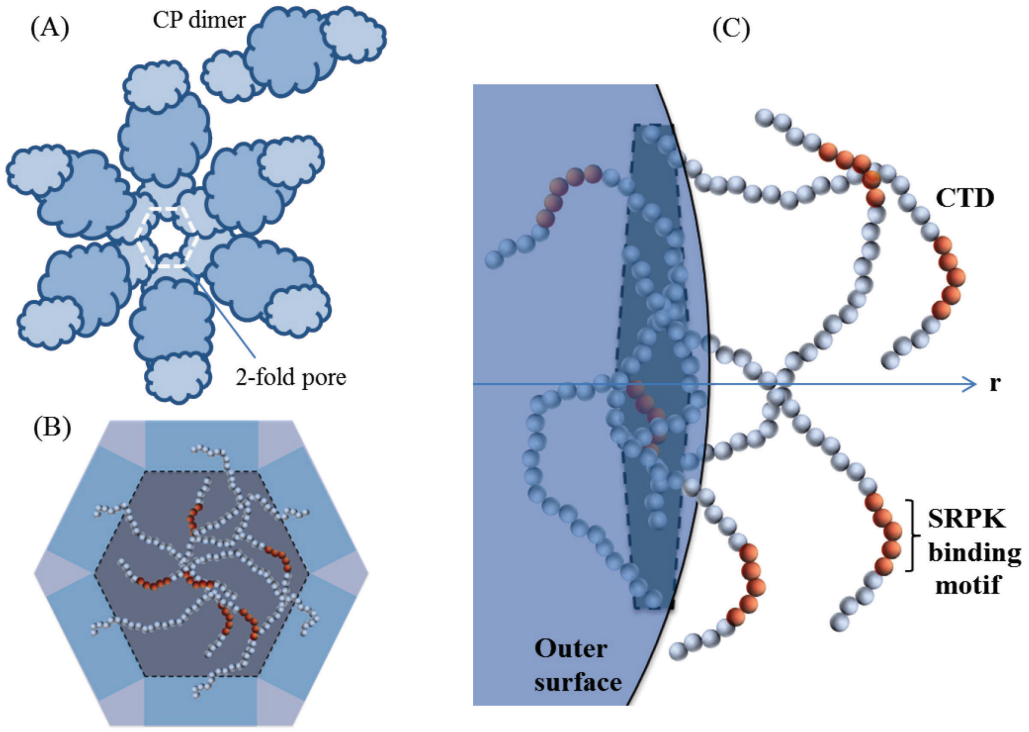
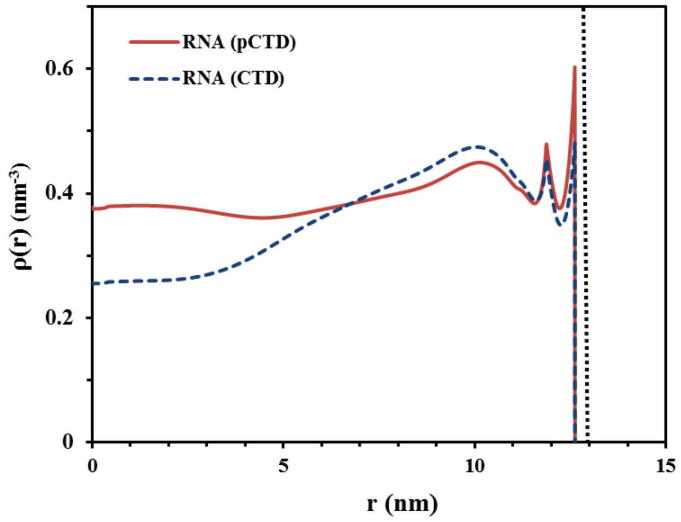
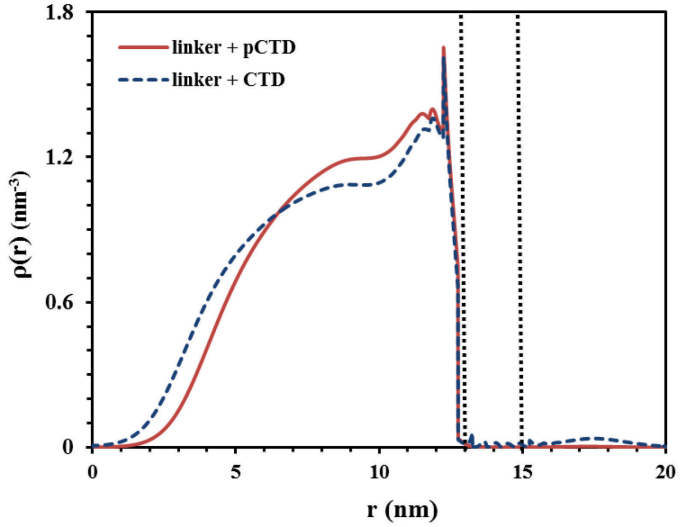


Figure 4-6. Effect of phosphorylation on the distributions of RNA and linker-CTD tails. (A) RNA distribution in phosphorylated (pCTD) and unphosphorylated (CTD) capsids. (B) The distributions of phosphorylated (pCTD) and unphosphorylated CTD tails.





Chapter 5. A thermodynamic model for genome packaging in hepatitis B virus

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 therapeutic applications. While the structure of encapsidated genomic materials has been routinely characterized with experimental techniques such as cryo-electron microscopy (cryo-EM) and X-ray diffraction, little 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 (pg) RNA encapsidation in human Hepatitis B virus (HBV) using a coarse-grained molecular model that captures the essential components of non-specific 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 the HBV nucleocapsid changing with the CTD length in excellent agreement with recent 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 studies. Our thermodynamic analysis of the nucleocapsid structure and stability indicates that about 10 % of the CTD residues are free from complexation with RNA, resulting in partially

exposed CTD tails that may play an important role in signaling HBV viral maturation. The thermodynamic model also predicts the free energy of complex formation between macromolecules, which corroborates experimental results for the impact of CTD truncation on 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 more than 350 million people alive today are chronically infected by HBV, current treatments are effective mostly in short terms and often plagued by adverse reactions such as drug resistance[3]. The life cycle of HBV replication starts with a single-stranded pre-genomic (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 DNA and then a partially double-stranded DNA in a matured capsid. It has been well 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 (a.a.) residues, which can be divided into an N-terminal assembly domain (140 a.a.), a nonapeptide linker (9 a.a.), and the CTD (34 a.a.) tails. It is known that CTD is rich in arginine residues and is indispensable for proper pgRNA packaging and DNA polymerization[6, 16, 208].

Previous mutagenesis studies indicate that pgRNA packaging is affected by the net charge of the CTD tails[18, 209]. A partial truncation of the flexible, highly charged domain reduces the size of packaged RNA, suggesting a charge balance for association

among macromolecules[209]. The regulative role of nonspecific electrostatic interactions has been indicated for many single-stranded viruses[210-218]. A conserved ratio (N/Q=1.61) of the genome size (N) to the overall charge of capsid proteins (Q) was proposed based on experimental results for over 30 wild-type viruses, and the change balance hypothesis has be partially justified with theoretical analysis[219]. However, the universal charge ratio is not applicable to HBV. The conserved N/Q value would predict the RNA size of ~4500 bases, much too large compared with the size of pgRNA (3400 bases) in wild-type HBV. An alternative model for predicting the optimal genome size has been proposed to account for not only electrostatic interactions but also multi-body correlation effects, such as particle size, chain connectivity, and intermolecular correlations[186]. It has been shown that genome packaging is influenced by the capsid properties including capsid shell charge. For HBV, genome packaging is also related to the partial exposure of the CTD tails[25-27]. From a biological perspective, CTD exposure changes the surface properties of the viral particle important for posttranslational processes such as nuclear entry, capsid envelopment and virion secretion [22, 29, 220-222].

Developing new therapeutic antiviral strategies for HBV can be helped with a good understanding of the molecular details of viral replication. One of the critical questions about the HBV replication process is how the RNA/DNA chains are associated with the flexible domains of the capsid proteins and how the macromolecular association stabilizes the nucleocapsid. It has been shown that, for many single-stranded (ss) DNA/RNA viruses, genome packaging is primarily driven by non-specific electrostatic interactions between the nucleic acids and oppositely charged, flexible domains of the

capsid proteins[219, 223]. Previous studies also suggest that the viral packaging conforms a fixed charge ratio between the nucleic acids and the capsid proteins[219, 223-226]. However, the charge balance model was challenged by more recent theoretical works indicating that genome packaging involves other mechanisms leading to nonlinearity in the macromolecular charges[227]. The charge balance hypothesis ignores the Donnan equilibrium[228] and specific effects of the genome structure[229].

Previously, we described the thermodynamic basis of RNA encapsidation using an ion-explicit coarse-grained model for HBV capsids[186]. The theoretical model accounts for important factors in regulating genome packaging including electrostatic interactions, molecular excluded-volume contributions, and correlation effects. It predicts the optimal size of packaged genome in fair agreement with existing experimental data for both wild type (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 to previous experimental references, we validate the thermodynamic model with more extensive *in vitro* data for partial RNA packaging in various HBV mutants that become available only recently. Although the universal ratio (N/Q) of the genome size to the charge of flexible tails is not applicable to HBV[219], both theoretical and experimental results indicate near linear dependency of the RNA size to the net charge of the CTD tails. The discrepancies among previous studies may be reconciled by accounting for by 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 quantify the relative stability of nucleocapsids at different levels of RNA encapsidation.

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 is not able to capture the atomic details or the secondary structure of pgRNA that may play a key role in the kinetics of viral genome encapsidation[229, 230], it keeps essential ingredients of non-specific interactions among macromolecules underlying nucleocapsid formation. From *in vitro* assembly of HBV nucleocapsids in *E. Coli*, it has been noted the RNA packaging is insensitive to the origin of the genome or the RNA sequence[16, 209]. The nonspecific nature of RNA encapsidation suggests that the essential features of HBV capsids can be faithfully described with a coarse-grained molecular model[112, 186, 231].

As in our previous work[186], 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 σ_R= 0.75 nm[187] and σ_C= 0.5 nm[232] for nucleic acids and amino-acid residues, respectively. We assume that each RNA segment bears a negative unit charge. At physiological conditions, a wild-type (WT) CTD tail contains 16 positively charged segments (all from arginine residues), 4 negative residues (1 glutamate and 3 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 before that the coarse-grained model is able to capture the essential features of macromolecular interactions[187, 233-236].

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

$$\exp[-\beta V_{RNA/CTD}(\mathbf{R})] = \prod_{i=1}^{M-1} \frac{\delta(|r_{i+1} - r_i| - \sigma_{R/C})}{4\pi\sigma_{R/C}^2}, \tag{5.1}$$

where $\beta = 1/(k_B T)$, T is the absolute temperature, and k_B is the Boltzmann constant; $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_M)$ represents the configuration of a coarse-grained RNA/CTD chain with the positions of M spherical segments specified by $\{\mathbf{r}_{i=1,2,\cdots,M}\}$. 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 non-bonded interactions among the RNA/CTD segments using the pairwise additive potential $u(r)$:

$$\beta u_{ij}(r) = \begin{cases} \infty, & r < \sigma_{ij} \\ \dfrac{Z_i Z_j l_B}{r}, & r \geq \sigma_{ij} \end{cases}, \tag{5.2}$$

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 σ_{Na^+} = 0.39 nm and σ_{Cl^-} = 0.36 nm, respectively[237].

As in our previous publications[112, 231], 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 semi-permeable 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

$$\phi_{RNA}(r) = \begin{cases} \infty, & r \geq R_C - 0.5\sigma_R \\ 0, & \text{otherwise} \end{cases}. \tag{5.3}$$

Each T4 HBV capsid is composed of 240 capsid proteins. Accordingly, 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 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:

$$\beta\phi_{tail}(r) = \begin{cases} \ln(0.1), & R_C - 0.5\sigma_t < r < R_C + d + 0.5\sigma_t \\ 0, & \text{otherwise} \end{cases}. \tag{5.4}$$

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

$$\Psi_i^C(r) = Z_i e\varphi_s(r), \tag{5.5}$$

where φ_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 $Q_c = 0.7e/nm^2$, assuming a uniform charge distribution[63]. The surface charge density is used to calculate the electrostatic capsid external potential Ψ^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. Figure 5-1 depicts schematically a packaged chain at the early stage of assembly. The RNA chain exposed will be eventually degraded by the cellular nuclease[209, 238], and the inside portion defines the optimal size of the genome.

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

$$F = F_{cap} + F_{out}, \tag{5.6}$$

where F_{cap} accounts for the free energy of the entire nucleocapsids, and F_{out} represents the energy of the exposed RNA in the local cellular environment outside capsid. At equilibrium, the optimal genome size in a stable nucleocapsid can be found by minimizing the overall potential

$$\frac{\partial F}{\partial N_{RNA}} = 0, \tag{5.7}$$

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

We may estimate the free energy of RNA outside capsid from the reversible work to insert the unpackaged RNA segments into the cellular environment. The free energy of

insertion is approximated by the volumetric work to create a cavity in a uniform electrolyte solution with the osmotic pressure of a cellular milieu ($\Pi \approx 5$ bar[239, 240])

$$F_{out} = \Pi v_R (N_{pgRNA} - N_{RNA}), \tag{5.8}$$

where v_R stands for the excluded volume of each RNA segment, and N_{pgRNA} 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_{RNA}, its derivative with respect to N_{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[186].

The thermodynamic potential within the nucleocapsid accounts for the genome-capsid interactions with the participation of salt ions. The corresponding semi-grand potential (F_{cap}) is estimated from the molecular composition of RNA and CTD tails, the chemical potential of small ions, and interaction potential between those components and the capsid shell:

$$\begin{aligned} F_{cap} = & \sum_{\alpha=\pm ions} \int d\mathbf{r} \rho_\alpha(\mathbf{r}) \left[\ln \rho_\alpha(\mathbf{r}) - 1 \right] \\ & + \sum_{\alpha=\pm ions} \int d\mathbf{r} \rho_\alpha(\mathbf{r}) \left[\Psi_\alpha(\mathbf{r}) - \mu_\alpha \right] \\ & + \sum_{M=R,T} \int d\mathbf{R} \rho_M(\mathbf{R}) \Psi_M(\mathbf{R}) + F^{ex} \end{aligned}, \tag{5.9}$$

where subscripts α 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; μ_α 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 potential due to intermolecular interactions, which 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[241, 242] and its successful applications to HBV nucleocapsids[112, 186, 231].

Results and discussion

The optimal genome size of HBV nucleocapsid

A newly formed WT HBV nucleocapsid contains pgRNA, which consists of about 3400 nucleotides. It has been shown that a similar amount of bacterial ssRNA can be encapsidated by expression of the full-length capsid proteins in *E. coli*.[10] When the CTD tails are truncated, however, the capsid contains a smaller RNA chain. Figure 5-2 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 (a.a.) 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.

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 encapsidated polymerase protein. For WT and mutants with low degree CTD truncations, nevertheless, the theoretical predictions are excellent. By contrast, the calculated RNA

size for mutants with 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.

The charge balance between packaged RNA and CTD tails

It has been speculated that, for single-stranded (ss) 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 packaged genome[18, 206, 209]. 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 the charge balance between those macromolecules[219]. However, previous investigations suggest different values for the charge ratio between the nucleic acids and capsid proteins[228].

Figure 5-3 presents the packaged RNA size (N_{RNA}) as a function of the net charge of CTD tails (Q_{tail}) of the HBV core proteins. Both the DFT results and experiment data for the optimal genome size (shown in Figure 5-2) are assorted according to Q_{tail} for WT and mutants. Although there is an overall trend that N_{RNA} increases with Q_{tail}, both experimental data and theoretical predictions show that the charge ratio (N_{RNA}/Q_{tail}) is not constant. For example, two mutants have same Q_{tail} but different N_{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 of viral particles. Hu *et al.* argued that the optimal RNA size is dependent on the contour length of peptide tails, and

obtained the charge ratio of N/Q=2.[223] Different values of fixed charge ratio (<1, 1 and 2) also have been predicted with an assumption of a uniform capsid charge placed at inner capsid surface[224-226]. Most notably, Belyi and Muthukumar determined N/Q ~ 1.6 based on the counterion-condensation theory for RNA and peptide tails[219]. The linear relation is compared in Figure 5-3 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 recent experiment also revealed the lack of universality for the charge ratio on RNA packaging in Brome Mosaic Virus (BMV)[227]. 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[186], the excluded volume and correlation effects of macromolecules and salt ions contribute in 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.

The stability of nucleocapsids

Earlier experimental work suggests that encapsidation of pgRNA stabilizes HBV nucleocapsids in compared to empty capsids[10]. Besides, the nucleocapsid stability changes during the reverse transcription from RNA to DNA, indicating its potential correlation with viral maturation[243]. Moreover, mutagenesis experiments reveal that partial truncation of CTD influences the RNA packaging and DNA synthesis. For

example, Le Pogam *et al.* 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 additional deletion of CTD residues critically impeded the DNA production[209]. Nucleocapsids of those mutants (more of 10 C-termini residues are truncated) 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 the capsid structure formation by protein subunits (F_{em}) and that due to the interactions among the flexible domains of macromolecules (F_{plex})[244]:

$$F_{NC} = F_{em} + F_{plex}. \tag{5.10}$$

In Eq.(5.10), the first term on the right side represents the energy change from diluted core protein subunits to assembled empty capsid, and the other is the complexation energy of RNA/DNA genome and CTD segments, which is determined from the free energy difference between confined macromolecules complex and isolated genome and CTDs chains at infinite dilution. Previously, this complexation free energy has been approximated by accounting for changes in the chain elastic energy, excluded volume effects, electrostatic interactions, and entropy of mixing[244]. Among HBV mutants, the contribution of empty capsid structure is same for all cases, and thus we only compare the free energy of chain complexation (F_{plex}) for mutants with different length of CTD tails.

Figure 5-4 shows the complex formation free energy versus the CTD chain length for the HBV mutations considered above. Here a negative value means that the overall contribution by the encapsidation of macromolecules is energetically favorable. As expected, the F_{plex} is negative for WT nucleocapsids (CP183) and others with a small degree truncation (CP179 and CP173), and thus the corresponding nucleocapsids are more stable than the empty capsid without macromolecules. However, the complexation free energy is positive for those with a large degree truncation. We note a sigmoidal change of the complexation energy with regard to size of CTD tails as shown by the solid line in this plot. The exponential transition occurs around CP173 to CP171, where F_{plex} changes from a negative to a positive value.

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

Structure of HBV nucleocapsids

Whereas the cryo-EM pictures are available for both WT and mutant HBV nucleocapsids[186], the molecular structures for the flexible components of the virus remain poorly understood. With CTD tails fully confined, our previous DFT calculation

predicted a thin layer of RNA chain and a brush-like distribution of the CTD segments. Here we conduct similar structural analyses by expanding the computation range from inside 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 semi-permeable characters are relevant to build the HBV capsid model[231].

Figure 5-5 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[186], the 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 brush-like structure. Noticeably, some CTD segments approach outside capsid region for both WT and CP179 mutant. The last panel of Figure 5-5 compares the average number of exposed CTD residues for different mutants. For WT nucleocapsids, 3-4 CTD residues per capsid protein are supposed in the outside region. Because these residues do not participate in the complex formation with encapsidated RNA, we expect about 10% of CTD segments are not essential for the genome packaging.

Our theoretical results for the multiregional distributions of CTD segments inside and outside capsid may help to explain how the CTD tails fulfill multiple roles the HBV life cycle. As discussed above, CTDs are indispensable for genome encapsidation and replication. Other roles have also been identified such as mediating the nuclear localization and maturating signal [22, 29, 220-222]. Our model suggests the WT capsids are

optimized to fulfill pgRNA packaging while maintaining a portion of exposed CTD segments to conduct a potent signaling role. Mutant CP179 has less than 2 residues per CTD tail exposed, and such exposure is not observed for mutants with shorten CTDs. Consequently, those mutants may lose the capability to deliver the maturation related 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 sizes 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 wild type (WT) and mutant HBV capsids with truncated C-termini of the core proteins. The thermodynamic model confirms that the 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.

The thermodynamic stability of nucleocapsid is affected not only by the overall charge balance between oppositely charged macromolecules, but also by the molecular excluded volume effects and charge distribution in 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. We have also investigated the spatial distributions of RNA and CTD segments in both WT and mutant capsids. 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 prefer to stay outside the HBV capsid. The multi-regional distribution may help to rationalize the multifunctional roles of the CTD tails in assembly, maturation, and envelopment of HBV capsids.

Figure 5-1. (A) Schematic of key components in a HBV nucleocapsid. Optimal genome size can be determined from partial encapsidation of a RNA chain that minimizes the thermodynamic potential of the nucleocapsid in an aqueous electrolyte solution. The unpackaged RNA fragments are subjected to degradation by cellular nuclease.

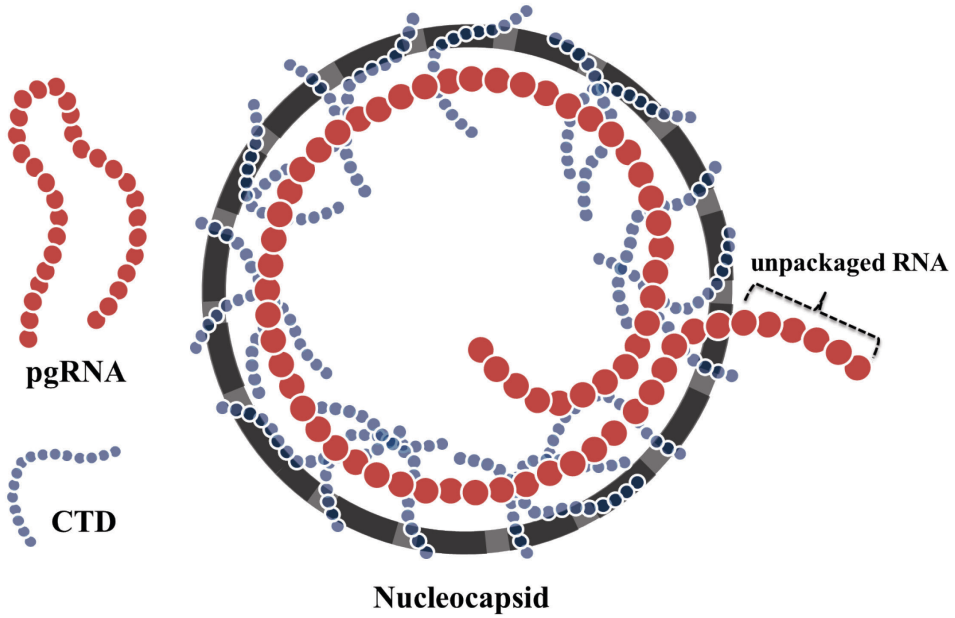


Figure 5-1. (B) The sequences of amino-acid residues in the C-terminal domains (CTD) of the wild type (WT) and mutant capsid proteins (from residue 150 to the end). In this and other figures, the WT and mutants of the core protein are distinguished with the number of residues (e.g., 183 179, 173, 171 and 164). The electrostatic status of each CTD residue is labeled with different colors (red, black and blue for positive (+), neutral and negative (-), respectively).

+++ +- +++ - ++++ - ++++ +-	
R R R G R S P R R R T P S P R R R R S Q S P R R R R S Q S R E S Q C	183 (WT)
R R R G R S P R R R T P S P R R R R S Q S P R R R R S Q S R	179
R R R G R S P R R R T P S P R R R R S Q S P R R	173
R R R G R S P R R R T P S P R R R R S Q S P	171
R R R G R S P R R R T P S P R	164

Figure 5-2. Theoretical predictions (black open square) and experimental data (Exp. 1[209], red circle; Exp. 2[245], blue triangle) for the optimal RNA length (kb or kilo-base) in WT and mutant HBV nucleocapsids.

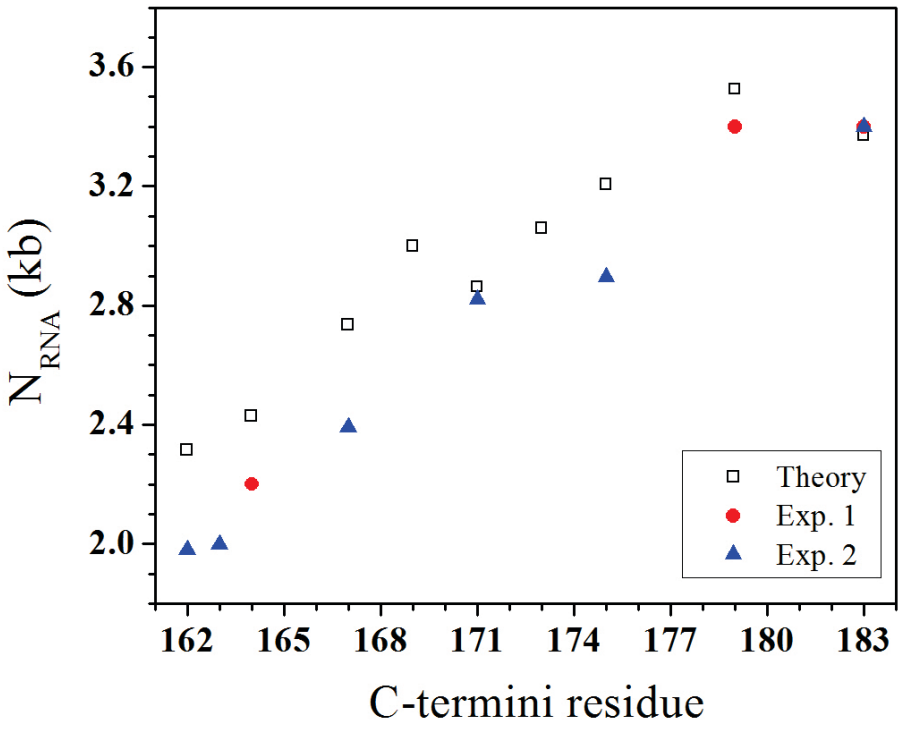


Figure 5-3. Correlation between the packaged RNA size (N_{RNA}) and the net charge of C-terminal tails (Q_{tail}). The solid line (N/Q = 1.61) was proposed by Belyi and Muthukumar [219].

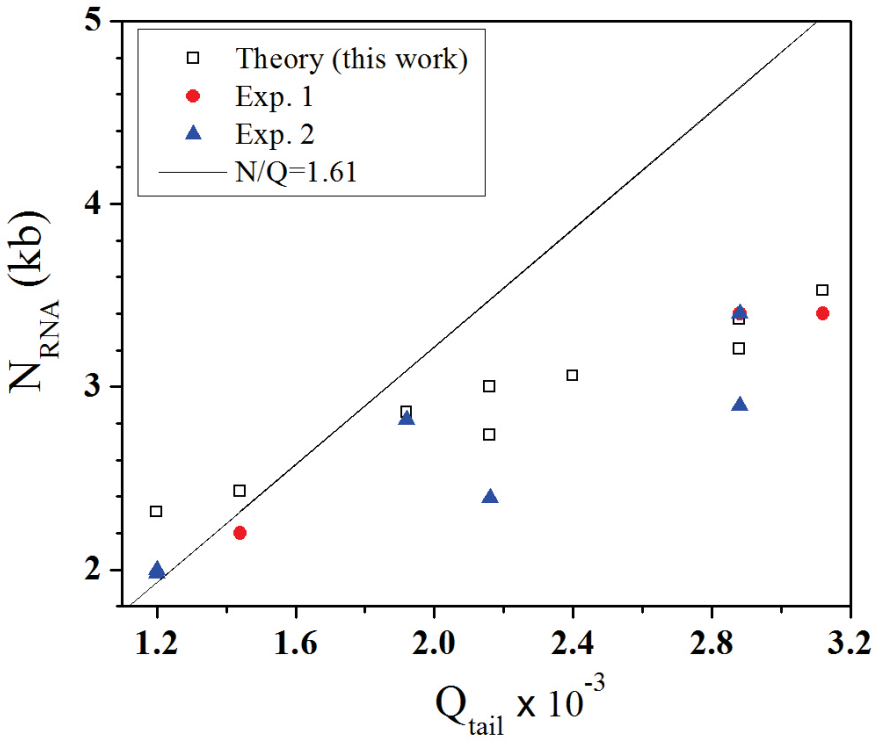


Figure 5-4. The free energy of RNA encapsidation depends on the CTD length. The complexation free energy (F_{plex}) of RNA and CTD chains is determined for WT and mutants. The sigmoidal line is for the guidance of the eye.

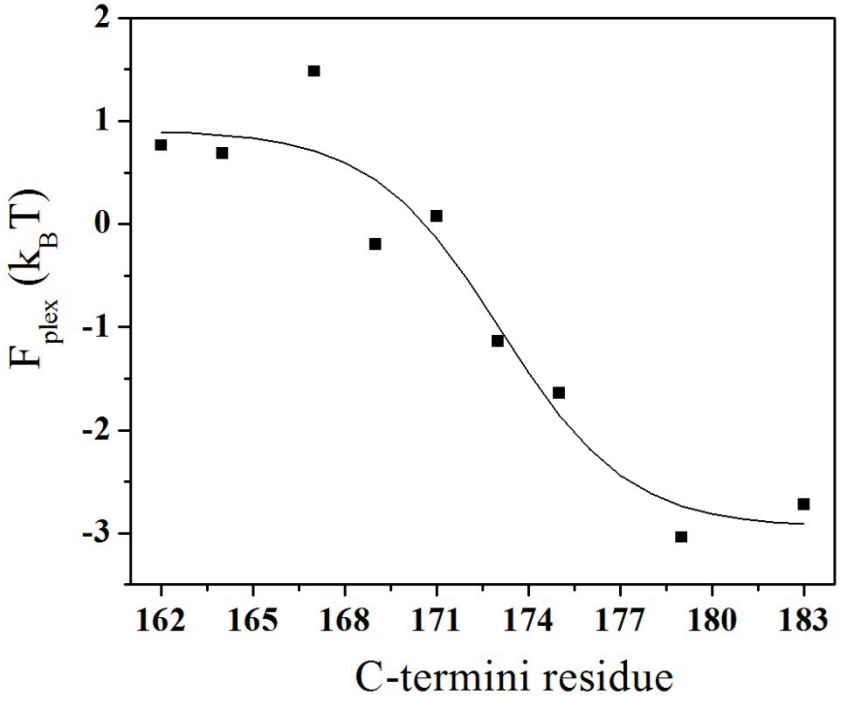
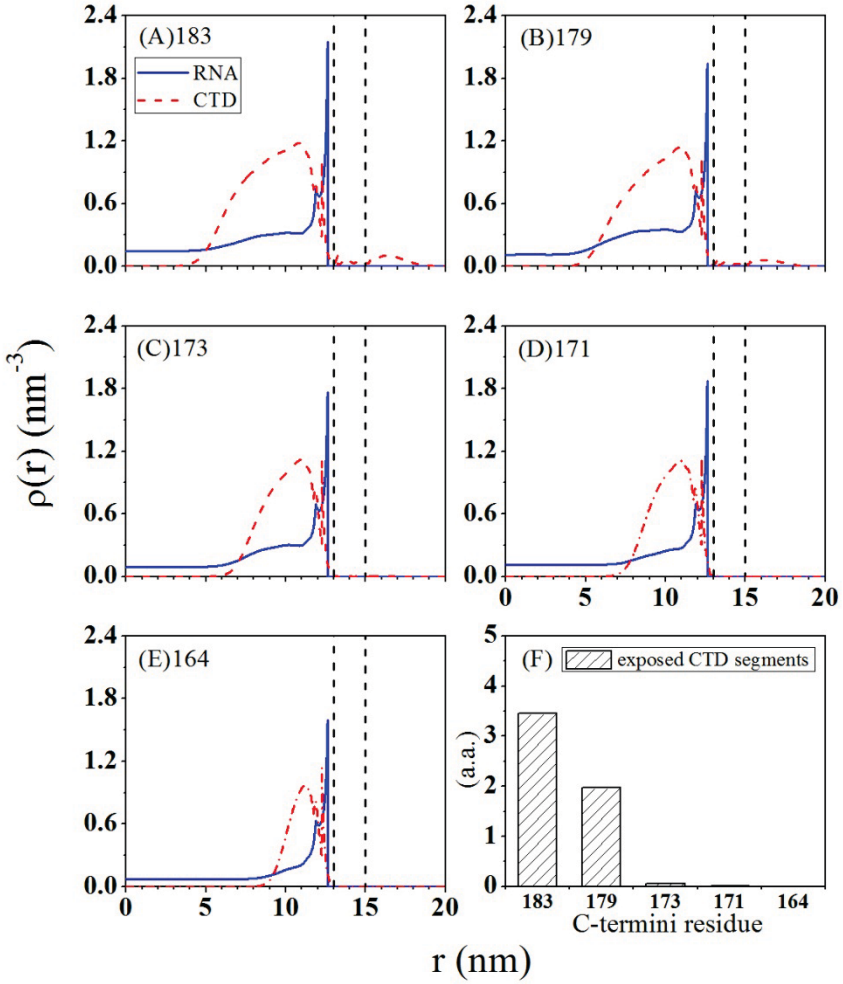


Figure 5-5. Radial distributions of RNA (blue solid line) and CTD chains (red dashed line) in WT and mutant HBV nucleocapsids (A-E). Here *r* stands for the radial distance from the capsid center; the perpendicular dashed lines indicate the position of the capsid shell. (F) The average number of amino acid (a.a.) residues from the CTD tails located outside the capsid wall for WT and mutants.



Chapter 6. Molecular recognition in envelopment of hepatitis B virus

Abstract

Hepatitis B virus (HBV) virions have a double-layered structure with an icosahedral nucleocapsid at the inner core and an outer shell of proteolipids. Although the internal domains of envelope proteins interacting with cytosolic nucleocapsids have been recognized, molecular details about the envelope-capsid association remains little known. In this study, we identify probable interaction sites between the Pre-S domain of large surface proteins (L-HBs) and the capsid core proteins (HBc) based on a combination of Molecular dynamic (MD) simulations and molecular docking. The theoretical results suggest that the specific binding of Pre-S domain of L-HBs with HBc arises from recognition between proteins domains consisting of 101-108th residues from L-HBs and 135-139th residues from HBc. A theoretical analysis based on the Molecular Mechanics-Poisson-Boltzmann surface area (MM/PBSA) method indicates that the approximated binding energy is -6.7 kcal/mol, which is weaker than that for typical cellular receptor-ligand binding. The capsid-envelope association is enhanced by multi-body cooperative effects and may be regulated by the spatial location of flexible C-termini of HBc. The theoretical model provides insights on selective envelopment of HBV capsids at different degrees of maturation.

Introduction

Hepatitis B virus (HBV) virions have a double-shell structure with an icosahedral capsid coated with a lipid membrane containing there types of surface proteins. The capsid consists of 180 or 240 copies of the core protein (HBc) arranged in an icosahedral structure with triangulation (T) number 3 or 4. The majority of wild-type (WT) virions contain T4 capsids, about 28 nm in outer diameter. After envelopment with the surface proteins, the out diameter of the entire virion (Dane particle) is about 44 nm[4-6]. The envelope is composed of the HBV surface proteins (HBs) imbedded in a lipid bilayer. The glycoproteins exist in three related forms, small, middle and large, defined according to extended N-terminal domains (Pre-S1 or Pre-S2) of the surface proteins after a common S domain. The small form (S-HBs) has 226 amino-acid (aa) residues; these proteins are associated with the lipid and have four trans-membrane domains. Addition of Pre-S2 (55 aa) to S-HBs become the middle form (M-HBs), and Pre-S1 (108 aa) plus Pre-S2 extension to S-HBs become the large surface protein (L-HBs). The Pre-S domains are flexible and not imbedded in the lipid membrane. The S and L forms of HBs are indispensable for HBV envelopment and infection but the M-HBs are not essential for virion formation[27, 30].

In addition to the HBV virion, additional viral particles are secreted from infected cells. The so-called subviral particles (SVP) exist either in long filaments with 22 nm diameter in cross-section diameter and variable length or spherical particles 22 nm in diameter. SVP mainly consist of lipid molecules and the envelope proteins[31]. While SVP outnumber virions in the serum of HBV patients, their biological function remains

unknown. It has been suggested that SVP contribute to a persistent infection by enhancing immune tolerance[32]. The spherical SVP are predominately composed of S-HBs, and the filaments include also L-HBs[33]. SVP induce shared immunological responses with the virions, and thus can be used as effective immunogens for HBV treatments. For example, the first human vaccine for HBV was developed from expression of s-HBs particles in yeast[34].

The HBV virions are generated from association of matured nucleocapsids with a HBs-containing Golgi membrane before undergoing the constitutive secretory pathway[32, 35]. It was shown that L-HBs plays a critical role in the envelopment of nucleocapsids[36], as well as in signaling cell binding with the receptors of hepatocytes[37]. The Pre-S domain of the L-HBs plays a critical role to mediate those interactions. L-HBs take an alternative topology for binding the capsid proteins and hepatocyte receptors, with the Pre-S domain located at inside and outside the viral membrane, respectively. It has also been suggested that the internal Pre-S domain mediates the interaction with cytosolic nucleocapsids[38, 39]. Mutagenesis studies reveal candidate residues in HBc that interact specifically with the envelope proteins. Those residues may be located at the tip of the HBc spikes [40] or around the base of 4-helix bundle where the C-terminus residues exist close to the capsid pores[41]. However, it is unclear how the envelope and the capsid proteins are engaged in a HBV virion. One key question is whether there are specific interactions between HBs and HBc. Although electron microscopy measurements point to a non-icosahedral arrangement of HBs in the envelope[42], Seitz *et al.* proposed a distinct packaging rule for HBs, which supports specific interactions of the surface proteins with the nucleocapsid[43].

The putative interactions of HBc with the Pre-S domain of HBs may be considered as a special type of regulatory cellular processes that are mediated by protein-peptide interactions. Although high-resolution analysis, such as NMR spectroscopy and X-ray crystallography, is able to uncover the atomic structure of many protein-peptide complexes, the structural information for the putative protein-peptide pair in HBV remains unavailable. Because the experimental procedure to determine the mechanism of recognition and binding processes is challenging, theoretical modeling and computations provide a valuable alternative. In this work, we identify probable interactions between the Pre S2 domain of L-HBs and the core protein HBc using MD simulation and molecular docking methods. We find a binding site in HBc that consists of amino acid (aa) residues from 139 to 142. It binds with a peptide made of 101-108 aa of L-HBs to form the HBc-HBs complex. The binding free energy, estimated from the Molecular Mechanics-Poisson-Boltzmann surface area (MM/PBSA) analysis, is -6.5kJ/mol for each HBc-HBs pair. Our theoretical analysis supports the specific binding of HBc with the Pre S domain. By docking with 4 residue sub-fragments of Pre S2 for HBc, we find that a Pre S2 fragment (105-108 aa) shows a strong affinity with a binding site (around 139-142 aa) from HBc that is directly connected with the liker (141-149) and C-terminal domains (CTD). The dynamic structures of CTD at different stages of viral replication suggest that HBV envelopment coordinates with genome maturation within the nucleocapsids.

Materials and Methods

Molecular Systems

In this work, we chose the initial configuration of the HBc protein from the crystal structure for the HBV capsid core protein composite, PDB 1QGT[191]. The complex is composed of 4 quasi-equivalent HBc monomers, labeled as A, B, C, and D. The structure for chain D is used as the initial input for the HBc monomer. The structure of HBs is currently not yet available; we use PEP-fold[246-248] to obtain the conformation of the Pre-S domain in L-HBs. For effective computation of HBc-HBs binding free energy, we consider only the sub-fragments of in the Pre-S domain, as listed in Table 1. Initial conformations of these peptide fragments were directly taken from their 3-D coordinates in the Pre-S structure predicted by PEP-fold.

MD Simulations

We performed MD simulations with an implicit water model, which requires much less computation power and enables systematic analysis of the unknown protein-peptide binding. All simulations were carried out with a standard package, Gromacs 4.5[249], with OPLA-AA force field[250].. Generalized-Born (GB) potentials were applied in representing implicit water[251], and no counter ions were added to neutralize the systems. After an energy minimization for 2000 steps, 10-ns production runs were performed for each system with HBc protein and a randomly positioned peptide fragment. The temperature was kept at 300 K by the velocity rescaling thermostat[252]. The resulting trajectories were collected every 2 ps with a 2-fs time step. For the analysis of the root-mean-square deviation (RMSD), we only considered trajectories from 2 – 10 ns.

To quantify the interaction free energy between HBc and the bound peptide fragment, we performed additional MD simulation with an explicit water model. We used the end point of previous MD run in implicit solvent as the starting input for the latter MD run in explicit water system. AMBER 10 package[253] was used to obtain the additional 2 ns production run, with TIP3P water model[254] and OPLS-AA force field[250]. To calculate the binding free energy, we applied the molecular mechanics/Poisson Boltzmann surface area (MM/PBSA) method[255, 256], which is implemented in the AMBER package.

Molecular Docking

We used Autodock 4.2 with explicit consideration of the flexibility of the protein receptor and peptide ligand[257]. The Autodock procedure includes two steps, autogrid run to calculation the map of interactions upon the selected binding site in receptor with general atoms, and autodock run to pose the ligand with respect to the map of interactions. We employed the same initial structures of HBc protein and peptide fragments used in the MD simulation as the receptor and ligand respectively. We chose 30 of probable grid maps with grid spacing of 0.375 Å and dimensions of 30 x 30 x 30 points. These grids were centered at varied positons to cover the whole surface area of the HBc. For initial input structures, Gasteiger partial charge was applied[258] and hydrogens were added by AutoDockTools. Lamarckian genetic algorithm (LGA) method was employed in the docking process[259] .

Results and Discussion

MD simulations for cases with HBc protein and each of 8-mer peptide were performed in implicit water system. 10 selected peptide fragments of pre-S1 domain in L-HBs are listed in Figure 6-1. Each of the peptide (P1~P10) was placed at random position around HBc monomer initially, then parallel MD runs for each case were performed. In most cases, the peptide moved around the protein without making a fixed binding with the protein. However, P4 peptide fragment is associated with protein in an early time (< 2ns), then the binding continued until the end point of simulation (100ns). Figure 6-2 shows snap shots of the P4 peptide- HBc protein conformation during MD run. Within 2ns time frame, the peptide made a contact to protein and stick on the surface of the protein, and such a bound state was preserved with a small position shift of the peptide on around similar binding domain in the protein. The binding domain is located at sequence 135-139 of the protein, which is relatively flexible loop region located at initial portion of the linker domain of the HBc.

Within 2ns, the P4 peptide fragment associates with HBc monomer. And the binding is mediated at the domain of 135-139 aa in HBc. Figure 6-3(A) shows distance change between P4's center of mass and that of putative binding domain (135-139) in HBc. In the plot, the initial up and down present the flowing around movement of the peptide. At ~1.1 ns, however, the distance converges within 2 nm. In comparison with RMSD data (Fig 3(B)), we can see the peptide get close to the protein at around 1ns, while its conformation converges into stable form slightly later as ~1.1ns. We can see a convergence of the peptide conformation within approximately 1.0 Å resulted from its

association with the protein at around 1.5 ns, while certain range of structural fluctuation exist indicating it may not reached to the final state of binding. In general, it has been indicated the equilibrated state of protein-peptide complex would be detected by quite long simulation analysis with several μs, which is order of magnitude beyond the time frame tested in this study. However, it is clear that there exist a selective binding between a certain fragment (P4, 101-108 aa) of pre-S1 domain in L-HBs and HBc protein, Although this binding is determined at the beginning step such as in the time frame, it still possesses general characteristics of the specific peptide-protein association.

To analyze the mechanism of binding, we determine the interaction energy between the P4 peptide and HBc monomer by MM/PBSA method. For that calculation, we performed additional MD run in explicit water system. As shown in Table 1, the approximated internal energy (ΔE_{tot}) in the peptide-protein complex formation is ~-36 kcal/mol, and corresponding entropic energy change ($T\Delta S$) is ~-29 kcal/mol. Accordingly, the binding free energy (ΔG) of -6.67 kcal/mol is obtained. Such an energy profile indicates the interaction is relatively weaker than general case of peptide-protein binding for ~8 mer peptide (~-10 kcal/mol[260]). However, in the cellular signaling process, such a weak ligand-receptor association is a specific aspect as shown in the T-cell response[261, 262], and it could be expected the envelope coating process also requires the weak protein association. A comparison of polar ($\Delta U_{Coul} + \Delta W_{PB}$ = 13.5 kcal/mol) and nonpolar ($\Delta U_{VDW} + \Delta W_{np}$ = -49.18 kcal/mol) contributions indicates that the overall binding is driven by the hydrophobic interactions. Such a favorable hydrophobic and

unfavorable electrostatic contribution in protein-peptide association has been observed in many promiscuous protein recognitions[263-265].

To further confirm this molecular recognition, we conducted the docking analysis on the P4 peptide- HBc association. It has been suggested a probable docking of flexible peptide is mostly allowed for a short peptide composed by less than 5 residues. Accordingly, we considered 4-mer sub-fragments of the P4, as listed in Fig. 1(C), and conducted parallel trials of the docking procedure on to overall surface of HBc protein. Among 20 of selected grid positions for each of 5 sub-P4 fragments, 3 pairs of successful docking with negative binding energy were obtained. Interestingly, one of them is the case of P4-3 docked on the domain of 135-139 aa in HBc protein. This complex formation is clearly overlapped with the P4-HBc binding conformation predicted by MD simulation. As seen in Figure 6-4, the P4-3 peptide, representing 103-106 aa in L-HBs, is supposed to be well adopted into a narrow loop region that is close to linker domain in HBc.

It is know that there is a selective mechanism in the HBV virion formation, such that only the mature nucleocapsids with dsDNA can be enveloped, while the immature nucleocapsids with ssRNA or ssDNA are excluded in that envelopment process. Regarding such a maturation depending mechanism, our results on the probable binding between L-HBs and HBc suggest a possible scenario. MD simulations and the docking analysis indicate the protein recognition is mediated by 135-139 aa domain in HBc, which is located beside the flexible C-terminal region. Accordingly, we consider that such a binding can be inhibited by C-terminal domain (CTD, 150-183 aa) of HBc. The

structural information of CTD is not currently available but the equilibrium structure of HBV nucleocapsid including the CTD chain location can be determined by the density functional theory (DFT) calculation[112, 266]. Figure 6-5 shows the DFT predictions on the radial distributions of CTD residues in 3 different capsids, the mature form with dsDNA; immature one with ssRNA; and empty capsid without any genome content. The CTD chains tend to be associated more strongly with dsDNA with all residues residing inside capsid. On the other hand, a marginal portion of CTD segments are to be located beyond inner capsid wall in the immature capsid. For empty capsid, even more portion of CTD residues are extruding outside capsid region. Such a varied CTD positioning with respect to genome contents strongly suggests the selective mechanism in the envelopment might depend on the CTD location. When CTD chains are totally internalized, i.e., in mature nucleocapsids, the putative core-envelope interaction should not be disturbed. However, if the CTD chains are placed close to inside capsid wall or block the pore area on the capsid, then the contact of L-HBs to the binding site in HBc (135-139 aa) could be easily inhibited. Such a scenario explains why only the mature nucleocapsid can be enveloped.

Conclusion

In this study, we investigate capsid-surface protein association using MD simulation and molecular docking. We find that HBV envelopment is mediated by the docking of Pre-S1 domain from L-HBs into pre-linker domain from HBc. The combined analysis of MD simulations using both implicit and explicit water models as well as molecular docking reveals specific peptide-protein binding underlying the HBs-HBc association. We found the domains of 101-108 aa in L-HBs and 135-139 aa in HBc are

critical in this molecular recognition. The L-HBs-HBc association has a lower binding free energy in comparison with typical cellular receptor-ligand binding, indicating that envelopment is a cooperatively driven multi-body process. Considering that the binding site is close to the C-terminal domain (CTD) of HBc, we expect that the core-envelope association is closely coordinated with the nucleocapsid maturation. In immature nucleocapsids, the putative docking of L-HBs to HBc is blocked because the CTD tails are partially exposed and cover the capsid surface. The dynamic nature of the CTD distribution explains the selective envelopment of HBV nucleocapsids during virion formation.

Figure 6-1. The sequence of amino-acid (a.s.) residues in the Pre-S domain of the large HBV surface protein (L-HBs). (A) N-terminal residues (1-174 aa) of L-HBs. (B) 8-mer peptides (P1 ~ P10) used in MD simulations. (C) 4-mer peptides (P4) used for molecular docking analysis.

(A)

1 MGGWSSKPRQ GMGTNLSVPN PLGFFPDHQL DPAFGANSNN PDWDFNPNKD HWPEAHQVGA
GAFGPGFTP**P HGGLLGWSPQ AQGILTTVPV APPPASTNRQ SGRQPTPISP PLRDSHPQAM**
QWNSTTFHQA LLDPKVRGLY FPAGGSSSGT VNPVPTTASP ISSIFSRTGD PAPN 174

(B)

P1 (89-96): **PVAPPPAS**
P2 (93-100): **PPASTNRQ**
P3 (97-104): **TNRQSGRQ**
P4(101-108): **SGRQPTPI**
P5(105-112): **PTPISPPL**
P6(109-116): **SPPLRDSH**
P7(113-120): **RDSHPQAM**
P8(117-124): **PQAMQWNS**
P9(121-128): **QWNSTTFH**
P10(125-132): **TTFHQALL**
P11(129-136): **QALLDPKV**
P12(133-140): **DPKVRGLY**

(C)

P4 (101-108): **QSGRQPTPIS**
P4-1 (99-102): **RQSG**
P4-2(101-104): **SGRQ**
P4-3(103-106): **RQPT**
P4-4(105-108): **PTPI**
P4-5(107-110): **PISP**

Figure 6-2. Snap shots from MD simulation trajectory for a P4 peptide interacting with the HBc protein. (A) Initial conformation with P4 randomly placed around HBc. (B) The P4-HBc bound structure at 2 ns.

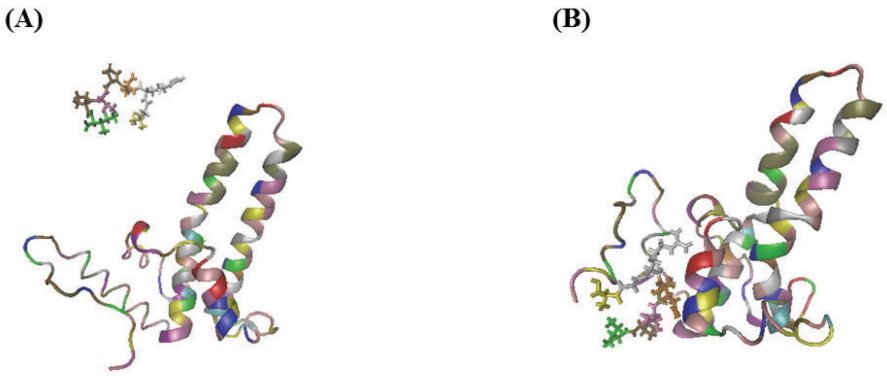


Figure 6-3. Dynamic properties of P4-HBc association. (A) Change of the distance between the averaged centers of P4 peptide and domains (130-134 aa; 135-139 aa) from HBc. (B) Root mean square distribution (RMSD) of the backbone atoms in the P4 peptide.

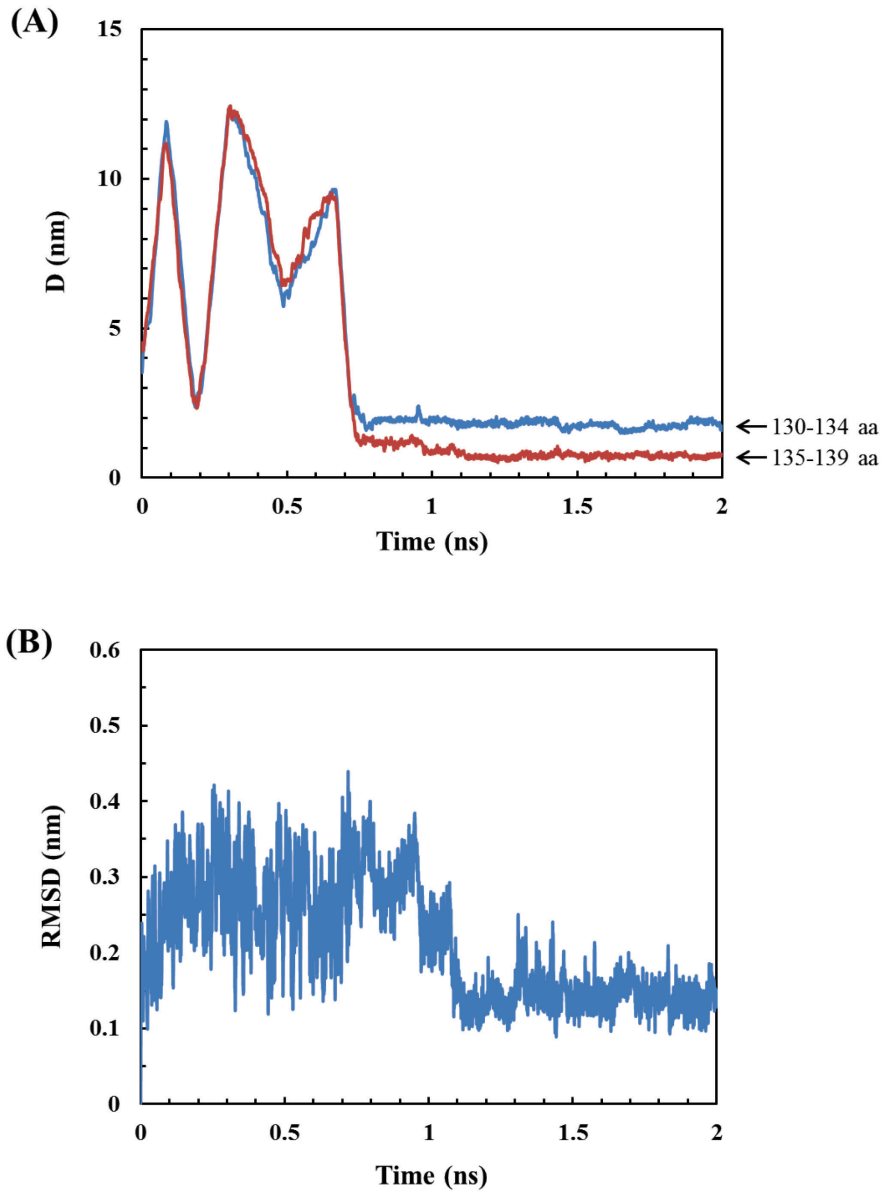


Figure 6-4. Docking of sub-fragment P4-3 onto HBc. Shown here are the docked conformation for P4-3 (101-103 aa of L-HBs) and the approximated binding energy.

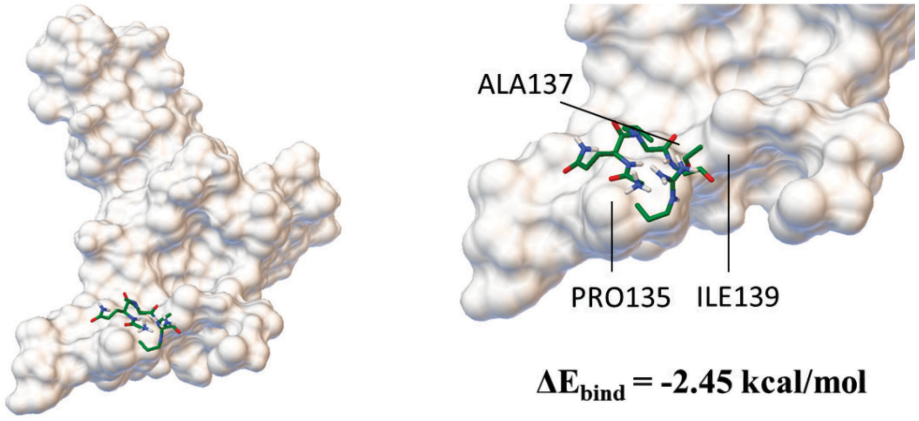


Figure 6-5. Radial distributions of CTD chains at different stages of maturation. Density profiles (ρ(r)) of CTD residues are compared for the mature (with dsDNA, red line), immature (with ssRNA, blue line) and empty (green line) capsids. Perpendicular dash lines indicate the capsid wall.

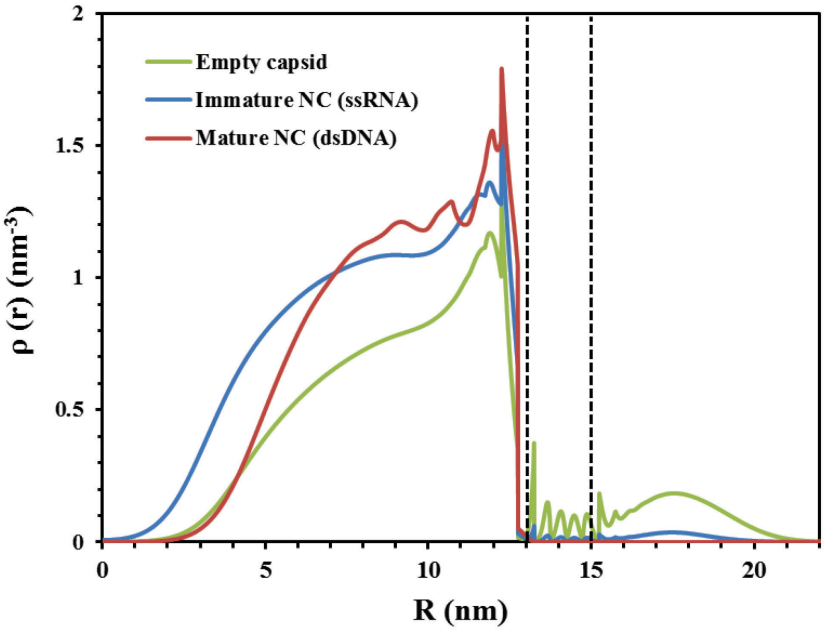


Table 6-1. Protein-peptide interaction energy (kcal/mol). ΔU_{VDW} is the van der Waals energy; ΔU_{Coul} is the electrostatic energy; ΔW_{PB} and ΔW_{np} are the polar and non-polar contributions to the the solvation energy; ΔE_{tot} is the total interaction energy, which is sum of ΔE_{polar} and ΔE_{np}; $T\Delta S$ is the entropic energy term; ΔG is the binding free energy.

ΔU_{VDW}	ΔU_{Coul}	ΔW_{PB}	ΔW_{np}	ΔE_{total}	$T\Delta S$	ΔG
-43.37	-292.19	305.69	-5.81	-35.66	-28.99	-6.67

Chapter 7. Thermodynamic and structural evidences for reduced hydrogen bonding among water molecules near small hydrophobic solutes

Abstract

The structure of water molecules near a hydrophobic solute remains elusive despite a long history of scrutiny. Here we reexamine the subtle issue by a combination of thermodynamic analysis for Henry's constants of several nonpolar gases over a broad range of temperatures and molecular dynamic simulations for the hydration energies using two popular molecular models of liquid water. Both the structural and thermodynamic data indicate that hydrophobic hydration reduces the degree of the hydrogen bonding among water molecules in particular at high temperature. Hydrogen bond formation is slightly hindered near a hydrophobic solute due to the anisotropic distribution of water molecules in the solvation shell, and the confinement effect becomes more significant as temperature increases. Reduction in the degree of hydrogen bonding is fully consistent with the positive contribution of a small hydrophobic solute to the solution heat capacity. As predicted by the scaled-particle theory, both Henry's constants and simulation results suggest that the hydration entropy is determined primarily by cavity formation in liquid water, with its magnitude increasing with the solute size but declining with temperature.

Introduction

Hydrophobic effect constitutes an essential component of biomolecular interactions and has been subjected to extensive studies [see, e.g.[154, 267-269] for recent

reviews]. While its macroscopic ramification is self-evident as immiscibility between water and oil or the tiny solubility of a nonpolar compound in liquid water[270], a precise description of hydrophobicity from a molecular perspective has been an enduring challenge. A common wisdom is that hydrophobicity is intimately affiliated with the disturbance of the hydrogen-bonding network among water molecules in the vicinity of nonpolar solutes or surfaces. However, direct evidence on the structural variation is elusive. It has been long debated whether a hydrophobic solute promotes or hinders hydrogen bonding among water molecules in the solvation shell[271-274].

Based on the thermodynamic data for the aqueous solubility of hydrophobic compounds, Frank and Evans first suggested in 1945 that "When a rare gas atom or a non-polar molecule dissolves in water at room temperature, it modifies the water structure in the direction of greater 'crystallinity' − the water, so to speak, builds a microscopic iceberg around it"[275]. Although Frank and Evans cautioned at the outset that "iceberg" is by no means ice-like, the iceberg picture is often literally interpreted as formation of a cage-like ordered structure of water molecules near a hydrophobic solute, concomitant with an increased propensity for hydrogen bonding. While such an interpretation readily explains the negative hydration enthalpy and the entropy loss associated with solvation of a small nonpolar solute at room temperature,[275-277] there have been ample experimental evidences revealing that hydrophobic hydration results in little change on hydrogen bonding among water molecules[278-280]. In stark contrast to the iceberg model, it has been suggested that, near the interface of an organic liquid and an aqueous solution, the degree of hydrogen bonding among water molecules decreases in

comparison to that in bulk water[281]. At the molecular scale, recent thermodynamic and neutron scattering studies render no direct evidence for enhanced ordering of water molecules near a hydrophobic solute[282-284].

Thermodynamics of hydrophobic hydration can be alternatively explained in terms of a balance between the loss in entropy due to cavity formation to accommodate a hydrophobic molecule and the gain in enthalpy due to van der Waals attraction between the solute and solvent molecules[285-289]. Cavity formation and van der Waals forces are equally applicable to solvation in polar as well as nonpolar solvents[290-292]. Such an explanation entails no specific information on the self-organization or hydrogen bonding among water molecules. It has been argued that the entropy affiliated with hydrophobic hydration is particularly negative owing to the small size of water molecules, little affected by the variation of hydrogen bonding within the solvation shell[293]. The reorganization of hydrogen bonds around the nonpolar solute has little effect on the hydration free energy[294].

The lack of ice-like tetrahedral structure around hydrophobic solutes does not necessarily preclude the increase of hydrogen bonding among water molecules within the hydration shell[267, 269, 271, 283, 289, 295-300]. There have been substantial experiments revealing a moderate slowdown of water molecules around nonpolar solutes, suggesting more hydrogen bonding[301-309]. A recent study based on molecular dynamics (MD) simulations shows enhanced hydrogen bonding at least for a subset of water molecules around a hydrophobic solute[272, 310]. Because MD simulation typically employs a semi-empirical

force field with hydrogen bonding depicting as electrostatic interactions, the pairwise additive potential does not reflect the quantum characteristics of bond formation.

Whereas a full quantum-mechanical calculation of hydrophobic solvation remains computationally demanding, hydrogen bonding is at least partially reflected in the thermodynamic data for the solvation of hydrophobic compounds. In this work, we investigate the solute effects on the degree of hydrogen bonding by a combination of MD simulation and experimental data for the solubility of several small nonpolar solutes in water over a broad temperature range. With the solvation free energy extracted from Henry's constants at different temperatures, we are able to obtain entropic and energetic properties of different solutes using standard thermodynamic relationships. In conjunction with MD simulation for obtaining the van der Waals attraction between solute-solvent molecules, the hybrid approach allows us to acquire both energetic and structural signatures of local hydrogen bond formation without direct computation of the hydrogen-bonding energy using the electrostatic models.

Molecular model and methods

Thermodynamics of solvation

Solvation free energy is conventionally defined as the reversible work to transfer a solute molecule from a hypothetical ideal-gas state into an ideal solution of the same temperature and the solute number density. Alternatively, it can be understood as the excess chemical potential of the solute molecule at infinite dilution in a solvent, here liquid water, at fixed temperature and pressure. In terms of the gas-liquid equilibrium, the solvation free energy (F_s) can be related to Henry's constant for the solute[284]:

$$F_s = -k_B T \ln(HRT), \tag{7.1}$$

In Eq.(7.1), H denotes Henry's constant in units of mol/(liter·atm), R = 0.082 liter·atm/(mol·K) is the gas constant, T is the absolute temperature (K), and k_B is the Boltzmann constant. The derivatives of the solvation free energy yield the entropy and enthalpy of solvation:

$$\Delta s = -\left(\frac{\partial F_s}{\partial T}\right), \tag{7.2}$$

$$\Delta h = \frac{\partial(\beta F_s)}{\partial \beta}, \tag{7.3}$$

where $\beta = (k_B T)^{-1}$. In Table 1, we tablet the experimental data for Henry's constants of four solutes, *viz.*, argon, neon, methane and neopentane, in liquid water at 1 atm over the temperature range from 0 to 50° C. Henry's constants for other noble gases [311] and low-molecular weight alkanes[312] are also available from the literature. Based on the experimental results, Eq.(1-3) allow us to determine the hydration free energy, hydration enthalpy, and hydration entropy of small hydrophobic molecules.

Scaled-particle theory

The scaled-particle theory (SPT) was originally developed for deriving the equation of state for hard-sphere fluids[313]. Similar concepts can be used to predict the free energy of cavity formation in any bulk liquid[291]. Augmented with a semi-empirical equation to estimate the solute-solvent van der Waals attractive energy, the SPT model has been successfully used to predict solubility, enthalpy of solvation, and partial molar volume for diverse gases in both organic and aqueous solvents[286].

According to SPT, the solvation free energy consists of two contributions, one due to cavity formation to accommodate a solute molecule, and the other due to the long-range interactions between the solute and solvent molecules. If the solute molecule is represented by a spherical particle of diameter σ_0, the free-energy cost for cavity formation in a liquid of molecular number density ρ and pressure $\boldsymbol{P}$ is given by[313]

$$\beta F_0 = -\ln(1-\eta) + \left(\frac{3\eta}{1-\eta}\right)R + \left[\frac{3\eta}{1-\eta} + \frac{9}{2}\left(\frac{\eta}{1-\eta}\right)^2\right]R^2 + \frac{\beta P\eta}{\rho}R^3. \qquad (7.4)$$

where $\eta = \pi\rho\sigma^3/6$ stands for the packing fraction of solvent molecules, σ represents the hard-sphere diameter of the solvent molecules, and $R = \sigma_0/\sigma$. In deriving Eq.(7.4), it is assumed that the excluded volumes of both the solute and solvent molecules can be represented by hard-sphere interactions; the cavity is devoid of any solvent molecule with its center of mass located within a radial distance less than $(\sigma_0+\sigma)/2$ from the center of the solute molecule.

In this work, we use SPT to investigate the effect of solute size on the hydration entropy. Specifically, we consider the cavity formation free energies for Ne, Ar and CH_4 in water at 1 atm. To calculate cavity solvation free energy from Eq.(7.4), we need experimental data for the bulk solvent density as the input. The solvent density from 273 to 323 K is obtained from an empirical correlation of the experimental data for water density[314]. At 298 K, for example, the number density of the solvent is $\boldsymbol{\rho = 33.36\,nm^{-3}}$, corresponding to the weight density of liquid water, $d = 997\,\mathrm{kg/m^3}$. The hard-sphere diameter for the solvent is assumed identical to the Lennard-Jones size parameter for

liquid water according to the SPC/E model, i.e., σ = 3.166 Å. The hard-sphere diameters for Ne, Ar and CH_4 are 3.09 Å, 3.56 Å and 3.73 Å, respectively. The cavity sizes for Ne and Ar are chosen according to Henry's constants predicted by SPT[315], while that for CH_4 is from the size parameter in the OPLS all-atom force field[250].

Molecular dynamic (MD) simulation

All simulations reported in this work were conducted using the GROMACS 4.5.1 program[249]. TIP3P[254] and SPC/E[316] water models were used to describe interactions between water molecules. The solute-solvent interactions were described with the Lennard-Jones (LJ) potential and the Lorentz-Berthelot combining rules. The LJ parameters for the solute molecules were obtained from the OPLS all-atom force field[250]. The simulation box consisted of a single non-polar solute fixed at the center and 466 water molecules under the isothermal–isobaric (NPT) condition. In all simulations, the pressure was fixed at P = 1 bar, and the temperature varies from 273.15 K to 323.15 K at 5 K interval. In MD simulation runs, we used the Berendsen thermostat for controlling the temperature with a coupling constant of 0.1 ps[317], and the Parrinello-Rahman extended-ensemble coupling for the pressure control with a coupling constant of 2 ps[318]. Long-range electrostatic interactions were treated with the particle-mesh Ewald method and periodic boundary conditions[319]. The residual electrostatic and van der Waals interaction cutoffs were set to 1 nm. The equations of motion were integrated with the leapfrog algorithm at a time step of 2 fs, and the coordinates were saved at every 2 ps. All simulation results were obtained by averaging 100 ns MD steps after 100 ns equilibrium steps at a given temperature and pressure for each solute.

Van der Waals' energy

We estimated the van der Waals energy between a spherical solute and water molecules based on the Weeks-Chandler-Anderson (WCA) potential, $\Gamma_{WCA}(r)$ [320]

$$F_{vdw} = \int d\mathbf{r}\rho(r)\Gamma_{WCA}(r), \tag{7.5}$$

where $\rho(r)$ represents the local density of water molecules at radial distance r from the solute center. For the solutes studied in this work, the LJ parameters are σ_{Ar}= 3.345 Å and ε_{Ar}/k_B= 125.7 K for Ar [321], σ_{Ne}= 2.782 Å and ε_{Ne}/k_B= 34.7 K for Ne [250], σ_{CH_4} = 3.73 Å and ε_{CH_4}/k_B= 147.9 K for methane[322], and $\sigma_{C(CH_3)_4}$= 5.759 Å and $\varepsilon_{C(CH_3)_4}/k_B$= 311.5 K for neopentane[322]. According to the SPC/E model[316], the LJ parameters for water are σ_{water}= 3.166 Å and ε_{water}/k_B= 78.2 K. The TIP3P model gives slightly different LJ parameters for water molecules, i.e., σ_{water}= 3.151 Å and ε_{water}/k_B=76.5 K. The Lorentz-Berthelot combining rules are used for water-solute interactions.

Number of hydrogen bonds

We calculated the average number of hydrogen bonds among water molecules using Gromacs command 'g_hbond'. A pair of water molecules are considered as hydrogen-bonded when the O-O distance R_{oo} is within 3.5 Å and the HO-O bond angle is no greater than 30°. The local hydrogen-bonding propensity is measured in terms of the average number of hydrogen bonds within equally separated co-centric shells from the solute surface.

Results and discussion

Distributions of water molecules around small hydrophobic solutes

We consider the solvation of 4 nonpolar compounds, Ne, Ar, CH_4 and $C(CH_3)_4$, in liquid water at P=1 atm and temperature varying from 273.15 K to 323.15 K. Both the SPC/E and TIP3P water models were used in our MD simulations. Figure 7-1 presents the radial distribution functions (RDF) of hydrogen and oxygen atoms of water molecules around each solute at 298.15 K. Approximately, the hydration shell for each solute may be identified as water molecules within the first peak of the RDF for oxygen atoms. Although inhomogeneity in distribution of water molecules is in evidence beyond the hydration shell, it is generally recognized that the long-range correlation makes a relatively minor contribution to hydrophobicity[284]. Qualitatively, the SPC/E and TIP3P models yield little difference in the radial distributions of hydrogen and oxygen atoms. Both models show that oxygen atoms are more preferred than hydrogen atoms in direct contact with a hydrophobic solute, and the trend becomes more distinctive as the solute size increases. The local density of hydrogen atoms also exhibits a maximum at the solute surface but the first peak is noticeably broader, more asymmetric, and smaller in magnitude in comparison to that for oxygen atoms. The RDFs of oxygen and hydrogen atoms show a near random angular distribution of water molecules in the solvation shell. Clearly, the orientation of water molecules near a nonpolar solute is distinctively different from that near a spherical ion[323].

To calculate the van der Waals energy between solute and water molecules, we divide the LJ potential into a repulsive and an attractive contribution following the WCA

method[320]. Figure 7-2 shows the total solute-solvent van der Waals energy versus temperature estimated from the RDF for the oxygen atoms and the WCA decomposition of the LJ potential. Here the error bars represent the discrepancies between the SPC/E and TIP3P models used in the simulation. The van der Waals energy for large solutes is more sensitive to the water models, especially for $C(CH_3)_4$ at low (273 K) and high (323 K) temperature.

Figure 7-2 indicates that the van der Waals solvation energy increases with the solute size but is little affected by temperature. For all solutes considered in this work, the overall change in van der Waals energy is less than 0.5 kJ/mol over the temperature range from 273.15 K to 323.15 K. At low temperature (~273 K), the solute-water energy is comparable to the solvation enthalpy, suggesting that, at such conditions, the hydrogen bonding makes negligible contribution to the solvation energy. Because the van der Waals energy is almost invariant with temperature, it has little effect on the solvation entropy and makes virtually no contribution to the solution heat capacity.

The solvation enthalpy includes contributions not only from the solute-solvent van der Waals interaction discussed above, but also from additional terms due to the change in the system volume at constant temperature and pressure ($-p\Delta V$), and the changes in van der Waals energy ($\Delta U^{vdW}_{water-water}$) and hydrogen bonding among water molecules (ΔU^{HB}). As indicated above, we can determine the solvation enthalpy (ΔH) from Henry's constant using the Gibbs-Helmholtz equation. Besides, MD simulation provides direct information for both the changes in system volume and van der Waals attractive energy among water molecules. Figure 7-3 shows the simulation results for the

volumetric work ($-p\Delta V$), the change in van der Waals energy among water molecules ($\Delta U^{vdW}_{water-water}$), and the solute-solvent van der Waals energy ($U^{vdW}_{solute-water}$) for the four solutes at three representative temperatures. As expected, the volumetric work and the change in van der Waals energy among water molecules are much smaller in comparison to the solute-solvent energy, confirming a previous assumption that these terms are negligible in predicting the solvation energy of small molecule[284]. Meanwhile, the change in the van der Waals interaction among water molecules owing to the solute insertion ($\Delta U^{vdW}_{water-water}$) is also relatively small in comparison to the change in the internal energy of solvation (ΔU).

Figure 7-4 shows the temperature dependence of the solvation enthalpy for the four nonpolar solutes considered in this work. As well documented, hydrophobic hydration results in a negative solvation enthalpy at low temperature, and its magnitude increases with the molecular diameter. At high temperature, the hydration enthalpy becomes positive for Ne, and those corresponding to other solutes remain negative because of stronger van der Waals attractions. In comparison to the experimental results deduced from Henry's constants, the MD simulation predicts the solvation enthalpy noticeably too large at low temperature but too small at high temperature for Ne, Ar and $C(CH_3)_4$. As a result, we expect that the classical models yield a heat capacity smaller than the experimental value over the temperature range considered in this work. Similar discrepancy was identified in a previous simulation for xenon solvation in SPC/E and TIP5P models of liquid water[324]. Evidently, both MD and experimental results show a positive hydration heat capacity for small hydrophobic solutes, indicating that the

solvation entropy is increased with temperature. The entropy increase has been interpreted as the "melting" of water structure in the hydration shell[277], rendering a key evidence for the formation of clathrate-like structure near small hydrophobic solutes.

Also shown in Figure 7-4 are different components of the hydration enthalpy by its decomposition into the energy changes due to van der Waals (VDW) and hydrogen bonding (HB) interactions plus the volumetric work:

$$\Delta H = U^{vdW}_{solute-water} + \Delta U^{vdW}_{water-water} + \Delta U^{HB} + p\Delta V, \tag{7.6}$$

where $U^{vdW}_{solute-water}$ denotes the solute-water van der Waals energy, $\Delta U^{vdW}_{water-water}$ and ΔU^{HB} are, respectively, the changes in van der Waals and hydrogen-bonding energies among water molecules upon introducing a solute molecule. As discussed above, the hydration enthalpy is dominated by $U^{vdW}_{solute-water}$ and ΔU^{HB}.

Figure 7-4 indicates that the temperature dependence of the solvation enthalpy can be attributed mainly to the potential-energy change of the water molecules in the hydrophobic hydration shell. Near the normal freezing point of liquid water (273 K), the hydration enthalpy is virtually identical to the solute-water van der Waals energy, suggesting a negligible change in the total interaction energy among water molecules. As temperature increases, ΔU^{HB} becomes more positive, with the magnitude increasing with the size of the solute. While the presence of a solute molecule may slightly disturb the strength of hydrogen bonds[325], a more positive energy for the self-interaction reveals less hydrogen bonding among water molecules. In other words, the degree of hydrogen bonding within the solvation shell should be reduced in comparison to that in bulk water. While the reduction in the number of hydrogen bonds is insignificant at low temperature,

it becomes more appreciable as temperature rises. Clearly, the change in self-energy among water molecules contradicts to the prediction of the iceberg model. If a small nonpolar molecule induces a cage-like water structure, we would expect a decrease in the self-energy of water molecules.

While the reorganization of water molecules within the solvation shell depends on temperature, it appears that the reduction in the degree of hydrogen bonding shows no qualitative difference between small and large hydrophobic molecules. It has been suggested before that neopentane defines a length scale for the observation of large-scale hydrophobicity, *i.e.,* water molecules within the first solvation shell are similar to those near a paraffin wall, in contrast to those surrounding a small nonpolar solute such as argon and methane[326]. However, Figure 7-4 indicates that the change in the hydrogen bonding energy renders no clear evidence on the size-dependent crossover for hydrophobic hydration. As notated by Graziano[327], no distinctive features can be identified as the solutes vary from neon to neopentane, the smallest and the largest solutes studied in this work. For both cases, the change in water self-energy is smaller than that corresponding to the formation or broken of a single hydrogen bond (~ 20 kJ/mol).

To provide further insights into water reorganization within the solvation shell, we determine the average number of hydrogen bonds per water molecule in spherical shells of equal width (0.1 nm) around each solute. At a given molecular configuration, the number of H-bonds is counted for following the criteria that a pair of water molecules is considered H-bonded when the O-O distance is within 3.5 Å while the angle between

the intramolecular O-H bond and the intermolecular O-O bond is less than 30°[328]. Figure 7-5 presents the number of hydrogen bonds per water molecule near different solutes at three representative temperatures. In consistent with previous MD simulations based on SPC, SPC/E and TIP5P water models[324, 326], the number of H-bonds per water molecule is reduced in comparison to that in the bulk. The magnitude of H-bond reduction is the highest for the largest solute, neopentane, and the effect becomes most significant at high temperature.

According to a recent study by Davis *et al* based on Raman scattering measurements on water structure in dilute solutions of alcohol solutes[274], hydrophobic hydration leads to less hydrogen bonding among water molecules as temperature increases within the entire liquid range (0 ~ 100 °C). While the temperature effect agrees qualitatively with the results shown in Figure 7-5, one of the key conclusions in that work is that, near room temperature, the hydration of alkyl group enhances H-bonding among water molecules in comparison to that in bulk water. The discrepancy may arise from the close connection of alkyl chains with the hydroxyl group. Although the latter effect was examined by a comparison between the hydration spectrum for aqueous solution of 1-hexanol and 1,2-hexanediol, the water structure near alcohol solutes can be distinctively different from that near alkanes, resulting in opposite effects on hydrogen bonding among water molecules. Recent MD simulations with a polarizable force field indicate that the degree of H-bonding is enhanced only in the first water layer of solvation but reduced in subsequent water layers[271]. However, the overall change in H-bonding due to the hydration of a small solute such as methane is almost zero at room temperature,

confirming that the magnitude of H-bonding change is too small to imply any cage-like water ordering[273].

Origin of the solvation entropy

The iceberg model attributes the negative entropy of hydrophobic hydration to the enhanced water ordering around the solute molecule, which contradicts to the positive energy change for the self-interaction among water molecules. Alternatively, the negative solvation entropy may be simply explained in terms of cavity formation, i.e., the entropy loss arises from excluding the water molecules to accommodate a hydrophobic solute rather than formation of a ice-like structure[288]. Intuitively, cavity creation decreases both configurational and translational entropy for water molecules, and such effects can be easily quantified with the SPT model. Figure 7-6 shows that, for a small hydrophobic like argon, neon or methane, the solvation entropy is a negative and its contribution to the solvation free energy ($-T\Delta S$) declines almost linearly with temperature. The temperature dependence of the solvation entropy coincides well with that for a cavity of similar size as predicted by SPT. The discrepancy arises from the solute-solvent attraction that is absent for cavity solvation[300, 329]. Apparently, the increase in the solvation entropy with temperature is not unique for liquid water.

Conclusions

We have analyzed the effects of hydrogen bonding on hydrophobic hydration by combining the experimental data for Henry's constants and MD simulations for the solvation of several hydrophobic solutes in water over a broad range of temperatures. Neither energetic nor structural evidence could be identified to support a clathrate-like

water structure around non-polar solutes. In consistent with the energetic data derived from Henry's constants, the MD simulation indicates that H-bonding among water molecules is slightly reduced by the hydrophobic solvation and the effect becomes most evident as temperature increases. The reduction in H-bonding extent makes a key contribution to the positive hydrophobic solvation heat capacity. As predicted by the scaled-particle theory, the negative solvation entropy is mainly originated from cavity formation to accommodate the solute molecule, not from higher ordering of water structure.

It has been suggested before that the impact of H-bonding among water molecules on the hydrophobic solvation is marginal because of the compensation of the corresponding enthalpy and entropy contributions[294, 330-332]. Supporting this view, we clarify a small but clear reduction of H-bonding due to the solvation of small nonpolar solutes. Because van der Waals interactions between the solute and solvent molecules are relatively insensitive to temperature, the reduction in the degree of H-bonding consists of a key contribution to the positive hydration heat capacity (ΔC_p), which is a characteristic feature of the hydrophobic solvation.

Despite extensive studies, the definition of H-bond in liquid water remains elusive due to different perspectives on bonding or non-bonding properties[333, 334]. According to conventional MD simulation and ultrafast infrared spectroscopy, non-H-bonded configurations persist only within a short life time (< 200 fs)[333]. Contradictory interpretations could be deduced from the same simulation or experimental data, depending on the precise definition of the H-bonds[335]. Besides, classical models assume

that H-bond configurations are continuous rather than disruptive such as 'intact' or 'broken'[336]. Disputed arguments had also been given on the reorientation dynamics of water molecules within the hydration shell. For example, experiments based on the polarization-resolved mid-infrared spectroscopy detected slower motion of water molecules around hydrophobic groups in comparison to that in bulk water[304], supporting the iceberg picture. However, it has been suggested that the slowdown of hydration water motion may be attributed to a slower H-bond exchange due to the exclude volume effects by the solute, irrelevant to ice-like dynamics in the hydration shell[337].

Figure 7-1. The radial distribution functions (RDF) of water oxygen (OW) and hydrogen (HW) atoms near Ne, Ar, CH_4 and $C(CH_3)_4$ molecules at 298 K and 1 atm predicted by MD simulation. The results from the SPC/E model are shown in black lines, and those from TIP3P are in red; the solid lines correspond to solute (S)-water oxygen (OW) distributions, and the dashed lines are for solute (S)-water hydrogen (HW) distributions.

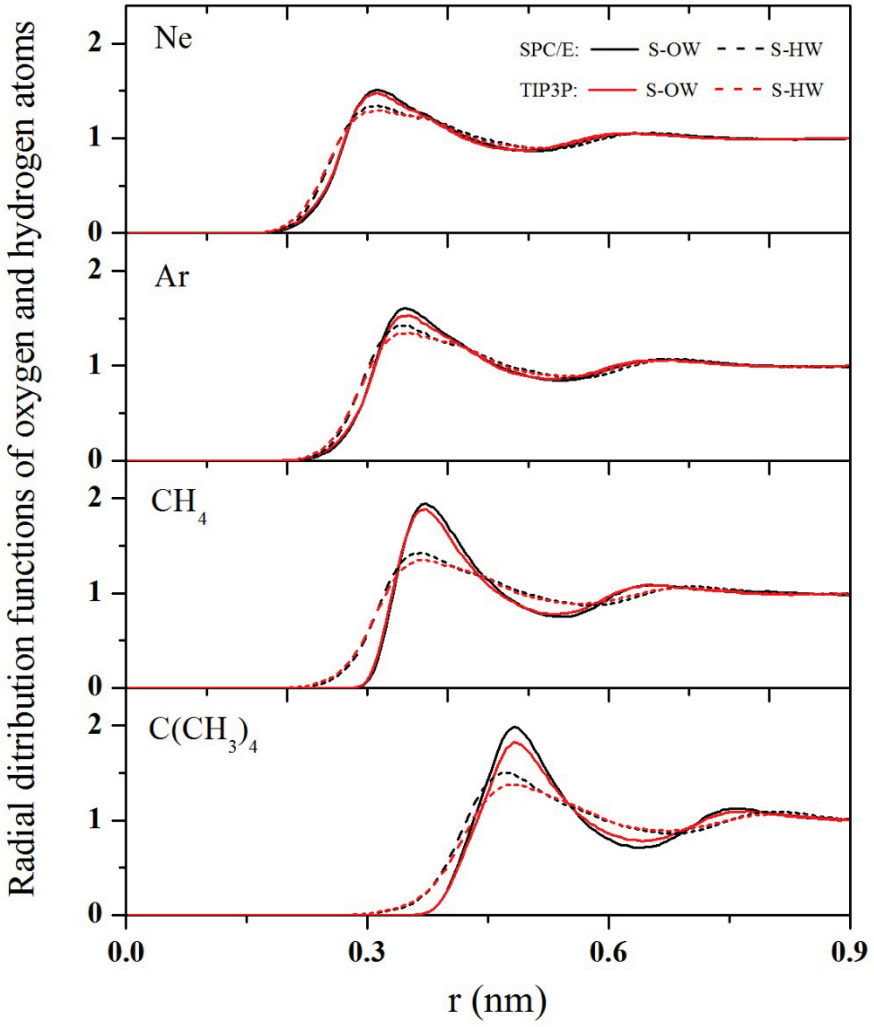


Figure 7-2. The temperature effect on the solute-solvent van der Waals energy for Ne, Ar, CH_4 and $C(CH_3)_4$. Error bars reflect the differences between SPC/E and TIP3P models.

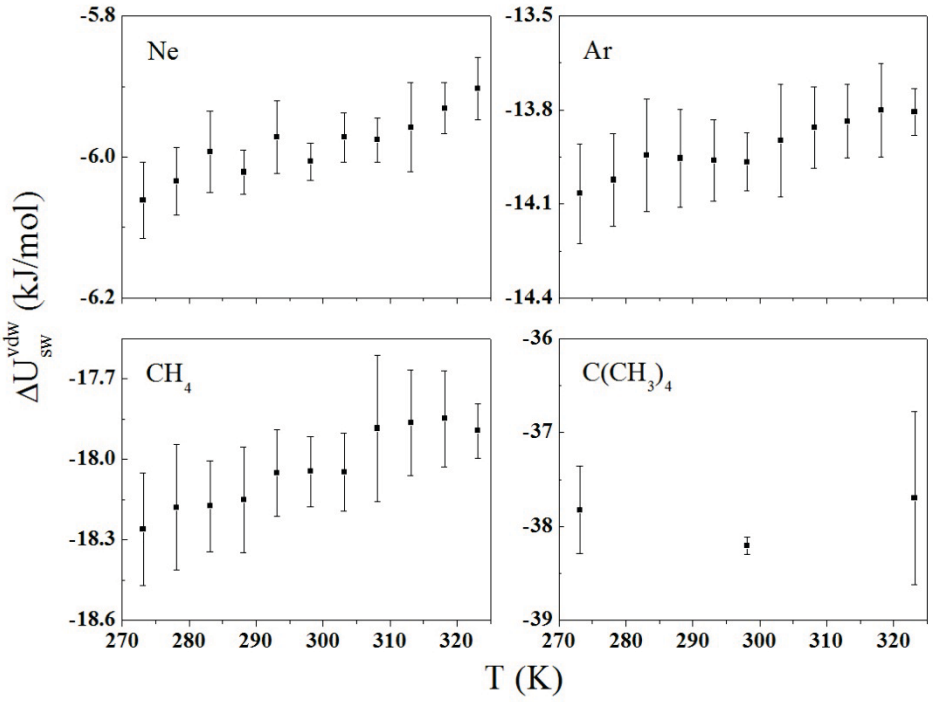


Figure 7-3. Different contributions to the hydration enthalpies of Ne, Ar, CH_4 and $C(CH_3)_4$ at three representative temperatures (273, 298 and 323 K): the energy corresponding to the change of solution volume ($p\Delta V$), the change in van der Waals energy among water molecules (ΔU_{ww}^{vdw}), and the van der Waals energy between the solute and water molecules (ΔU_{sw}^{vdw}).

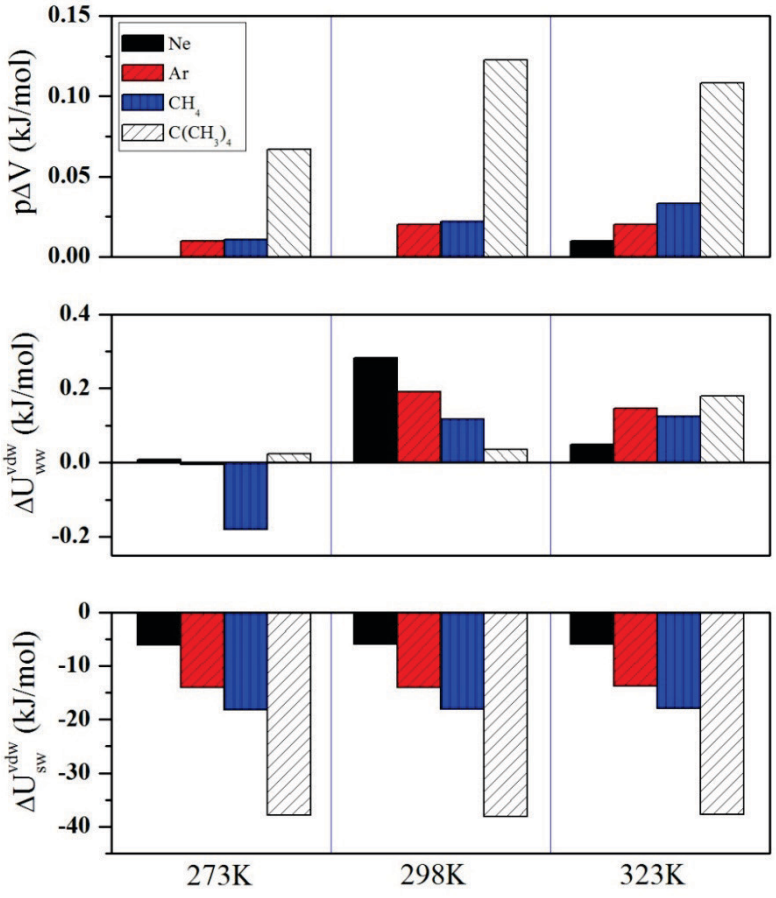


Figure 7-4. Energetic properties of hydration for Ne, Ar, CH_4 and $C(CH_3)_4$ from 270 K to 325 K. In each panel, the black line with filled squares represents the total hydration enthalpy deduced from Henry's constants, $\Delta H(\mathbf{Exp})$; the black line with opened squares represents the corresponding results from MD simulation, $\Delta H(\mathbf{MD})$; the green line with triangles designates the solute-solvent van der Waals energy, ΔU_{sw}^{vdw}; the blue line with diamonds is for the hydration enthalpy due to the overall change in the self-interaction among water molecules, ΔU_{ww}, and the red line with circles denotes the contribution to the hydration enthalpy due to the variation in hydrogen bonding, ΔU^{HB}.

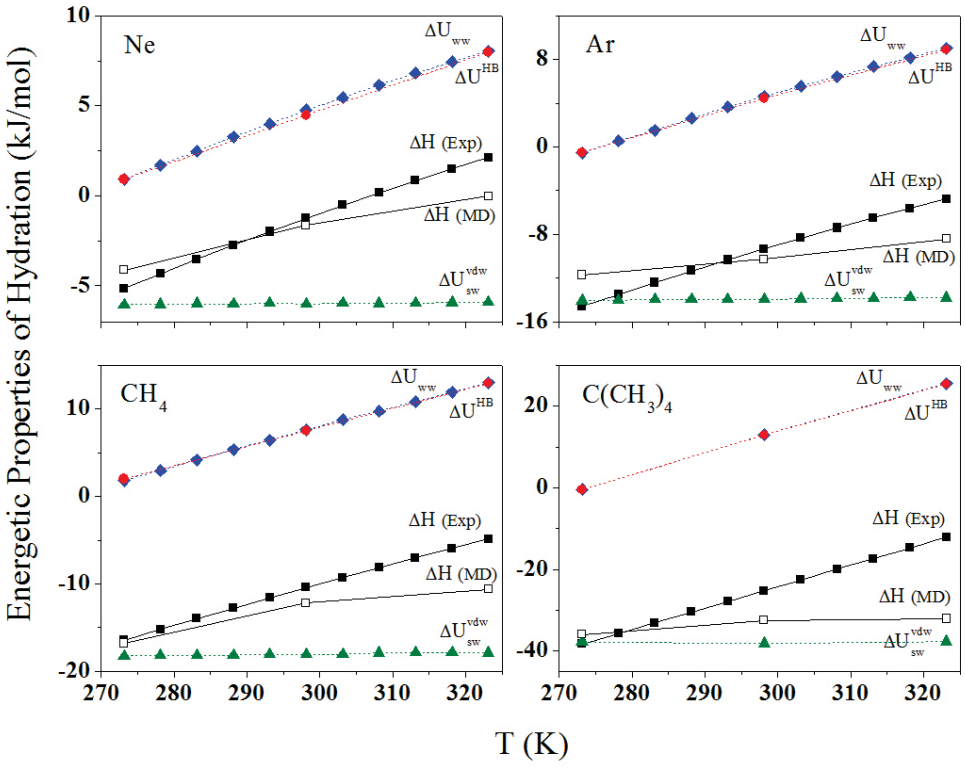


Figure 7-5. The average number of H-bonds per water molecule in the hydration shell of Ne (black squares), Ar (red circles), CH_4 (blue triangles) and $C(CH_3)_4$ (green diamonds) at three representative temperatures, 273, 298 and 323 K.

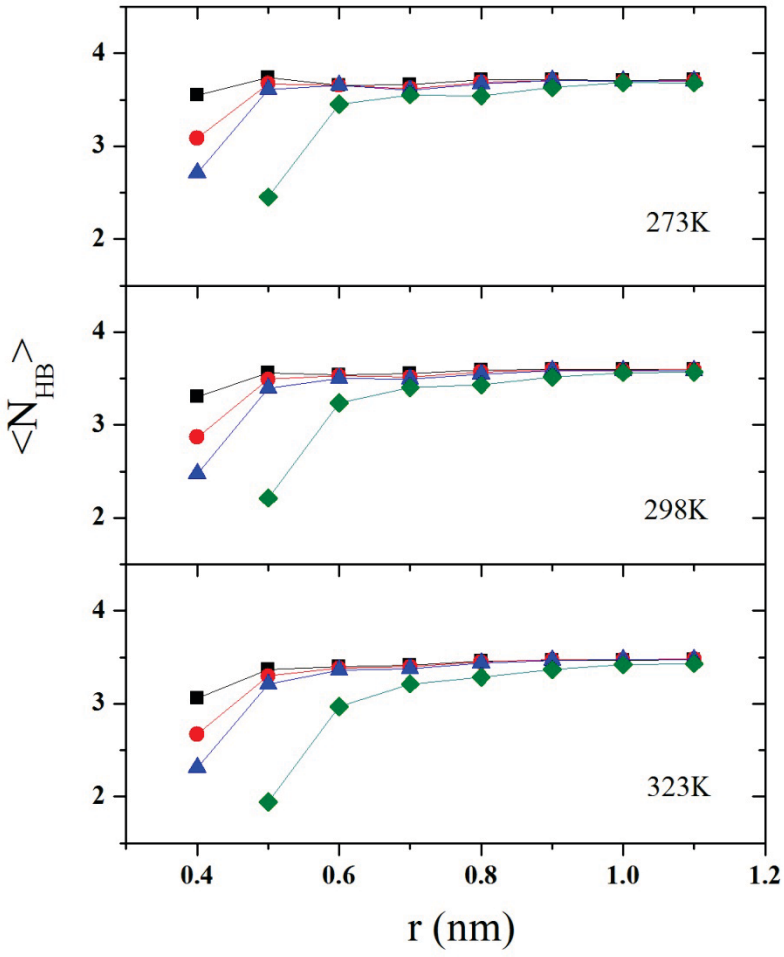


Figure 7-6. Effect of temperature on the hydration entropies deduced from Henry's constants for Ne (black squire), Ar (red circle) and CH_4 (blue triangle) and from the scaled-particle theory (solid lines).

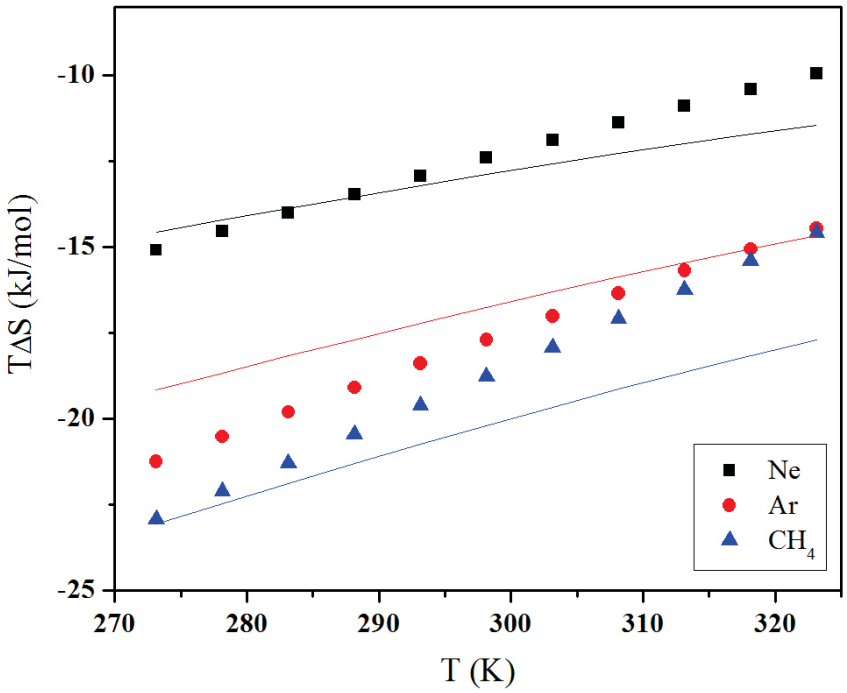


Table 7-1. Experimental data for Henry's constant (*H*) used in the thermodynamic analysis.

	H (M/atm)			
T (K)	Ne	Ar	CH_4	$C(CH_3)_4$
273	0.000517	0.002219	0.002556	0.002125
278	0.000502	0.002010	0.002210	0.001555
283	0.000487	0.001828	0.001938	0.001174
288	0.000474	0.001667	0.001722	0.000912
293	0.000462	0.001525	0.001549	0.000727
298	0.000450	0.001400	0.001409	0.000595
303	0.000439	0.001289	0.001295	0.000499
308	0.000428	0.001189	0.001202	0.000427
313	0.000419	0.001100	0.001126	0.000374
318	0.000409	0.001020	0.001063	0.000334
323	0.000400	0.000949	0.001012	0.000303

Chapter 8. Conclusions

In this dissertation, I have developed theoretical approaches for describing capsid self-assembly, genome encapsidation and maturation, and nucleocapsid envelopment of hepatitis virus (HBV). The theoretical results provide novel insights into the life cycle of HBV replication and fresh ideas toward developing new antiviral strategy.

I have conducted a systematic investigation of the different stages of the viral structure using various molecular modeling techniques spanning different scales from macroscopic range to atomic level. A combination of the multi-scale methods allows us to reveal the thermodynamic basis for the capsid assembly, genome packaging, dynamic capsid structure, and capsid envelopment. Specifically, we use coarse-grained models to represent key viral components and intermolecular interactions under realistic intracellular environment. I demonstrate that the coarse-grained models provide an accurate description of the thermodynamic stability of HBV capsids at different stages of replication. Different from an earlier model for HBV stability, the new model accounts for genome contents, the semi-permeability nature of the capsid shell, and effects of salt ions explicitly. The theoretical results correspond well with experimental data and help to explain stability modulation of HBV capsids by the solution conditions such as the ionic strength. The theoretical model provides valuable information on how to regulate virus formation by controlling capsid shell assembly and genome encapsidation.

In comparison to alternative theoretical methods for investigating the equilibrium properties of HBV capsids at different stages of the replication, the density functional

theory (DFT) has the advantage of numerical efficiency. In particular, it provides direct information on thermodynamic properties that reflect various non-bonded intermolecular interactions beyond conventional mean-field approximations. By extensive comparison with experimental data, I find that the DFT is able to quantify the structure and thermodynamic properties of empty or genome containing nucleocapsids.

The encapsidated HBV genome and the flexible tails of the core proteins undergo structural transitions during maturation of the nucleocapsid. The structure change is sensitive to phosphorylation of serine residues in the flexible tails of the core proteins. We predict the optimal RNA lengths in quantitative agreement with experiments for both the WT virus and mutants with C-terminally truncated core proteins. While genome packaging is dominated by nonspecific electrostatic interaction, we find substantial contributions from other factors including the inhomogeneous charge distribution and molecular excluded volume effects.

The surface structure of a nucleocapsid is intimately related to the signaling of its maturation for viral amplification and envelopment with surface proteins for egression. To unravel the molecular mechanism behind the surface-core protein recognition, I combine MD simulations and molecular docking to examine capsid-envelope association due to the interaction of a large surface protein and a capsid core protein. The simulation methods allow me to identify the specific binding sites on each protein and elucidate the capsid envelopment mechanism in terms of the location of CTD tails. Based on the numerical results, I provide a probable scenario about how the viral maturation is dependent on HBV nucleocapsid formation.

The molecular models for various stages HBV replication established in this thesis will contribute to developing novel biophysical schemes to terminate the HBV life cycle. For example, we have demonstrated that the stability of HBV capsids requires a subtle balance of hydrophobic and electrostatic interactions. New therapeutic strategies may be developed based on their stability at various solution conditions. Also the thermodynamic model for RNA encapsidation could facilitate the design of new experimental procedure for *in vitro* reconstitution of RNA-containing nucleocapsids. In addition, we identify new biophysical schemes to regulate HBV replication by controlling the charge statuses and the positions of CTD tails. The nucleocapsid formation and genome packaging depend on the charge density of CTDs. As a result, charge neutralization on the CTDs may block pgRNA encapsidation and its reverse transcription in the nucleocapsid. Moreover, we study the distributions of amino-acid residues in CTD tails at different replication stages and identify the possible signatures of a mature nucleocapsid that differentiates it from an immature one. We expect that the CTD tails provide a good target for modulating the post-transcriptional process of the viral particles. For the envelopment of the mature nucleocapsids, we have identified the specific binding site in the capsid proteins that mediates their association with the large surface proteins in the envelope shell. We believe that the particular interactive sequence in the surface protein renders a possible candidate for the design of polypeptide drugs as an inhibiter of the viral envelopment.

In conclusion, the fundamental research presented in this thesis allows us not only to attain a deeper understanding of the basic principles of viral replication but also to

identify unconventional targets for future therapeutic developments. The theoretical analyses have been validated with extensive experimental results and lead to insightful predictions that could be tested with new biochemical experiments. It is expected that such synergic efforts from theoretical and experimental investigations will facilitate most efficient discovery of powerful new therapeutic treatments against HBV infection.

References

1. Ganem, D.; Prince, A. M., Mechanisms of disease: Hepatitis B virus infection - Natural history and clinical consequences. *New England Journal of Medicine* **2004,** 350, (11), 1118-1129.

2. Seeger, C.; Mason, W. S., Hepatitis B virus biology. *Microbiology and Molecular Biology Reviews* **2000,** 64, (1), 51.

3. Locarnini, S.; Mason, W. S., Cellular and virological mechanisms of HBV drug resistance. *Journal of Hepatology* **2006,** 44, (2), 422-431.

4. Kenney, J. M.; Vonbonsdorff, C. H.; Nassal, M.; Fuller, S. D., Evolutionary conservation in the hepatitis-B virus core structure - comparison of human and duck cores. *Structure* **1995,** 3, (10), 1009-1019.

5. Zlotnick, A.; Cheng, N.; Conway, J. F.; Booy, F. P.; Steven, A. C.; Stahl, S. J.; Wingfield, P. T., Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry* **1996,** 35, (23), 7412-7421.

6. Zlotnick, A.; Cheng, N.; Stahl, S. J.; Conway, J. F.; Steven, A. C.; Wingfield, P. T., Localization of the C terminus of the assembly domain of hepatitis B virus capsid protein: Implications for morphogenesis and organization of encapsidated RNA. *Proceedings of the National Academy of Sciences of the United States of America* **1997,** 94, (18), 9556-9561.

7. Nassal, M., 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. *Journal of Virology* **1992,** 66, (7), 4107-4116

8. Petit, M. A.; Pillot, J., HBc and HBe antigenicity and DNA-binding activity of major core protein P22 in hepatitis B virus core particles isolated from the cytoplasm of human liver cells. *Journal of Virology* **1985,** 53, (2), 543-551

9. Wingfield, P. T.; Stahl, S. J.; Williams, R. W.; Steven, A. C., Hepatitis core antigen produced in Escherichia coli: subunit composition, conformational analysis, and in vitro capsid assembly. *Biochemistry* **1995,** 34, (15), 4919-4932

10. Porterfield, J. Z.; Dhason, M. S.; Loeb, D. D.; Nassal, M.; Stray, S. J.; Zlotnick, A., Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. *Journal of Virology* **2010,** 84, (14), 7174-7184.

11. Lan, Y. T.; Li, J.; Liao, W. Y.; Ou, J. H., Roles of the three major phosphorylation sites of hepatitis B virus core protein in viral replication. *Virology* **1999,** 259, (2), 342-348.

12. Lewellyn, E. B.; Loeb, D. D., Serine phosphoacceptor sites within the core protein of hepatitis B virus contribute to genome replication pleiotropically. *Plos One* **2011,** 6, (2), e17202.

13. Perlman, D. H.; Berg, E. A.; O'Connor, P. B.; Costello, C. E.; Hu, J. M., Reverse transcription-associated dephosphorylation of hepadnavirus nucleocapsids. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, (25), 9020-9025.

14. Basagoudanavar, S. H.; Perlman, D. H.; Hu, J., Regulation of hepadnavirus reverse transcription by dynamic nucleocapsid phosphorylation. *Journal of Virology* **2007,** 81, (4), 1641-1649.

15. Gazina, E. V.; Fielding, J. E.; Lin, B.; Anderson, D. A., Core protein phosphorylation modulates pregenomic RNA encapsidation to different extents in human and duck hepatitis B viruses. *Journal of Virology* **2000,** 74, (10), 4721-4728.

16. Kock, J.; Nassal, M.; Deres, K.; Blum, H. E.; von Weizsacker, F., Hepatitis B virus nucleocapsids formed by carboxy-terminally mutated core proteins contain spliced viral genomes but lack full-size DNA. *Journal of Virology* **2004,** 78, (24), 13812-13818.

17. Melegari, M.; Wolf, S. K.; Schneider, R. J., Hepatitis B virus DNA replication is coordinated by core protein serine phosphorylation and HBx expression. *Journal of Virology* **2005,** 79, (15), 9810-9820.

18. Chua, P. K.; Tang, F.-M.; Huang, J.-Y.; Suen, C.-S.; Shih, C., Testing the Balanced Electrostatic Interaction Hypothesis of Hepatitis B Virus DNA Synthesis by Using an In Vivo Charge Rebalance Approach. *Journal of Virology* **2010,** 84, (5), 2340-2351.

19. Kann, M.; Sodeik, B.; Vlachou, A.; Gerlich, W. H.; Helenius, A., Phosphorylation-dependent binding of hepatitis B virus core particles to the nuclear pore complex. *Journal of Cell Biology* **1999,** 145, (1), 45-55.

20. Schlicht, H. J.; Bartenschlager, R.; Schaller, H., The duck hepatitis B virus core protein contains a highly phosphorylated C terminus that is essential for replication but not for RNA packaging. *Journal of Virology* **1989,** 63, (7), 2995-3000

21. Seifer, M.; Standring, D. N., A protease-sensitive hinge linking the two domains of the hepatitis B virus core protein is exposed on the viral capsid surface. *Journal of Virology* **1994,** 68, (9), 5548-5555

22. Rabe, B.; Vlachou, A.; Pante, N.; Helenius, A.; Kann, M., Nuclear import of hepatitis B virus capsids and release of the viral genome. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** 100, (17), 9849-9854.

23. Guo, H.; Mao, R.; Block, T. M.; Guo, J.-T., Production and function of the cytoplasmic deproteinized relaxed circular DNA of hepadnaviruses. *Journal of Virology* **2010,** 84, (1), 387-396.

24. Wang, J. C. Y.; Dhason, M. S.; Zlotnick, A., Structural organization of pregenomic RNA and the carboxy-terminal domain of the capsid protein of hepatitis B virus. *Plos Pathogens* **2012,** 8, (9), e1002919.

25. Bartenschlager, R.; Schaller, H., Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *Embo Journal* **1992,** 11, (9), 3413-3420.

26. Watts, N. R.; Conway, J. F.; Cheng, N. Q.; Stahl, S. J.; Belnap, D. M.; Steven, A. C.; Wingfield, P. T., The morphogenic linker peptide of HBV capsid protein forms a mobile array on the interior surface. *Embo Journal* **2002,** 21, (5), 876-884.

27. Bruss, V., Hepatitis B virus morphogenesis. *World Journal of Gastroenterology* **2007,** 13, (1), 65-73.

28. Schmitz, A.; Schwarz, A.; Foss, M.; Zhou, L.; Rabe, B.; Hoellenriegel, J.; Stoeber, M.; Pante, N.; Kann, M., Nucleoporin 153 arrests the nuclear import of hepatitis B virus capsids in the nuclear basket. *Plos Pathogens* **2010,** 6, (1), e1000741.

29. Lewellyn, E. B.; Loeb, D. D., The arginine clusters of the carboxy-terminal domain of the core protein of hepatitis B virus make pleiotropic contributions to genome replication. *Journal of Virology* **2011,** 85, (3), 1298-1309.

30. Bruss, V.; Ganem, D., The role of envelope proteins in hepatitis-B virus assembly. *Proceedings of the National Academy of Sciences of the United States of America* **1991,** 88, (3), 1059-1063.

31. Heermann, K. H.; Goldmann, U.; Schwartz, W.; Seyffarth, T.; Baumgarten, H.; Gerlich, W. H., Large surface-proteins of hepatitis-b virus containing the pre-s sequence. *Journal of Virology* **1984,** 52, (2), 396-402.

32. Ganem, D., ASSEMBLY OF HEPADNAVIRAL VIRIONS AND SUBVIRAL PARTICLES. *Current Topics in Microbiology and Immunology* **1991,** 168, 61-83.

33. Gilbert, R. J. C.; Beales, L.; Blond, D.; Simon, M. N.; Lin, B. Y.; Chisari, F. V.; Stuart, D. I.; Rowlands, D. J., Hepatitis B small surface antigen particles are octahedral. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, (41), 14783-14788.

34. McAleer, W. J.; Buynak, E. B.; Maigetter, R. Z.; Wampler, D. E.; Miller, W. J.; Hilleman, M. R., Human hepatitis-B vaccine from recombinant yeast. *Nature* **1984,** 307, (5947), 178-180.

35. Huovila, A. P. J.; Eder, A. M.; Fuller, S. D., Hepatitis-B surface-antigen assemblies in a post-er, pre-golgi compartment. *Journal of Cell Biology* **1992,** 118, (6), 1305-1320.

36. Bruss, V., A short linear sequence in the pre-S domain of the large hepatitis B virus envelope protein required for virion formation. *Journal of Virology* **1997,** 71, (12), 9350-9357.

37. Glebe, D.; Urban, S., Viral and cellular determinants involved in hepadnaviral entry. *World Journal of Gastroenterology* **2007,** 13, (1), 22-38.

38. Bruss, V.; Vieluf, K., Functions of the internal pre-s domain of the large surface protein in hepatitis-B virus particle morphogenesis. *Journal of Virology* **1995,** 69, (11), 6652-6657.

39. Poisson, F.; Severac, A.; Hourioux, C.; Goudeau, A.; Roingeard, P., Both pre-S1 and S domains of hepatitis B virus envelope proteins interact with the core particle. *Virology* **1997,** 228, (1), 115-120.

40. Bottcher, B.; Tsuji, N.; Takahashi, H.; Dyson, M. R.; Zhao, S.; Crowther, R. A.; Murray, K., Peptides that block hepatitis B virus assembly: analysis by cryomicroscopy, mutagenesis and transfection. *Embo Journal* **1998,** 17, (23), 6839-6845.

41. Ponsel, D.; Bruss, V., Mapping of amino acid side chains on the surface of hepatitis B virus capsids required for envelopment and virion formation. *Journal of Virology* **2003,** 77, (1), 416-422.

42. Dryden, K. A.; Wieland, S. F.; Whitten-Bauer, C.; Gerin, J. L.; Chisari, F. V.; Yeager, M., Native hepatitis B virions and capsids visualized by electron cryornicroscopy. *Molecular Cell* **2006,** 22, (6), 843-850.

43. Seitz, S.; Urban, S.; Antoni, C.; Boettcher, B., Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions. *Embo Journal* **2007,** 26, (18), 4160-4167.

44. Weiser, B.; Ganem, D.; Seeger, C.; Varmus, H. E., Closed circular viral-DNA and asymmetrical heterogeneous forms in livers from animals infected with ground-squirrel hepatitis-virus. *Journal of Virology* **1983,** 48, (1), 1-9.

45. Mason, W. S.; Aldrich, C.; Summers, J.; Taylor, J. M., Asymmetric replication of duck hepatitis-B virus-DNA in liver-cells - free minus-strand DNA. *Proceedings of the*

National Academy of Sciences of the United States of America-Biological Sciences **1982,** 79, (13), 3997-4001.

46. Summers, J.; Mason, W. S., Replication of the genome of a hepatitis-B-like virus by reverse transcription of an RNA intermediate. *Cell* **1982,** 29, (2), 403-415.

47. Yu, M. S.; Summers, J., A domain of the hepadnavirus capsid protein is specifically required for DNA maturation and virus assembly. *Journal of Virology* **1991,** 65, (5), 2511-2517.

48. Wei, Y.; Tavis, J. E.; Ganem, D., Relationship between viral DNA synthesis and virion envelopment in hepatitis B viruses. *Journal of Virology* **1996,** 70, (9), 6455-6458.

49. Gerelsaikhan, T.; Tavis, J. E.; Bruss, V., Hepatitis B virus nucleocapsid envelopment does not occur without genomic DNA synthesis. *Journal of Virology* **1996,** 70, (7), 4269-4274.

50. Roseman, A. M.; Berriman, J. A.; Wynne, S. A.; Butler, P. J. G.; Crowther, R. A., A structural model for maturation of the hepatitis B virus core. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, (44), 15821-15826.

51. Yu, M. S.; Summers, J., Multiple functions of capsid protein-phosphorylation in duck hepatitis-B virus-replication. *Journal of Virology* **1994,** 68, (7), 4341-4348.

52. Yu, M. S.; Summers, J., Phosphorylation of the duck hepatitis-B virus capsid protein associated with conformational-changes in the C-terminus. *Journal of Virology* **1994,** 68, (5), 2965-2969.

53. Yuan, T. T. T.; Sahu, G. K.; Whitehead, W. E.; Greenberg, R.; Shih, C., The mechanism of an immature secretion phenotype of a highly frequent naturally occurring missense mutation at codon 97 of human hepatitis B virus core antigen. *Journal of Virology* **1999,** 73, (7), 5731-5740.

54. Yuan, T. T. T.; Shih, C., A frequent, naturally occurring mutation (P130T) of human hepatitis B virus core antigen is compensatory for immature secretion phenotype of another frequent variant (I97L). *Journal of Virology* **2000,** 74, (10), 4929-4932.

55. Chu, C. K.; Ma, T. W.; Shanhmuganathan, K.; Wang, C. G.; Xiang, Y. J.; Pai, S. B.; Yao, G. Q.; Sommadossi, J. P.; Cheng, Y. C., Use of 2'-fluoro-5-methyl-beta-l-arabinofuranosyluracil as a novel antiviral agent for hepatitis-B virus and Epstein-Barr-virus. *Antimicrobial Agents and Chemotherapy* **1995,** 39, (4), 979-981.

56. de Man, R. A.; Wolters, L. M. M.; Nevens, F.; Chua, D.; Sherman, M.; Lai, C. L.; Gadano, A.; Lee, Y.; Mazzotta, F.; Thomas, N.; DeHertogh, D., Safety and efficacy of

oral entecavir given for 28 days in patients with chronic hepatitis B virus infection. *Hepatology* **2001,** 34, (3), 578-582.

57. Dienstag, J. L.; Perrillo, R. P.; Schiff, E. R.; Bartholomew, M.; Vicary, C.; Rubin, M., A preliminary trial of lamivudine for chronic hepatitis-B infection. *New England Journal of Medicine* **1995,** 333, (25), 1657-1661.

58. Lai, C. L.; Lim, S. G.; Brown, N. A.; Zhou, X. J.; Lloyd, D. M.; Lee, Y. M.; Yuen, M. F.; Chao, G. C.; Myers, M. W., A dose-finding study of once-daily oral telbivudine in HBeAg-positive patients with chronic hepatitis B virus infection. *Hepatology* **2004,** 40, (3), 719-726.

59. Chong, Y. H.; Chu, C. K., Understanding the unique mechanism of L-FMAU (clevudine) against hepatitis B virus: Molecular dynamics studies. *Bioorganic & Medicinal Chemistry Letters* **2002,** 12, (23), 3459-3462.

60. Yadav, V.; Chu, C. K., Molecular mechanisms of adefovir sensitivity and resistance in HBV polymerase mutants: a molecular dynamics study. *Bioorganic & Medicinal Chemistry Letters* **2004,** 14, (16), 4313-4317.

61. Yu, N.; Hagan, M. F., Simulations of HIV capsid protein dimerization reveal the effect of chemistry and topography on the mechanism of hydrophobic protein association. *Biophysical Journal* **2012,** 103, (6), 1363-1369.

62. Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K., Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* **2006,** 14, (3), 437-449.

63. Kegel, W. K.; van der Schoot, P., Competing hydrophobic and screened-Coulomb interactions in hepatitis B virus capsid assembly. *Biophysical Journal* **2004,** 86, (6), 3905-3913.

64. Joshi, H.; Ortoleva, P., Thermal, chemical and mechanical stability of HPV L1 VLP vaccines: a multiscale all-atom simulation study. *Biophysical Journal* **2011,** 100, (3), 196-196.

65. Joshi, H.; Cheluvaraja, S.; Somogyi, E.; Brown, D. R.; Ortoleva, P., A molecular dynamics study of loop fluctuation in human papillomavirus type 16 virus-like particles: A possible indicator of immunogenicity. *Vaccine* **2011,** 29, (51), 9423-9430.

66. Lidon-Moya, M. C.; Barrera, F. N.; Bueno, M.; Perez-Jimenez, R.; Sancho, J.; Mateu, M. G.; Neira, J. L., An extensive thermodynamic characterization of the dimerization domain of the HIV-1 capsid protein. *Protein Science* **2005,** 14, (9), 2387-2404.

67. del Alamo, M.; Neira, J. L.; Mateu, M. G., Thermodynamic dissection of a low affinity protein- protein interface involved in human immunodeficiency virus assembly. *Journal of Biological Chemistry* **2003,** 278, (30), 27923-27929.

68. del Alamo, M.; Mateu, M. G., Electrostatic repulsion, compensatory mutations, and long-range non-additive effects at the dimerization interface of the HIV capsid protein. *Journal of Molecular Biology* **2005,** 345, (4), 893-906.

69. Ceres, P.; Zlotnick, A., Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **2002,** 41, (39), 11525-11531.

70. Rollins, G. C.; Petrov, A. S.; Harvey, S. C., The role of DNA twist in the packaging of viral genomes. *Biophysical Journal* **2008,** 94, (5), L38-L40.

71. Petrov, A. S.; Locker, C. R.; Harvey, S. C., Characterization of DNA conformation inside bacterial viruses. *Physical Review E* **2009,** 80, (2).

72. Petrov, A. S.; Lim-Hing, K.; Harvey, S. C., Packaging of DNA by bacteriophage Epsilon15: Structure, forces, and thermodynamics. *Structure* **2007,** 15, (7), 807-812.

73. Petrov, A. S.; Harvey, S. C., Packaging double-helical DNA into viral capsids: Structures, forces, and energetics. *Biophysical Journal* **2008,** 95, (2), 497-502.

74. Forrey, C.; Muthukumar, M., Langevin dynamics simulations of genome packing in bacteriophage. *Biophysical Journal* **2006,** 91, (1), 25-41.

75. Leforestier, A.; Livolant, F., The bacteriophage genome undergoes a succession of intracapsid phase transitions upon DNA ejection. *Journal of Molecular Biology* **2010,** 396, (2), 384-395.

76. Leforestier, A.; Livolant, F., Structure of toroidal DNA collapsed inside the phage capsid. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** 106, (23), 9157-9162.

77. Arkhipov, A.; Freddolino, P. L.; Schulten, K., Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure* **2006,** 14, (12), 1767-1777.

78. Tama, F.; Brooks, C. L., Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *Journal of Molecular Biology* **2005,** 345, (2), 299-314.

79. Roos, W. H.; Gertsman, I.; May, E. R.; Brooks, C. L., III; Johnson, J. E.; Wuite, G. J. L., Mechanics of bacteriophage maturation. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** 109, (7), 2342-2347.

80. Miao, Y.; Johnson, J. E.; Ortoleva, P. J., All-atom multiscale simulation of cowpea chlorotic mottle virus capsid swelling. *Journal of Physical Chemistry B* **2010,** 114, (34), 11181-11195.

81. Zlotnick, A.; Johnson, J. M.; Wingfield, P. W.; Stahl, S. J.; Endres, D., A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry* **1999,** 38, (44), 14644-14652.

82. Zlotnick, A., Theoretical aspects of virus capsid assembly. *Journal of Molecular Recognition* **2005,** 18, (6), 479-490.

83. Stray, S. J.; Bourne, C. R.; Punna, S.; Lewis, W. G.; Finn, M. G.; Zlotnick, A., A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, (23), 8138-8143.

84. Bourne, C. R.; Finn, M. G.; Zlotnick, A., Global structural changes in hepatitis B virus capsids induced by the assembly effector HAP1. *Journal of Virology* **2006,** 80, (22), 11055-11061.

85. Katen, S. P.; Chirapu, S. R.; Finn, M. G.; Zlotnick, A., Trapping of hepatitis B virus capsid assembly intermediates by phenylpropenamide assembly accelerators. *Acs Chemical Biology* **2010,** 5, (12), 1125-1136.

86. Wu, J., *Molecular Thermodynamics of Complex Systems*. Springer: Berlin, 2009; Vol. 131, p 1-73.

87. Rowlinson, J. S., Translation of J. D. van der Waals' "The thermodynamik theory of capillarity under the hypothesis of a continuous variation of density". *Journal of Statistical Physics* **1979,** 20, (2), 197-200.

88. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Physical Review* **1964,** 136, (3B), B864-B871.

89. Wu, J.; Li, Z., Density-Functional Theory for Complex Fluids. *Annual Review of Physical Chemistry* **2007,** 58, (1), 85-112.

90. Yu, Y.-X.; Wu, J., Density functional theory for inhomogeneous mixtures of polymeric fluids. *Journal of Chemical Physics* **2002,** 117, (5), 2368-2376.

91. Yu, Y. X.; Wu, J. Z., Structures of hard sphere fluids from a modified fundamental measure theory. *Journal of Chemical Physics* **2002,** 117, (22), 10156-10164.

92. Roth, R.; Evans, R.; Lang, A.; Kahl, G., Fundamental measure theory for hard-sphere mixtures revisited: the White Bear version. *Journal of Physics-Condensed Matter* **2002,** 14, (46), 12063-12078.

93. Rosenfeld, Y., Free-energy model for the inhomogeneous hard-sphere fluid mixture and density-functional theory of freezing. *Physical Review Letters* **1989,** 63, (9), 980-983.

94. Mansoori, G. A.; Carnahan, N. F.; Starling, K. E.; T. W. Leland, J., Equilibrium Thermodynamic Properties of the Mixture of Hard Spheres. *Journal of Chemical Physics* **1971,** 54, (4), 1523-1525.

95. Wertheim, M. S., Fluids with highly directional attractive forces. I. Statistical thermodynamics. *Journal of Statistical Physics* **1984,** 35, (1), 19-34.

96. Yu, Y. X.; Wu, J. Z., Density functional theory for inhomogeneous mixtures of polymeric fluids. *Journal of Chemical Physics* **2002,** 117, (5), 2368-2376.

97. Blum, L., Mean spherical model for asymmetric electrolytes .1. method of solution. *Molecular Physics* **1975,** 30, (5), 1529-1535.

98. Tomasi, J.; Persico, M., Molecular-interactions in solution - an overview of methods based on continuous distributions of the solvent. *Chemical Reviews* **1994,** 94, (7), 2027-2094.

99. Yu, Y. X.; Wu, J. Z.; Gao, G. H., Density-functional theory of spherical electric double layers and zeta potentials of colloidal particles in restricted-primitive-model electrolyte solutions. *Journal of Chemical Physics* **2004,** 120, (15), 7223-7233.

100. Li, Z. D.; Wu, J. Z., Density-functional theory for the structures and thermodynamic properties of highly asymmetric electrolyte and neutral component mixtures. *Physical Review E* **2004,** 70, (3), 031109.

101. Li, Z.; Wu, J., Density Functional Theory for Planar Electric Double Layers: Closing the Gap between Simple and Polyelectrolytes. *Journal of Physical Chemistry B* **2006,** 110, (14), 7473-7484.

102. Yu, Y.-X.; Wu, J.; Gao, G.-H., Density-functional theory of spherical electric double layers and zeta potentials of colloidal particles in restricted-primitive-model electrolyte solutions. *Journal of Chemical Physics* **2004,** 120, (15), 7223-7233.

103. Wertheim, M. S., Exact Solution of the Percus-Yevick Integral Equation for Hard Spheres. *Physical Review Letters* **1963,** 10, (8), 321-323.

104. Jiang, T.; Li, Z. D.; Wu, J. Z., Structure and swelling of grafted polyelectrolytes: Predictions from a nonlocal density functional theory. *Macromolecules* **2007,** 40, (2), 334-343.

105. Li, Z. D.; Wu, J. Z., Density-functional theory for polyelectrolytes near an oppositely-charged surface. *Physical Review Letters* **2006,** 96, 048302.

106. Li, Z. D.; Wu, J. Z., Density functional theory for planar electric double layers: Closing the gap between simple and polyelectrolytes. *Journal of Physical Chemistry B* **2006,** 110, 7473-7484.

107. Wu, J. Z.; Jiang, T., Ionic effects in collapse of polyelectrolyte brushes. *Journal of Physical Chemistry B* **2008,** 112, (26), 7713-7720.

108. Jiang, T.; Wu, J. Z., Self-organization of multivalent counterions in polyelectrolyte brushes. *Journal of Chemical Physics* **2008,** 129, (8), 10.1063/1.2966359.

109. Wang, L.; Liang, H. J.; Wu, J. Z., Electrostatic origins of polyelectrolyte adsorption: Theory and Monte Carlo simulations. *Journal of Chemical Physics* **2010,** 133, (4), 10.1063/1.3463426.

110. Jiang, T.; Wang, Z. G.; Wu, J. Z., Electrostatic Regulation of Genome Packaging in Human Hepatitis B Virus. *Biophysical Journal* **2009,** 96, (8), 3065-3073.

111. Li, Z. D.; Wu, J. Z.; Wang, Z. G., Osmotic pressure and packaging structure of caged DNA. *Biophysical Journal* **2008,** 94, (3), 737-746.

112. Meng, D.; Hjelm, R. P.; Hu, J.; Wu, J., A theoretical model for the dynamic structure of hepatitis B nucleocapsid. *Biophysical Journal* **2011,** 101, (10), 2476-2484.

113. Woodward, C. E.; Forsman, J., Density functional theory for flexible and semiflexible polymers of infinite length. *Physical Review E* **2006,** 74, (1).

114. Mateu, M. G., Assembly, stability and dynamics of virus capsids. *Archives of Biochemistry and Biophysics* **2013,** 531, (1-2), 65-79.

115. Katen, S.; Zlotnick, A., The thermodynamics of virus capsid assembly. *Methods in Enzymology* **2009,** 455, 395-417.

116. Belyi, V. A.; Muthukumar, M., Electrostatic origin of the genome packing in viruses. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** 103, 17174-17178

117. Bruinsma, R. F.; Gelbart, W. M.; Reguera, D.; Rudnick, J.; Zandi, R., Viral self-assembly as a thermodynamic process. *Physical Review Letters* **2003,** 90, 248101.

118. Kegel, W. K.; van der Schoot, P., Competing hydrophobic and screened-Coulomb interactions in hepatitis B virus capsid assembly. *Biophysical Journal* **2004,** 86, 3905-3913.

119. Siber, A.; Podgornik, R., Nonspecific interactions in spontaneous assembly of empty versus functional single-stranded RNA viruses. *Physical Review E* **2008,** 78, 051915.

120. Hagan, M. F., Modeling viral capsid assembly. *Advances in Chemical Physics* **2013,** 155.

121. Siber, A.; Bozic, A. L.; Podgornik, R., Energies and pressures in viruses: contribution of nonspecific electrostatic interactions. *Physical Chemistry Chemical Physics* **2012,** 14, (11), 3746-3765.

122. Zhang, R.; Wernersson, E.; Linse, P., Icosahedral capsid formation by capsomer subunits and a semiflexible polyion. *Rsc Advances* **2013,** 3, (47), 25258-25267.

123. Liang, T. J., Hepatitis B: The virus and disease. *Hepatology* **2009,** 49, (5), S13-S21.

124. Beck, J.; Nassal, M., Hepatitis B virus replication. *World Journal of Gastroenterology* **2007,** 13, (1), 48-64.

125. Zlotnick, A., Viruses and the physics of soft condensed matter. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** 101, (44), 15549-15550.

126. Wingfield, P. T.; Stahl, S. J.; Williams, R. W.; Steven, A. C., Hepatitis core antigen produced in Escherichia coli: subunit composition, conformational analysis, and in vitro capsid assembly. *Biochemistry* **1995,** 34, 4919-4932

127. Bartenschlager, R.; Schaller, H., Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *Embo Journal* **1992,** 11, 3413-3420

128. Hirsch, R. C.; Lavine, J. E.; Chang, L. J.; Varmus, H. E.; Ganem, D., Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. *Nature* **1990,** 344, 552-555

129. Nassal, M., Hepatitis B viruses: Reverse transcription a different way. *Virus Research* **2008,** 134, (1-2), 235-249.

130. Liu, N.; Tian, R.; Loeb, D. D., Base pairing among three cis-acting sequences contributes to template switching during hepadnavirus reverse transcription. *Proceedings*

of the National Academy of Sciences of the United States of America **2003,** 100, (4), 1984-1989.

131. Lewellyn, E. B.; Loeb, D. D., Base pairing between cis-acting sequences contributes to template switching during plus-strand DNA synthesis in human hepatitis B virus. *Journal of Virology* **2007,** 81, (12), 6207-6215.

132. Ning, X. J.; Nguyen, D.; Mentzer, L.; Adams, C.; Lee, H.; Ashley, R.; Hafenstein, S.; Hu, J. M., Secretion of genome-free hepatitis B virus - single strand blocking model for virion morphogenesis of Para-retrovirus. *Plos Pathogens* **2011,** 7, (9), Artn E1002255.

133. van der Schoot, P.; Bruinsma, R., Electrostatics and the assembly of an RNA virus. *Physical Review E* **2005,** 71, 061928.

134. Nassal, M., 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. *Journal of Virology* **1992,** 66, 4107-4116

135. Perlman, D. H.; Berg, E. A.; O'Connor, P. B.; Costello, C. E.; Hu, J. M., Reverse transcription-associated dephosphorylation of hepadnavirus nucleocapsids. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** 102, 9020-9025.

136. Ceres, P.; Zlotnick, A., Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **2002,** 41, 11525-11531

137. Kegel, W. K.; van der Schoot, P., Physical regulation of the self-assembly of tobacco mosaic virus coat protein. *Biophysical Journal* **2006,** 91, (4), 1501-1512.

138. Siber, A.; Podgornik, R., Role of electrostatic interactions in the assembly of empty spherical viral capsids. *Physical Review E* **2007,** 76, (6).

139. Stray, S. J.; Ceres, P.; Zlotnick, A., Zinc ions trigger conformational change and oligomerization of hepatitis B virus capsid protein. *Biochemistry* **2004,** 43, (31), 9989-9998.

140. Alexander, C. G.; Jurgens, M. C.; Shepherd, D. A.; Freund, S. M. V.; Ashcroft, A. E.; Ferguson, N., Thermodynamic origins of protein folding, allostery, and capsid formation in the human hepatitis B virus core protein. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** 110, (30), E2782-E2791.

141. Ceres, P.; Stray, S. J.; Zlotnick, A., Hepatitis B virus capsid assembly is enhanced by naturally occurring mutation F97L. *Journal of Virology* **2004,** 78, (17), 9538-9543.

142. Newman, M.; Suk, F. M.; Cajimat, M.; Chua, P. M.; Shih, C., Stability and morphology comparisons of self-assembled virus-like particles from wild-type and mutant human hepatitis B virus capsid proteins. *Journal of Virology* **2003,** 77, (24), 12950-12960.

143. Packianathan, C.; Katen, S. P.; Dann, C. E.; Zlotnick, A., Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly. *Journal of Virology* **2010,** 84, (3), 1607-1615.

144. Newman, M.; Chua, P. K.; Tang, F. M.; Su, P. Y.; Shih, C., Testing an electrostatic interaction hypothesis of hepatitis B virus capsid stability by using an in vitro capsid disassembly/reassembly system. *Journal of Virology* **2009,** 83, (20), 10616-10626.

145. Porterfield, J. Z.; Dhason, M. S.; Loeb, D. D.; Nassal, M.; Stray, S. J.; Zlotnick, A., Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. *Journal of Virology* **2010,** 84, 7174-7184

146. Cui, X.; Ludgate, L.; Ning, X. J.; Hu, J. M., Maturation-associated destabilization of hepatitis B virus nucleocapsid. *Journal of Virology* **2013,** 87, (21), 11494-11503.

147. Dhason, M. S.; Wang, J. C. Y.; Hagan, M. F.; Zlotnick, A., Differential assembly of hepatitis B virus core protein on single- and double-stranded nucleic acid suggest the dsDNA-filled core is spring-loaded. *Virology* **2012,** 430, (1), 20-29.

148. Zlotnick, A.; Ceres, P.; Singh, S.; Johnson, J. M., A small molecule inhibits and misdirects assembly of hepatitis B virus capsids. *Journal of Virology* **2002,** 76, 4848-4854

149. Crowther, R. A.; Kiselev, N. A.; Bottcher, B.; Berriman, J. A.; Borisova, G. P.; Ose, V.; Pumpens, P., Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell* **1994,** 77, 943-950.

150. de Vries, R.; Stuart, M. C., Theory and simulations of macroion complexation. *Current Opinion in Colloid & Interface Science* **2006,** 11, (5), 295-301.

151. Yu, X.; Jin, L.; Jih, J.; Shih, C.; Zhou, Z. H., 3.5A cryoEM structure of hepatitis B virus core assembled from full-length core protein. *PLoS One* **2013,** 8, (9), e69729.

152. Crick, F. H. C.; Watson, J. D., Structure of small viruses. *Nature* **1956,** 177, (4506), 473-475.

153. Wynne, S. A.; Crowther, R. A.; Leslie, A. G. W., The crystal structure of the human hepatitis B virus capsid. *Molecular Cell* **1999,** 3, 771-780.

154. Baldwin, R. L., The new view of hydrophobic free energy. *Febs Letters* **2013,** 587, (8), 1062-1066.

155. Wu, J.; Prausnitz, J. M., Pairwise-additive hydrophobic effect for alkanes in water. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** 105, 9512-9515

156. Hagan, M. F., Controlling viral capsid assembly with templating. *Physical Review E* **2008,** 77, 051904.

157. Israelachvili, J. N., *Intermolecular and surface forces*. 3rd ed.; Academic Press: Burlington, MA, 2011.

158. Farnum, M.; Zukoski, C., Effect of glycerol on the interactions and solubility of bovine pancreatic trypsin inhibitor. *Biophysical Journal* **1999,** 76, (5), 2716-2726.

159. Weast, R. C., Handbook of chemistry and physics. *CRC Press, Boca Ranton, FL* **1984**.

160. Perlmutter, J. D.; Qiao, C.; Hagan, M. F., Viral genome structures are optimal for capsid assembly. *eLife* **2013,** 2, e00632.

161. van der Schoot, P.; Zandi, R., Impact of the topology of viral RNAs on their encapsulation by virus coat proteins. *Journal of Biological Physics* **2013,** 39, 289-299.

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

163. Prinsen, P.; van der Schoot, P.; Gelbart, W. M.; Knobler, C. M., Multishell structures of virus coat proteins. *Journal of Physical Chemistry B* **2010,** 114, 5522-5533.

164. Netz, R. R.; Andelman, D., Neutral and charged polymers at interfaces. *Physics Reports-Review Section of Physics Letters* **2003,** 380, 1-95.

165. Zandi, R.; van der Schoot, P., Size regulation of ss-RNA viruses. *Biophysical Journal* **2009,** 96, 9-20

166. Overbeek, J. T. G.; Voorn, M. J., Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *J Cell and Comp Physiol* **1957,** 49, 7-26

167. Yoffe, A. M.; Prinsen, P.; Gopal, A.; Knobler, C. M.; Gelbart, W. M.; Ben-Shaul, A., Predicting the sizes of large RNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** 105, 16153-16158.

168. Hyeon, C.; Dima, R. I.; Thirumalai, D., Size, shape, and flexibility of RNA structures. *Journal of Chemical Physics* **2006,** 125, 194905.

169. Carrillo, J.-M. Y.; Dobrynin, A. V., Polyelectrolytes in salt solutions: molecular dynamics simulations. *Macromolecules* **2011,** 44, 5798-5816.

170. Caliskan, G.; Hyeon, C.; Perez-Salas, U.; Briber, R. M.; Woodson, S. A.; Thirumalai, D., Persistence length changes dramatically as RNA folds. *Physical Review Letters* **2005,** 95, 268303.

171. Uetrecht, C.; Versluis, C.; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R., Stability and shape of hepatitis B virus capsids in vacuo. *Angewandte Chemie-International Edition* **2008,** 47, 6247-6251.

172. Meng, D.; Hjelm, R. P.; Hu, J.; Wu, J., A theoretical model for the dynamic structure of hepatitis B nucleocapsid. *Biophysical Journal* **2011,** 101, 2476-2484

173. Ting, C. L.; Wu, J.; Wang, Z.-G., Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** 108, 16986-16991.

174. Chen, C.; Wang, J. C.-Y.; Zlotnick, A., A kinase chaperones hepatitis B virus capsid assembly and captures capsid dynamics in vitro. *Plos Pathogens* **2011,** 7, e1002388.

175. Kenney, J. M.; Vonbonsdorff, C. H.; Nassal, M.; Fuller, S. D., Evolutionary conservation in the hepatitis B virus core structure: comparison of human and duck cores. *Structure* **1995,** 3, 1009-1019

176. He, L.; Porterfield, Z.; van der Schoot, P.; Zlotnick, A.; Dragnea, B., Hepatitis virus capsid polymorph stability depends on encapsulated cargo size. *Acs Nano* **2013,** 7, (10), 8447-8454.

177. WHO, Hepatitis B vaccines. *Weekly Epidemiological Record* **2004,** 79, 255-263.

178. Hirsch, R. C.; Lavine, J. E.; Chang, L. J.; Varmus, H. E.; Ganem, D., Polymerase gene-products of hepatitis-B viruses are required for genomic RNA packaging as well as for reverse transcription. *Nature* **1990,** 344, (6266), 552-555.

179. Kenney, J. M.; Vonbonsdorff, C. H.; Nassal, M.; Fuller, S. D., Evolutionary conservation in the hepatitis B virus core structure: comparison of human and duck cores. *Structure* **1995,** 3, (10), 1009-1019

180. Kau, J. H.; Ting, L. P., A serine-kinase-containing protein complex interacts with the terminal protein domain of polymerase of hepatitis B virus. *Journal of Biomedical Science* **1997,** 4, (4), 155-161.

181. Daub, H.; Blencke, S.; Habenberger, P.; Kurtenbach, A.; Dennenmoser, J.; Wissing, J.; Ullrich, A.; Cotten, M., Identification of SRPK1 and SRPK2 as the major cellular protein kinases phosphorylating hepatitis B virus core protein. *Journal of Virology* **2002,** 76, (16), 8124-8137.

182. Chen, C.; Wang, J. C.-Y.; Zlotnick, A., A kinase chaperones Hepatitis B virus capsid assembly and captures capsid dynamics in vitro. *Plos Pathogens* **2011,** 7, (11), e1002388.

183. Roseman, A. M.; Borschukova, O.; Berriman, J. A.; Wynne, S. A.; Pumpens, P.; Crowther, R. A., Structures of hepatitis B virus cores presenting a model epitope and their complexes with antibodies. *Journal of Molecular Biology* **2012,** 423, (1), 63-78.

184. Yu, X.; Jin, L.; Jih, J.; Shih, C.; Zhou, Z. H., 3.5 angstrom cryoEM structure of Hepatitis B virus core assembled from full-length core protein. *Plos One* **2013,** 8, (9), e69729.

185. Ning, X.; Nguyen, D.; Mentzer, L.; Adams, C.; Lee, H.; Ashley, R.; Hafenstein, S.; Hu, J., Secretion of genome-free hepatitis B virus - single strand blocking model for virion morphogenesis of para-retrovirus. *Plos Pathogens* **2011,** 7, (9), e1002255.

186. Jiang, T.; Wang, Z.-G.; Wu, J., Electrostatic Regulation of Genome Packaging in Human Hepatitis B Virus. *Biophysical Journal* **2009,** 96, (8), 3065-3073.

187. Zhang, D. Q.; Konecny, R.; Baker, N. A.; McCammon, J. A., Electrostatic interaction between RNA and protein capsid in cowpea chlorotic mottle virus simulated by a coarse-grain RNA model and a Monte Carlo approach. *Biopolymers* **2004,** 75, (4), 325-337.

188. Kolinski, A.; Godzik, A.; Skolnick, J., A general method for the prediction of the three dimensional structure and folding pathway of globular proteins: Application to designed helical proteins. *Journal of Chemical Physics* **1993,** 98, (9), 7420-7433.

189. Li, Z.; Wu, J.; Wang, Z.-G., Osmotic pressure and packaging structure of caged DNA. *Biophysical Journal* **2008,** 94, (3), 737-746.

190. Bruss, V., Envelopment of the hepatitis B virus nucleocapsid. *Virus Research* **2004,** 106, (2), 199-209.

191. Wynne, S. A.; Crowther, R. A.; Leslie, A. G. W., The crystal structure of the human hepatitis B virus capsid. *Molecular Cell* **1999,** 3, (6), 771-780.

192. Kim, J.; Wu, J., A molecular thermodynamic model for the stability of hepatitis B capsids. *Journal of Chemical Physics* **2014,** 140, 235101.

193. Li, Z.; Wu, J., Density functional theory for planar electric double layers: Closing the gap between simple and polyelectrolytes. *Journal of Physical Chemistry B* **2006,** 110, (14), 7473-7484

194. Li, Z.; Wu, J., Density functional theory for polyelectrolytes near oppositely charged surfaces. *Physical Review Letters* **2006,** 96, (4), 048302.

195. Wu, J. Z., Density functional theory for chemical engineering: From capillarity to soft materials. *Aiche Journal* **2006,** 52, (3), 1169-1193.

196. Wu, J.; Li, Z., Density-functional theory for complex fluids. *Annual Review of Physical Chemistry* **2007,** 58, 85-112.

197. Gui, J. F.; Lane, W. S.; Fu, X. D., A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature* **1994,** 369, (6482), 678-682.

198. Colwill, K.; Feng, L. L.; Yeakley, J. M.; Gish, G. D.; Caceres, J. F.; Pawson, T.; Fu, X. D., SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors. *Journal of Biological Chemistry* **1996,** 271, (40), 24569-24575.

199. Yu, M. S.; Summers, J., Phosphorylation of the duck hepatitis B virus capsid protein associated with conformational changes in the C terminus. *Journal of Virology* **1994,** 68, (5), 2965-2969

200. Hilmer, J. K.; Zlotnick, A.; Bothner, B., Conformational equilibria and rates of localized motion within hepatitis B virus capsids. *Journal of Molecular Biology* **2008,** 375, (2), 581-594.

201. Kann, M.; Schmitz, A.; Rabe, B., Intracellular transport of hepatitis B virus. *World Journal of Gastroenterology* **2007,** 13, (1), 39-47.

202. Pugh, J.; Zweidler, A.; Summers, J., Characterization of the major duck hepatitis B virus core particle protein. *Journal of Virology* **1989,** 63, (3), 1371-1376

203. Zheng, Y. Y.; Fu, X. D.; Ou, J. H. J., Suppression of hepatitis B virus replication by SRPK1 and SRPK2 via a pathway independent of the phosphorylation of the viral core protein. *Virology* **2005,** 342, (1), 150-158.

204. Li, H.-C.; Huang, E.-Y.; Su, P.-Y.; Wu, S.-Y.; Yang, C.-C.; Lin, Y.-S.; Chang, W.-C.; Shih, C., Nuclear export and import of human hepatitis B virus capsid protein and particles. *Plos Pathogens* **2010,** 6, (10), e1001162.

205. Yeh, C. T.; Liaw, Y. F.; Ou, J. H., The arginine-rich domain of hepatitis-B virus precore and core proteins contains a signal for nuclear transport. *Journal of Virology* **1990,** 64, (12), 6141-6147.

206. Newman, M.; Chua, P. K.; Tang, F.-M.; Su, P.-Y.; Shih, C., Testing an electrostatic interaction hypothesis of hepatitis B virus capsid stability by using an in vitro capsid disassembly/reassembly system. *Journal of Virology* **2009,** 83, (20), 10616-10626.

207. Vanlandschoot, P.; Van Houtte, F.; Serruys, B.; Leroux-Roels, G., The arginine-rich carboxy-terminal domain of the hepatitis B virus core protein mediates attachment of nucleocapsids to cell-surface-expressed heparan sulfate. *Journal of General Virology* **2005,** 86, 75-84.

208. Beames, B.; Lanford, R. E., Carboxy-terminal truncations of the HBV core protein affect capsid formation and the apparent size of encapsidated HBV RNA. *Virology* **1993,** 194, (2), 597-607.

209. Le Pogam, S.; Chua, P. K.; Newman, M.; Shih, C., 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). *Journal of Virology* **2005,** 79, (3), 1871-1887.

210. Bringas, R., Folding and assembly of hepatitis B virus core protein: A new model proposal. *Journal of Structural Biology* **1997,** 118, (3), 189-196.

211. Dong, X. F.; Natarajan, P.; Tihova, M.; Johnson, J. E.; Schneemann, A., Particle polymorphism caused by deletion of a peptide molecular switch in a quasiequivalent icosahedral virus. *Journal of Virology* **1998,** 72, (7), 6024-6033.

212. Kaplan, I. B.; Zhang, L.; Palukaitis, P., Characterization of cucumber mosaic virus - V. Cell-to-cell movement requires capsid protein but not virions. *Virology* **1998,** 246, (2), 221-231.

213. Marshall, D.; Schneemann, A., Specific packaging of nodaviral RNA2 requires the N-terminus of the capsid protein. *Virology* **2001,** 285, (1), 165-175.

214. Schmitz, I.; Rao, A. L. N., 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* **1998,** 248, (2), 323-331.

215. Schneemann, A.; Marshall, D., Specific encapsidation of nodavirus RNAs is mediated through the C terminus of capsid precursor protein alpha. *Journal of Virology* **1998,** 72, (11), 8738-8746.

216. Scott, K. P.; Farmer, M. J.; Robinson, D. J.; Torrance, L.; Murant, A. F., Comparison of the coat protein of groundnut rosette assistor virus with those of other luteoviruses. *Annals of Applied Biology* **1996,** 128, (1), 77-83.

217. Smith, T. J.; Chase, E.; Schmidt, T.; Perry, K. L., The structure of cucumber mosaic virus and comparison to cowpea chlorotic mottle virus. *Journal of Virology* **2000,** 74, (16), 7578-7586.

218. Tihova, M.; Dryden, K. A.; Le, T. V. L.; Harvey, S. C.; Johnson, J. E.; Yeager, M.; Schneemann, A., Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. *Journal of Virology* **2004,** 78, (6), 2897-2905.

219. Belyi, V. A.; Muthukumar, M., Electrostatic origin of the genome packing in viruses. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** 103, (46), 17174-17178.

220. Nassal, M., 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. *Journal of Virology* **1992,** 66, (7), 4107-4116.

221. Ning, X.; Nguyen, D.; Mentzer, L.; Adams, C.; Lee, H.; Ashley, R.; Hafenstein, S.; Hu, J., Secretion of genome-free hepatitis B virus - single strand blocking model for virion morphogenesis of para-retrovirus. *Plos Pathogens* **2011,** 7, (9).

222. Schmitz, A.; Schwarz, A.; Foss, M.; Zhou, L.; Rabe, B.; Hoellenriegel, J.; Stoeber, M.; Pante, N.; Kann, M., Nucleoporin 153 Arrests the Nuclear Import of Hepatitis B Virus Capsids in the Nuclear Basket. *Plos Pathogens* **2010,** 6, (1).

223. Hu, T.; Zhang, R.; Shkovskii, B. I., Electrostatic theory of viral self-assembly. *Physica a-Statistical Mechanics and Its Applications* **2008,** 387, (12), 3059-3064.

224. Angelescu, D. G.; Bruinsma, R.; Linse, P., Monte Carlo simulations of polyelectrolytes inside viral capsids. *Physical Review E* **2006,** 73, (4).

225. Siber, A.; Podgornik, R., Nonspecific interactions in spontaneous assembly of empty versus functional single-stranded RNA viruses. *Physical Review E* **2008,** 78, (5).

226. van der Schoot, P.; Bruinsma, R., Electrostatics and the assembly of an RNA virus. *Physical Review E* **2005,** 71, (6).

227. Ni, P.; Wang, Z.; Ma, X.; Das, N. C.; Sokol, P.; Chiu, W.; Dragnea, B.; Hagan, M.; Kao, C. C., An examination of the electrostatic interactions between the N-terminal tail of the brome mosaic virus coat protein and encapsidated RNAs. *Journal of Molecular Biology* **2012,** 419, (5), 284-300.

228. Ting, C. L.; Wu, J.; Wang, Z.-G., Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation. *Proceedings of the National Academy of Sciences of the United States of America* **2011,** 108, (41), 16986-16991.

229. Perlmutter, J. D.; Qiao, C.; Hagan, M. F., Viral genome structures are optimal for capsid assembly. *Elife* **2013,** 2.

230. van der Schoot, P.; Zandi, R., Impact of the topology of viral RNAs on their encapsulation by virus coat proteins. *Journal of Biological Physics* **2013,** 39, (2), 289-299.

231. Kim, J.; Wu, J., A theoretical study of SRPK Interaction with the flexible domains of hepatitis B capsids. *Biophysical Journal* **2014,** 107, (6), 1453-1461.

232. Kolinski, A.; Godzik, A.; Skolnick, J., A general method for the prediction of the three dimensional structure and folding pathway of globular proteins: Application to designed helical proteins *Journal of Chemical Physics* **1993,** 98, (9), 7420-7433.

233. Angelescu, D. G.; Linse, P., Modelling of icosahedral viruses. *Current Opinion in Colloid & Interface Science* **2008,** 13, (6), 389-394.

234. Elrad, O. M.; Hagan, M. F., Encapsulation of a polymer by an icosahedral virus. *Physical Biology* **2010,** 7, (4).

235. Hagan, M. F.; Chandler, D., Dynamic pathways for viral capsid assembly. *Biophysical Journal* **2006,** 91, (1), 42-54.

236. Petrov, A. S.; Harvey, S. C., Structural and thermodynamic principles of viral packaging. *Structure* **2007,** 15, (1), 21-27.

237. Simonin, J. P.; Bernard, O.; Blum, L., Ionic solutions in the binding mean spherical approximation: Thermodynamic properties of mixtures of associating electrolytes. *Journal of Physical Chemistry B* **1999,** 103, (4), 699-704.

238. Ostrow, K. M.; Loeb, D. D., Underrepresentation of the 3 ' region of the capsid pregenomic RNA of duck hepatitis B virus. *Journal of Virology* **2004,** 78, (5), 2179-2186.

239. Jiang, H.; Sun, S. X., Cellular pressure and volume regulation and implications for cell mechanics. *Biophysical Journal* **2013,** 105, (3), 609-619.

240. Tinevez, J.-Y.; Schulze, U.; Salbreux, G.; Roensch, J.; Joanny, J.-F.; Paluch, E., Role of cortical tension in bleb growth. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** 106, (44), 18581-18586.

241. Jiang, T.; Li, Z.; Wu, J., Structure and swelling of grafted polyelectrolytes: Predictions from a nonlocal density functional theory. *Macromolecules* **2007,** 40, (2), 334-343.

242. Li, Z. D.; Wu, J. Z., Density functional theory for polyelectrolytes near oppositely charged surfaces. *Physical Review Letters* **2006,** 96, (4).

243. Cui, X.; Ludgate, L.; Ning, X.; Hu, J., Maturation-associated destabilization of hepatitis B virus nucleocapsid. *Journal of Virology* **2013,** 87, (21), 11494-11503.

244. Kim, J.; Wu, J., A molecular thermodynamic model for the stability of hepatitis B capsids. *Journal of Chemical Physics* **2014,** 140, (23), 235101.

245. Sominskaya, I.; Skrastina, D.; Petrovskis, I.; Dishlers, A.; Berza, I.; Mihailova, M.; Jansons, J.; Akopjana, I.; Stahovska, I.; Dreilina, D.; Ose, V.; Pumpens, P., 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* **2013,** 8, (9).

246. Maupetit, J.; Derreumaux, P.; Tuffery, P., PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Research* **2009,** 37, W498-W503.

247. Maupetit, J.; Derreumaux, P.; Tuffery, P., A Fast Method for Large-Scale De Novo Peptide and Miniprotein Structure Prediction. *Journal of Computational Chemistry* **2010,** 31, (4), 726-738.

248. Thevenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tuffery, P., PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Research* **2012,** 40, (W1), W288-W293.

249. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **2008,** 4, (3), 435-447.

250. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **1996,** 118, (45), 11225-11236.

251. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *Journal of Physical Chemistry A* **1997,** 101, (16), 3005-3014.

252. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *Journal of Chemical Physics* **2007,** 126, (1).

253. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005,** 26, (16), 1668-1688.

254. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **1983,** 79, (2), 926-935.

255. Honig, B.; Nicholls, A., Classical electrostatics in biology and chemistry. *Science* **1995,** 268, (5214), 1144-1149.

256. Sitkoff, D.; Sharp, K. A.; Honig, B., Accurate calculation of hydration free-energies using macroscopic solvent models. *Journal of Physical Chemistry* **1994,** 98, (7), 1978-1988.

257. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of Computational Chemistry* **2009,** 30, (16), 2785-2791.

258. Gasteiger, J.; Marsili, M., Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron* **1980,** 36, (22), 3219-3228.

259. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* **1998,** 19, (14), 1639-1662.

260. Logean, A.; Sette, A.; Rognan, D., Customized versus universal scoring functions: Application to class I MHC-peptide binding free energy predictions. *Bioorganic & Medicinal Chemistry Letters* **2001,** 11, (5), 675-679.

261. Corse, E.; Gottschalk, R. A.; Allison, J. P., Strength of TCR-Peptide/MHC Interactions and In Vivo T Cell Responses. *Journal of Immunology* **2011,** 186, (9), 5039-5045.

262. Germain, R. N.; Stefanova, I., The dynamics of T cell receptor signaling: Complex orchestration and the key roles of tempo and cooperation. *Annual Review of Immunology* **1999,** 17, 467-522.

263. Basdevant, N.; Weinstein, H.; Ceruso, M., Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study. *Journal of the American Chemical Society* **2006,** 128, (39), 12766-12777.

264. Hou, T.; Chen, K.; McLaughlin, W. A.; Lu, B.; Wang, W., Computational analysis and prediction of the binding motif and protein interacting partners of the Abl SH3 domain. *Plos Computational Biology* **2006,** 2, (1), 46-55.

265. Huang, Y.-m. M.; Kang, M.; Chang, C.-e. A., Mechanistic Insights into Phosphopeptide-BRCT Domain Association: Preorganization, Flexibility, and Phosphate Recognition. *Journal of Physical Chemistry B* **2012,** 116, (34), 10247-10258.

266. Kim, J.; Wu, J. Z., A Theoretical Study of SRPK Interaction with the Flexible Domains of Hepatitis B Capsids. *Biophysical Journal* **2014,** 107, (6), 1453-1461.

267. Ball, P., Water as an active constituent in cell biology. *Chemical Reviews* **2008,** 108, (1), 74-108.

268. Chandler, D., Interfaces and the driving force of hydrophobic assembly. *Nature* **2005,** 437, (7059), 640-647.

269. Ashbaugh, H. S.; Pratt, L. R., Colloquium: Scaled particle theory and the length scales of hydrophobicity. *Reviews of Modern Physics* **2006,** 78, (1), 159-178.

270. Tanford, C., *The hydrophobic effect: formation of micelles and biological membranes*. Wiley: New York,, 1973; p viii, 200 p.

271. Galamba, N., Water's Structure around Hydrophobic Solutes and the Iceberg Model. *Journal of Physical Chemistry B* **2013,** 117, (7), 2153-2159.

272. Galamba, N., Reply to "Comment on 'Water's structure around hydrophobic solutes and the iceberg model'". *Journal of Physical Chemistry B* **2014,** 118, 2600-2603.

273. Graziano, G., Comment on "Water's structure around hydrophobic solutes and the iceberg model". *Journal of Physical Chemistry B* **2014,** 118, 2598–2599.

274. Davis, J. G.; Gierszal, K. P.; Wang, P.; Ben-Amotz, D., Water structural transformation at molecular hydrophobic interfaces. *Nature* **2012,** 491, (7425), 582-585.

275. Frank, H. S.; Evans, M. W., Free volume and entropy in condensed systems III. Entropy in binary liquid mixtures; Partial molal entropy in dilute solutions; Structure and thermodynamics in aqueous electrolytes. *Journal of Chemical Physics* **1945,** 13, (11), 507-532

276. Kauzmann, W., Some factors in the interpretation of protein denaturation. *Advances in Protein Chemistry* **1959,** 14, 1-63.

277. Kauzmann, W., Thermodynamics of unfolding. *Nature* **1987,** 325, (6107), 763-764

278. Bowron, D. T.; Soper, A. K.; Finney, J. L., Temperature dependence of the structure of a 0.06 mole fraction tertiary butanol-water solution. *Journal of Chemical Physics* **2001,** 114, (14), 6203-6219.

279. Dixit, S.; Soper, A. K.; Finney, J. L.; Crain, J., Water structure and solute association in dilute aqueous methanol. *Europhysics Letters* **2002,** 59, (3), 377-383.

280. Soper, A. K.; Finney, J. L., Hydration of methanol in aqueous solution. *Physical Review Letters* **1993,** 71, (26), 4346-4349

281. Scatena, L. F.; Brown, M. G.; Richmond, G. L., Water at hydrophobic surfaces: Weak hydrogen bonding and strong orientation effects. *Science* **2001,** 292, (5518), 908-912.

282. Buchanan, P.; Aldiwan, N.; Soper, A. K.; Creek, J. L.; Koh, C. A., Decreased structure on dissolving methane in water. *Chemical Physics Letters* **2005,** 415, (1-3), 89-93.

283. Graziano, G.; Lee, B., On the intactness of hydrogen bonds around nonpolar solutes dissolved in water. *Journal of Physical Chemistry B* **2005,** 109, (16), 8103-8107.

284. Wu, J.; Prausnitz, J. M., Pairwise-additive hydrophobic effect for alkanes in water. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** 105, (28), 9512-9515.

285. Lucas, M., Size effect in transfer of nonpolar solutes from gas or solvent to another solvent with a view on hydrophobic behavior. *Journal of Physical Chemistry* **1976,** 80, (4), 359-362.

286. Pierotti, R. A., A scaled particle theory of aqueous and nonaqueous solutions. *Chemical Reviews* **1976,** 76, (6), 717-726

287. Pratt, L. R.; Chandler, D., Theory of the hydrophobic effect. *Journal of Chemical Physics* **1977,** 67, 3683.

288. Lee, B., The physical origin of the low solubility of nonpolar solutes in water. *Biopolymers* **1985,** 24, (5), 813-823.

289. Ashbaugh, H. S.; Truskett, T. M.; Debenedetti, P. G., A simple molecular thermodynamic theory of hydrophobic hydration. *Journal of Chemical Physics* **2002,** 116, (7), 2907-2921.

290. Harano, Y.; Roth, R.; Chiba, S., A Morphometric Approach for the Accurate Solvation Thermodynamics of Proteins and Ligands. *Journal of Computational Chemistry* **2013,** 34, (23), 1969-1974.

291. Wu, J., Solvation of a Spherical Cavity in Simple Liquids: Interpolating between the Limits. *Journal of Physical Chemistry B* **2009,** 113, (19), 6813-6818.

292. Jin, Z. H.; Kim, J.; Wu, J. Z., Shape Effect on Nanoparticle Solvation: A Comparison of Morphometric Thermodynamics and Microscopic Theories. *Langmuir* **2012,** 28, (17), 6997-7006.

293. Madan, B.; Lee, B., Role of hydrogen bonds in hydrophobicity: the free energy of cavity formation in water models with and without the hydrogen bonds. *Biophysical Chemistry* **1994,** 51, (2-3), 279-289

294. Lee, B., Enthalpy-entropy compensation in the thermodynamics of hydrophobicity. *Biophysical Chemistry* **1994,** 51, (2-3), 271-278.

295. Ben-Amotz, D.; Underwood, R., Unraveling water's entropic mysteries: A unified view of nonpolar, polar, and ionic hydration. *Accounts of Chemical Research* **2008,** 41, (8), 957-967.

296. Blokzijl, W.; Engberts, J., Hydrophobic effects. Opinions and facts. *Angewandte Chemie-International Edition* **1993,** 32, (11), 1545-1579

297. Graziano, G., On the size dependence of hydrophobic hydration. *Journal of the Chemical Society-Faraday Transactions* **1998,** 94, (22), 3345-3352.

298. Lee, B.; Graziano, G., A two-state model of hydrophobic hydration that produces compensating enthalpy and entropy changes. *Journal of the American Chemical Society* **1996,** 118, (22), 5163-5168.

299. Southall, N. T.; Dill, K. A.; Haymet, A. D. J., A view of the hydrophobic effect. *Journal of Physical Chemistry B* **2002,** 106, (3), 521-533.

300. Lee, B., Solvent reorganization contribution to the transfer thermodynamics of small nonpolar molecules. *Biopolymers* **1991,** 31, (8), 993-1008.

301. Bakulin, A. A.; Pshenichnikov, M. S.; Bakker, H. J.; Petersen, C., Hydrophobic molecules slow down the hydrogen-bond dynamics of water. *Journal of Physical Chemistry A* **2011,** 115, (10), 1821-1829.

302. Haselmeier, R.; Holz, M.; Marbach, W.; Weingartner, H., Water dynamics near a dissolved noble gas. First direct experimental evidence for a retardation effect. *Journal of Physical Chemistry* **1995,** 99, (8), 2243-2246.

303. Qvist, J.; Halle, B., Thermal signature of hydrophobic hydration dynamics. *Journal of the American Chemical Society* **2008,** 130, (31), 10345-10353.

304. Rezus, Y. L. A.; Bakker, H. J., Observation of immobilized water molecules around hydrophobic groups. *Physical Review Letters* **2007,** 99, (14).

305. Rossato, L.; Rossetto, F.; Silvestrelli, P. L., Aqueous solvation of methane from first principles. *Journal of Physical Chemistry B* **2012,** 116, (15), 4552-4560.

306. Rossky, P. J.; Karplus, M., Solvation. A molecular dynamics study of a dipeptide in water. *Journal of the American Chemical Society* **1979,** 101, (8), 1913-1937.

307. Silvestrelli, P. L., Are there immobilized water molecules around hydrophobic groups? Aqueous solvation of methanol from first principles. *Journal of Physical Chemistry B* **2009,** 113, (31), 10728-10731.

308. Wachter, W.; Buchner, R.; Hefter, G., Hydration of tetraphenylphosphonium and tetraphenylborate ions by dielectric relaxation spectroscopy. *Journal of Physical Chemistry B* **2006,** 110, (10), 5147-5154.

309. Zichi, D. A.; Rossky, P. J., Solvent molecular dynamics in regions of hydrophobic hydration. *Journal of Chemical Physics* **1986,** 84, (5), 2814-2822.

310. Bin, S.; Molinero, V., Thermodynamic and structural signatures of water-driven methane-methane attraction in coarse-grained mW water. *Journal of Chemical Physics* **2013,** 139, (5), 054511

311. Krause, D.; Benson, B. B., The solubility and isotopic fractionation of gases in dilute aqueous solution. IIa. Solubilities of the noble gases. *Journal of Solution Chemistry* **1989,** 18, (9), 823-872

312. Yaws, C. L.; Yang, H. C., Henry's law constant for compound in water. Thermodynamic and physical property data. In Gulf Publishing Company: Houston, TX, 1992; pp 181-206.

313. Reiss, H.; Frisch, H. L.; Lebowitz, J. L., Statistical Mechanics of Rigid Spheres. *Journal of Chemical Physics* **1959,** 31, (2), 369-380.

314. Kell, G. S., Density, thermal expansivity, and compressibility of liquid water from 0.deg. to 150.deg.. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. *Journal of Chemical and Engineering Data* **1975,** 20, (1), 97-105.

315. Schulze, G.; Prausnitz, J. M., Solubilities of gases in water at high temperatures. *Industrial & Engineering Chemistry Fundamentals* **1981,** 20, (2), 175-177.

316. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The missing term in effective pair potentials. *Journal of Physical Chemistry* **1987,** 91, (24), 6269-6271.

317. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **1984,** 81, (8), 3684-3690

318. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981,** 52, (12), 7182-7190

319. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *Journal of Chemical Physics* **1995,** 103, (19), 8577-8593.

320. Weeks, J. D.; Chandler, D.; Andersen, H. C., Role of repulsive forces in determining the equilibrium structure of simple liquids. *Journal of Chemical Physics* **1971,** 54, (12), 5237- 5247.

321. White, J. A., Lennard-Jones as a model for argon and test of extended renormalization group calculations. *Journal of Chemical Physics* **1999,** 111, (20), 9352-9356.

322. Jorgensen, W. L.; Madura, J. D.; Swenson, C. J., Optimized intermolecular potential functions for liquid hydrocarbons. *Journal of the American Chemical Society* **1984,** 106, (22), 6638-6646.

323. Rajamani, S.; Ghosh, T.; Garde, S., Size dependent ion hydration, its asymmetry, and convergence to macroscopic behavior. *Journal of Chemical Physics* **2004,** 120, (9), 4457-4466.

324. Paschek, D., Heat capacity effects associated with the hydrophobic hydration and interaction of simple solutes: A detailed structural and energetical analysis based on molecular dynamics simulations. *Journal of Chemical Physics* **2004,** 120, (22), 10605-10617.

325. Sharp, K. A.; Madan, B., Hydrophobic effect, water structure, and heat capacity changes. *Journal of Physical Chemistry B* **1997,** 101, (21), 4343-4348.

326. Huang, X.; Margulis, C. J.; Berne, B. J., Do molecules as small as neopentane induce a hydrophobic response similar to that of large hydrophobic surfaces? *Journal of Physical Chemistry B* **2003,** 107, (42), 11742-11748.

327. Graziano, G., Comment on "Do molecules as small as neopentane induce a hydrophobic response similar to that of large hydrophobic surfaces?" *Journal of Physical Chemistry B* **2004,** 108, (26), 9371-9372.

328. Luzar, A.; Chandler, D., Hydrogen-bond kinetics in liquid water. *Nature* **1996,** 379, (6560), 55-57.

329. Graziano, G.; Lee, B., Entropy convergence in hydrophobic hydration: a scaled particle theory analysis. *Biophysical Chemistry* **2003,** 105, (2-3), 241-250.

330. Grunwald, E.; Steel, C., Solvent reorganization and thermodynamic enthalpy-entropy compensation. *Journal of the American Chemical Society* **1995,** 117, (21), 5687-5692.

331. Qian, H.; Hopfield, J. J., Entropy-enthalpy compensation: Perturbation and relaxation in thermodynamic systems. *Journal of Chemical Physics* **1996,** 105, (20), 9292-9298.

332. Yu, H. A.; Karplus, M., A thermodynamic analysis of solvation. *Journal of Chemical Physics* **1988,** 89, (4), 2366-2379.

333. Eaves, J. D.; Loparo, J. J.; Fecko, C. J.; Roberts, S. T.; Tokmakoff, A.; Geissler, P. L., Hydrogen bonds in liquid water are broken only fleetingly. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, (37), 13019-13022.

334. Wernet, P.; Nordlund, D.; Bergmann, U.; Cavalleri, M.; Odelius, M.; Ogasawara, H.; Naslund, L. A.; Hirsch, T. K.; Ojamae, L.; Glatzel, P.; Pettersson, L. G. M.; Nilsson, A., The structure of the first coordination shell in liquid water. *Science* **2004,** 304, (5673), 995-999.

335. Eisenberg, D.; Kauzmann, W., *The structure and properties of water*. Oxford University Press: New York, 1969.

336. Smith, J. D.; Cappa, C. D.; Wilson, K. R.; Cohen, R. C.; Geissler, P. L.; Saykally, R. J., Unified description of temperature-dependent hydrogen-bond rearrangements in liquid water. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, (40), 14171-14174.

337. Laage, D.; Stirnemann, G.; Hynes, J. T., Why water reorientation slows without iceberg formation around hydrophobic solutes. *Journal of Physical Chemistry B* **2009,** 113, (8), 2428-2435.