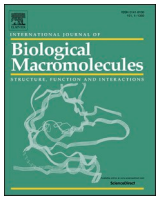
International Journal of Biological Macromolecules 263 (2024) 130365

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac



Impact of DNA on interactions between core proteins of Hepatitis B virus-like particles comprising different C-terminals

Srdjan Pusara, Wolfgang Wenzel, Mariana Kozlowska *

Institute of Nanotechnology, Karlsruhe Institute of Technology KIT, Kaiserstraße 12, 76131 Karlsruhe, Germany

A B S T R A C T

Hepatitis B virus (HBV) virus-like particles (VLPs) are promising therapeutic agents derived from HBV core proteins (Cp). This study investigates the assembly dynamics of HBV VLPs, which is crucial for their potential as drug carriers or gene delivery systems. Coarse-grained molecular dynamics simulations explore the impact of C-terminal domain length (in the Cp ranging from Cp149 to wild-type Cp183) on Cp assembly and stability, particularly in the presence of DNA. Our findings reveal that the C-terminal nucleic acid binding region significantly influences Cp assembly and stability of trimers comprising Cp dimers. Shorter C-terminal domains (Cp164, Cp167) enhance stability and protein-protein interactions, while interactions between naturally occurring Cp183 are destabilized in the absence of DNA. Interestingly, DNA addition further stabilizes Cp assemblies, and this effect is influenced by the length of the nucleic acid binding region. Shorter C-terminal domains show less dependency on DNA content. This stabilization is attributed to electrostatic forces between positively charged C-terminal chains and negatively charged nucleic acids. Our study sheds light on the molecular mechanisms governing protein-protein and protein-DNA interactions in HBV VLP assembly, providing insights into Cp processability and informing the development of efficient gene therapy carriers using VLP technology.

1. Introduction

A virus capsid is a protein shell that encapsulates the viral genome and provides protection for the virus. Virus-like particles (VLPs) are nanoparticles that mimic the structure of viral capsids, but are assembled from core proteins without the viral genome, making them noninfectious [1,2]. Due to their structural stability, VLPs are promising candidates for next-generation vaccines and drug delivery systems with high immunogenicity and ability to encapsulate therapeutic nucleic acids [1,3]. Hepatitis B virus (HBV) core protein-based VLPs are particularly attractive for gene therapy and drug delivery applications because they can self-assemble in various expression systems, making them versatile and efficient nanocarriers [4].

The HBV capsid is composed of core proteins (Cp) and a lipid envelope containing other types of proteins [5,6]. The mature HBV capsid has an icosahedral structure with quasi-symmetry ($T = 4$, see Fig. 1a) and is composed of 120 dimers of Cp, with a smaller fraction of capsids (~5 %) also having $T = 3$ symmetry, formed by 90 copies of Cp dimers [7]. The overall diameters of those two types of HBV capsids are 36 nm and 32 nm, respectively, allowing the further encapsulation of the desired molecules. The native Cp monomer consists of 183 amino acids (Cp183) and contains two separate domains: the N-terminal assembly domain at amino acid (aa) residues 1–140 (as depicted in Fig. 1), and the C-terminal (C-ter) domain at residues 150–183 responsible for binding of nucleic acids and intracellular trafficking [4]. The hinge peptide, formed between 141 and 149 aa, links these domains and can perform morphogenic functions [8]. Cp proteins exist like dimers in specific solution conditions [7], and they can undergo self-assembly in vitro conditions giving capsids of indistinguishable morphology compared to the natural ones. Dimers of truncated Cp are also capable of assembly, provided that truncation is up to 140 aa [4,9–12]. It is known that Cp of 149 aa can self-assemble into the empty VLP, however, the full length Cp (i.e. 183 aa) allows the VLP assembly only in higher salt conditions (≥0.25 M) or in the presence of nucleic acids or other types of polyanions [13]. This is caused by the fact that the C-ter domain of the Cp protein contains positively charged arginine residues, which disable VLP formation in the absence of counterions due to the electrostatic repulsion interactions [14]. Therefore, C-ter is responsible for binding the negatively charged nucleic acids, e.g. viral genome. Unlike the assembly domain, which has a stable tertiary structure and is responsible for the capsid's surface charge, the C-ter domain lacks a defined tertiary structure, making it highly flexible [15]. During the HBV life cycle, the C-ter domain is primarily located in the capsid interior, but it can also be found on the capsid's outer surface, penetrating through capsid pores [16]. The unique flexibility and adaptability of the C-ter domain are critical for its involvement in several crucial interactions, including genome packaging or capsid assembly and disassembly [4].

In vitro assembly of HBV capsids begins with a slow nucleation step,

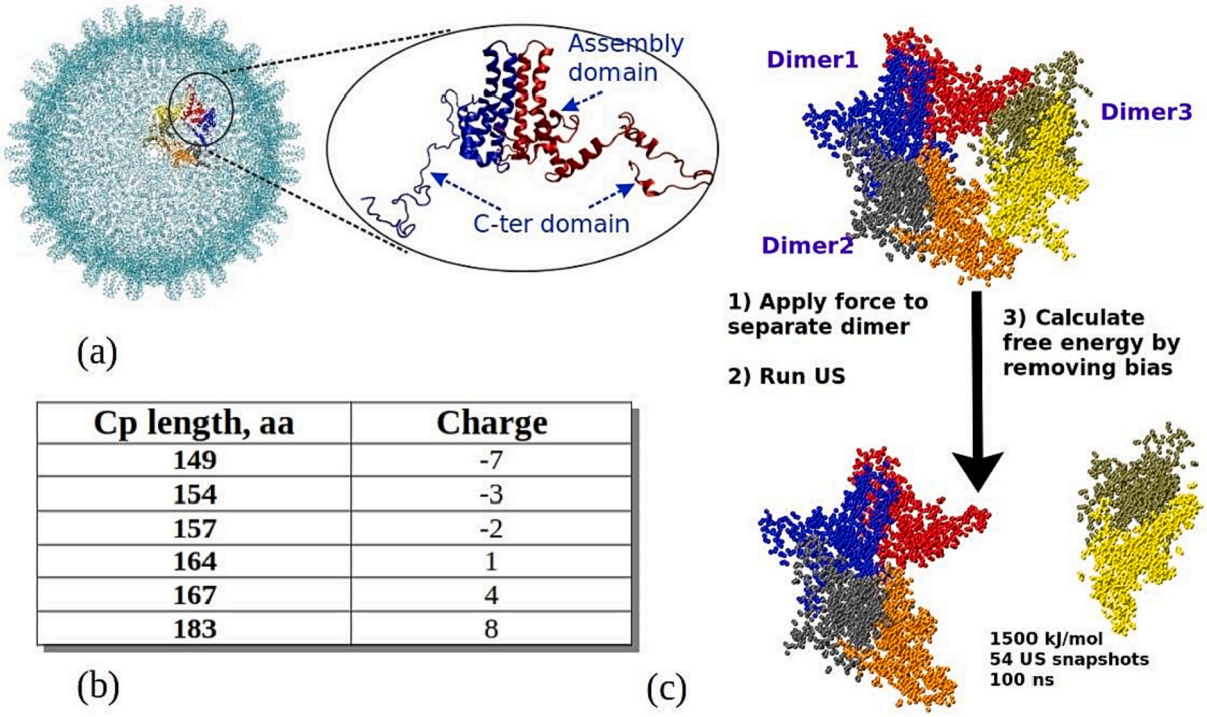
* Corresponding author.
E-mail address: mariana.kozlowska@kit.edu (M. Kozlowska).

https://doi.org/10.1016/j.ijbiomac.2024.130365
Received 28 November 2023; Received in revised form 14 February 2024; Accepted 20 February 2024
Available online 23 February 2024

forming a trimer of dimers, followed by a rapid elongation phase whose mechanism differs in the presence and absence of nucleic acids [7,17]. The assembly is primarily driven by weak hydrophobic interactions at dimer-dimer interfaces, counterbalanced by electrostatic repulsion [18]. These weak interactions minimize the possibility of mis-assembly (known as kinetic traps) by allowing incorrectly bound subunits to disassemble and reassemble correctly, especially in the case, when assembly occurs without the presence of the polyions [19–21]. The assembly process is allosterically regulated and can be modulated by changes in solution conditions, such as pH, temperature, or ionic strength, which reduce electrostatic repulsions between dimers and possibly induce a conformational change in Cp to an assembly-active state [7,22]. Truncated Cp149 capsids have been extensively studied, with numerous experiments investigating the kinetics, thermodynamics, and mechanism of capsid assembly or disassembly [23–29]. In addition, some studies focused on modulating Cp149 capsid stability or preventing assembly by introducing mutations at certain residues [30–33] or in the presence of small molecule antivirals [34–36]. Several experimental studies have also investigated the in vitro assembly of empty, as well as RNA-containing HBV capsids using full-length Cp183 proteins. They have found some notable differences from the assembly of empty truncated Cp149 capsids [13,14,37,38]. These studies have demonstrated that assembly of the full-length Cp183 dimers occurs only at higher ionic strengths ($\geq$0.25 M) due to the strong electrostatic repulsion among the positively charged C-ter domain [13]. In both cases the VLPs formed are rather stable, which is a natural function of the protein shell protecting genome.

Due to the structural stability of the HBV capsids, its VLPs have garnered attention for their potential in biomedical applications as carriers of therapeutic nucleic acids [4]. Up to this point, VLPs composed of naturally occurring Cp183 dimers have shown the ability to encapsulate a diverse array of RNA molecules [13,39,40], and similar results have been observed with VLPs derived from phosphorylated and mutated Cp183 variants [13]. However, their practical use on a large scale is hindered by the necessity to conduct their processing in severe conditions (e.g. high urea concentration and alkaline conditions) [41] due to the low solubility of Cp183 dimers [13] and the high tendency to bind the host genetic material, used for their expression [4]. As a result, researchers have explored the use of truncated Cp dimers in their quest to identify optimal HBV VLP systems that would allow them to find a balance between nucleic acid binding and Cp dimers solubility, as demonstrated by studies conducted by Sominskaya et al. [12] and Valentic et al. [42]. These studies have revealed that variations in the lengths of nucleic acid binding regions (i.e. C-ter) and the inclusion of host nucleic acids within HBV VLPs affect capsid stability, thus exerting influence on phase behavior, purification processes, and the yield of purified Cp dimers [42]. The production of HBV VLPs starts with expression in a host organism, such as *E. coli.* During this process, Cp dimers form capsids within host cells, inadvertently encapsulating host cell nucleic acids, which is an undesired feature of this procedure [4,13,42]. Subsequently, the cells undergo lysis, followed by a series of purification steps, which include disassembling host-assembled capsids and extracting dimers while preventing capsid reassembly, as well as purifying from host nucleic acids [41]. Once purified, dimers can be assembled under controlled conditions to encapsulate target RNA or DNA molecules. The understanding of the interplay of diverse interactions between proteins and RNA or DNA molecules during these processes is essential for the optimization of processing conditions.

The formation and stability of HBV VLPs is governed by a complex interplay of various intermolecular interactions, with electrostatic forces often being the most dominant. Thus, theoretical studies are essential for comprehending the interplay between protein-protein and protein-nucleic acid interactions, particularly in light of the challenges encountered in experiments. This knowledge can, in turn, facilitate the achievement of large-scale VLP production for nucleic acid delivery. However, the current state-of-the-art is scarce in the simulation VLPs formed from Cp with various C-ter. Theoretical studies on HBV capsids have predominantly been focused on VLPs formed from truncated



Cp length, aa	Charge
149	-7
154	-3
157	-2
164	1
167	4
183	8

Fig. 1. a) All-atom structure of the $T = 4$ HBV capsid (PDB 6htx). For the sake of illustration, the trimer of dimers is highlighted within the capsid structure. The structure of the dimer of core proteins (Cp_2) is magnified on the right side with marked assembly and C-ter domains. b) Charges of core proteins (per one Cp monomer) for full-length Cp183 and Cp with truncated C-ter. c) Computational scheme of umbrella sampling simulations performed on the trimer of Cp dimers by using SIRAH force field in the present work.

Cp149 dimers as model systems. For instance, Zlotnick et al. developed a theoretical model for HBV capsid assembly, which can predict assembly kinetics and extract thermodynamic parameters from experimental data [43]. All-atom (AA) and coarse-grained (CG) molecular dynamics simulations have been used to study Cp149 capsids, including flexibility and sodium ion distribution [44], disassembly under mechanical stress [45] and irreversible deformation mimicking nanoindentation experiments [46], as well as stability when complexed with antiviral compounds [47–53]. However, full-length capsids have received relatively less attention compared to Cp149 truncated ones. Thermodynamic models for Cp183 capsids have been developed, with Cp183 dimers represented by low-resolution coarse-grained structures (1–3 CG beads for the assembly-competent region and one CG bead per residue for the flexible C-terminal domain) [54–57]. RNA chains were represented by charged spheres with identical diameter, valence, and electrolyte was taken into account explicitly. Despite the absence of atomic-level resolution, these models were able to capture experimental trends in modulating the stability of capsids as a function of ionic strength and temperature, as well as the change of stability due to single point mutations and T3/T4 dimorphism [54]. Moreover, these studies were able to predict the uniform distribution of nucleic acids within the capsid, as well as the level of exposure of the C-terminal domain on the outer capsid surface. Additionally, by minimizing the free energy of the capsid system, these models permit to determine the optimal amount of encapsidated nucleic acids giving the most stable capsids [55–57]. However, prior theoretical studies did not focus on the quantification of protein-protein interactions involving Cp proteins, particularly in the context of core length modification. Furthermore, insufficient emphasis was placed on investigating Cp interactions with nucleic acids, despite their importance in the assembly and disassembly of VLPs, where a complex interplay between protein-protein and protein-nucleic acids is crucial. This gap in research hinders our theoretical understanding of these processes.

To address this gap and to provide the molecular bases of the early-step assembly or insights of disassembly of Cp nuclei in the presence of the nucleic acids, in this study, we perform CG molecular dynamics (MD) simulations and umbrella sampling (US) simulations. The trimers of the Cp dimers, named here as Cp_2, i.e. trimers of Cp_2, are used as the smallest nuclei representing the fragment of the VLP. The binding energy between the Cp_2 dimers with different C-ter lengths (Cp183, Cp167, Cp164, Cp157, Cp154 and Cp149) in an explicit water environment at low ionic strength conditions (37.5 mM NaCl) and 300 K temperature in the presence and absence of DNA is explored. For that, the SIRAH CG force field for molecular systems studied is applied. To understand the differences in the Cp dimer interactions within the trimer and the stabilization effects in the presence of nucleic acids, the model DNA of 23 base pairs is investigated. We aim to enhance our understanding of HBV capsid assembly and stability factors, and the delicate balance between protein-protein and protein-nucleic acids interactions.

2. Computational methods

2.1. System preparation and MD simulations

The all-atom structure of HBV Cp protein was obtained from the PDB website (code 6htx) [58] and missing atoms or residues were reconstructed using the Swiss-Model program (v9.15) [59]. Based on the 6htx crystal structure of Cp, the protein structure was modeled up to residue 183. The conformation of the flexible C-ter domain at residues 150–183, which is not resolved in any of the available PDB structures of HBV Cp, was set by superimposing it with the reconstructed Cp dimer structure obtained by Ingemar et al. [37]. These authors used Bayesian inference to obtain ensembles of conformations that match experimental Small-Angle X-ray Scattering (SAXS) spectra measured at low protein concentration, where Cp dimers exist in an unassembled state. To achieve computational efficiency, we simulated a smaller capsid fragment, a trimer of Cp dimers (as depicted in Fig. 1a). The crystallographic symmetry operations (given in PDB 6htx) were then applied to the dimer of two Cp, i.e. Cp_2183 dimer, to obtain the structure of the whole capsid. A trimer of dimers, which was reported to be the smallest stable intermediate of HBV (also in the case of the assembly in the presence of nucleic acids [37]), was then taken from this structure and used for all MD simulations. To investigate the effect of the C-ter length of Cp on the dimer interface stability and DNA binding, truncated trimers of dimers were obtained by extracting the protein chains at the desired positions (i.e. Cp149, Cp154, Cp157, Cp164, and Cp167, respectively).

The MD simulations were conducted using GROMACS program (version 2019.2) [60] and the SIRAH coarse-grained force field version 2.0 [61] for both types of macromolecules considered, i.e. proteins and DNA. SIRAH models proteins with high structural resolution by preserving the positions of backbone C and N atoms while modeling amino acid side chains with one or two additional beads depending on group size. It allows hydrogen bond-like interactions, which stabilize the secondary structure, therefore, SIRAH enables the simulation of proteins without introducing bias to fix the secondary structure. For DNA, SIRAH preserves the positions of phosphorus atoms, representing the phosphate group, along with the C1' and C5' atoms of nucleotides, as well as the atoms involved in Watson-Crick interactions, forming hydrogen bonds [62,63]. The SIRAH model for DNA has been demonstrated to accurately replicate experimental structures, breathing dynamics, and conformational transitions of DNA molecules with a level of accuracy comparable to that of all-atom simulations [62,63]. Finally, it models long-range electrostatic interactions using the particle mesh Ewald method (PME), allowing accounting of the ionic strength and electrostatics explicitly. Since it is the most complete force field for CG and multiscale simulations of complex biological systems, it was utilized in this study.

The simulations were prepared following the protocol recommended by SIRAH developers. Firstly, the reconstructed proteins were protonated at pH 7 using the PROPKA method via the PDB2PQR online web server [64], using Amber force field parameters and atom naming scheme. The fully protonated all-atom structures were then mapped onto CG structures using SIRAH tools [65] and the created PQR files. The proteins with the CG representation were placed in a rectangular box with dimensions of 26.0 × 22.0 × 22.0 nm [3]. The trimers of dimers were aligned to the x-axis with respect to their rigid core (1–149 aa), and the system was solvated using pre-equilibrated water molecules according to the WT4 water model from the SIRAH force field (one WT4 molecule consists of four linked beads and represents eleven all-atom water molecules) [66]. The WT4 water molecules located within 0.3 nm of the protein were removed to facilitate the relaxation of protein side chains during minimization. Local water gradients were resolved in the subsequent equilibration stage. Finally, Na^+ and Cl^- ions were added at a concentration of 37.5 mM (according to experiments reported by Valentic et al. [42]), including neutralizing counterions.

Each system underwent a two-step energy minimization protocol using the GROMACS package. In the first step, the steepest descent algorithm was employed for 5000 steps, with protein backbone atoms harmonically restrained with 1000 kJ/mol·nm. The second step consisted of an unrestrained minimization for another 5000 steps. Equilibration was performed under the NVT ensemble at 300 K, using the V-rescale thermostat [67]. For the first 5 ns, all protein CG beads were harmonically restrained with a force constant of 1000 kJ/mol·nm. In the subsequent 25 ns, only backbone beads of proteins were weakly restrained with a force constant of 100 kJ/mol·nm, while solvent molecules and protein side chains were allowed to move and relax. To prepare the starting structures for subsequent pulling and US simulations, 25 ns of unrestrained MD runs were performed under isothermal-isobaric (NPT) conditions, at 300 K and 1 bar pressure, using the V-rescale thermostat and Parrinello-Rahman barostat with isotropic pressure coupling [68]. A timestep of 20 fs was used in all simulations, and the neighbor lists were updated every 10 steps. Electrostatic interactions were calculated using PME [69], with a direct cutoff of 1.2 nm and grid

spacing of 0.2 nm. Van der Waals interactions were calculated with a cutoff of 1.2 nm.

2.2. Umbrella sampling calculations

In order to estimate protein-protein interactions between HBV Cp with various C-ter lengths and investigate the impact of nucleic acids on the stability of Cp_2 trimer, a series of US simulations was conducted. The reaction coordinate selected was the distance between the center of mass (COM) of the two Cp_2 dimers and the remaining Cp_2 dimer, which was pulled from the trimer (see Fig. 1c). To generate starting configurations for the respective US windows, pulling simulations were conducted using the final snapshot after MD simulations. The pulling simulations involved moving the Cp dimer away from the trimer of dimers along the vector that connects the COM distances of the two groups. During pulling, the two dimers were restrained harmonically, while a force was applied to separate the third dimer. The applied pulling force was 500 kJ/mol·nm, with a pulling rate of 0.001 nm/ps. This force constant was carefully selected to facilitate a gradual protein movement during simulations, thereby minimizing structural deformations and preventing the adoption of unnatural configurations. Initial US windows were taken from the snapshots obtained during the pulling simulations. The US windows were created with an asymmetric distribution, where a spacing of 0.0625 nm was used for the first 35 umbrella windows (i.e. up to COM of ~8 nm), and a spacing of 0.125 nm was used for the remaining windows. A total of 101 US windows per each US run were generated, spanning a distance range from the initial COM distance R_0 to approximately $R_0 + 5$ nm. The adequacy of this sampling is demonstrated by US histograms depicted in Fig. S1. In each window, the system was equilibrated for 3.5 ns using the NPT ensemble at 300 K and 1 bar, followed by a 100 ns MD run using the NPT ensemble with a V-rescale thermostat and Parrinello–Rahman barostat. A bias harmonic potential of 1500 kJ/mol·nm was applied in each window. The force constant was chosen to optimize histogram overlap (see Fig. S1). While exploring force constants from 1000 to 1750 kJ/mol·nm, we found that 1500 kJ/mol·nm produced superior overlap for all trimer-DNA US simulations. The potential of mean force (PMF) was calculated by applying the weighted histogram analysis method (WHAM) to remove the effect of applied bias [70].

To investigate the impact of nucleic acids on the strength of protein interactions in the Cp_2 trimer, we performed simulations with a DNA molecule (PDB code 1hw2) attached to the trimer. To find the possible binding of the DNA to the trimer, the DNA molecule was firstly docked onto the trimer of dimers using the HDOCK web server (available at http://hdock.phys.hust.edu.cn/) [71]. 100 best docking structures with the best docking score, representing the best binding affinity in the system, were simulated using MD for 50 ns each. The binding energy between the DNA and the Cp_2 trimers was calculated using the Molecular Mechanics Generalised Born Surface method (MM-GBSA) [72], and analyzed using the gmx_MMPBSA analysis tool [73]. The binding energy between the Cp_2 trimer and DNA was determined using a single trajectory approach, by subtracting the total molecular mechanics and solvation energy of the complex from the total energy of the single components. The best five structures of the Cp_2 trimer-DNA complex were used for US simulations. In this case, the Cp dimer was separated from the trimer of dimers as described above, but with the DNA molecule attached in the vicinity to determine its impact on trimer stability. The DNA molecule was not part of any pulling group and was not restrained during pulling and US simulations. The summary scheme of the MD simulations and US calculations performed for the Cp_2 trimer-DNA complexes is depicted in Fig. 2.

3. Results and discussion

To understand the subtle interactions, impacting VLP purification and molecule loading, we employed US calculations to investigate the

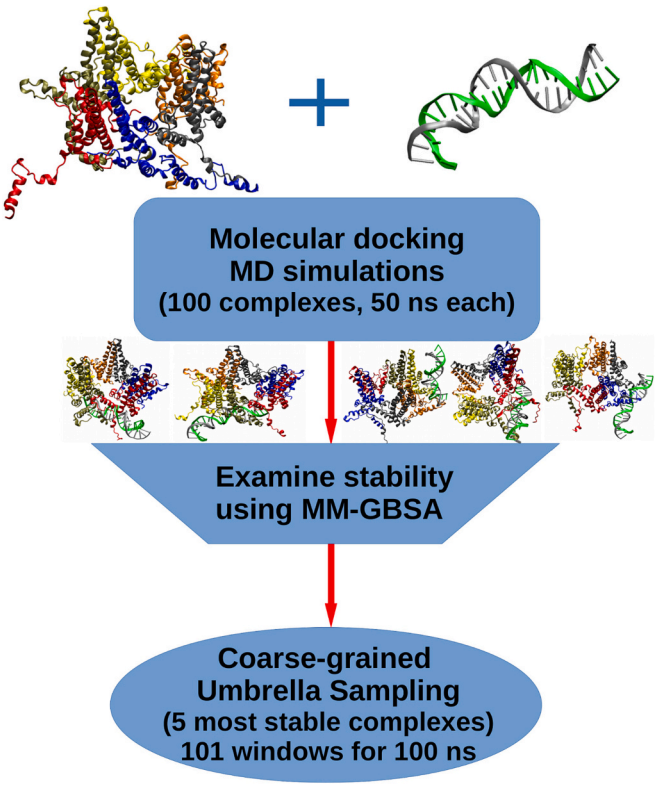


Fig. 2. Schematic representation of the MD simulations and US calculations performed for each of the Cp_2 trimer-DNA complexes. Five parallel US runs for the five selected complexes for trimers comprising $Cp_2$154, $Cp_2$157, $Cp_2$164, $Cp_2$167, and $Cp_2$183 with bonded DNA were simulated.

stability of the trimer of dimers, known as a crucial intermediate nucleate formed during capsid assembly or disassembly [24,36], in the presence and absence of the genetic material. Here, the DNA molecule is considered due to the parameterization of the CG force field applied. We utilized six Cp sequences (Cp149, Cp154, Cp157, Cp164, Cp167, Cp183), known for their capacity to generate VLPs for delivering nucleic acids [12,42], as model systems. The addition of residues at the C-ter domain, from Cp149 to Cp183, results in a change in the total protein charge (as shown in Fig. 1b), with the Cp149 dimer having a charge of -14e (i.e. one Cp is -7e charged) and the full length Cp183 dimer having a charge of +16e. This shift from negative to positive charge is due to the addition of positively charged arginine residues, which are abundantly present in the flexible 150-183aa C-ter located inside the capsid. Such a change will significantly impact protein-protein interactions (PPIs) related to trimer formation. Therefore, to explore it, we first calculated the free energy of binding between the Cp dimer within the trimer of dimers. We achieved this by separating Cp_2 from the dimeric Cp units forming a trimer, as illustrated in Fig. 1c. Considering various core protein lengths enables us to estimate the change of PPIs driving assembly.

3.1. Binding energy between core proteins

Fig. 3a shows the PMF of Cp_2 binding in the trimer formed by core proteins of different C-ter lengths. Dimers of all six proteins showed a negative free energy difference between the assembled and disassembled states, indicating overall attractive PPIs, even upon the change in the Cp charge, and thermodynamic favorability of assembly. Calculated binding energies and associated statistical uncertainties obtained from US simulations (Fig. 3a) are summarized in Table S1. For the Cp with the shortest sequence length (i.e. Cp149), the free energy of binding obtained in the US is -7.90 ± 0.78 kcal/mol. To corroborate the

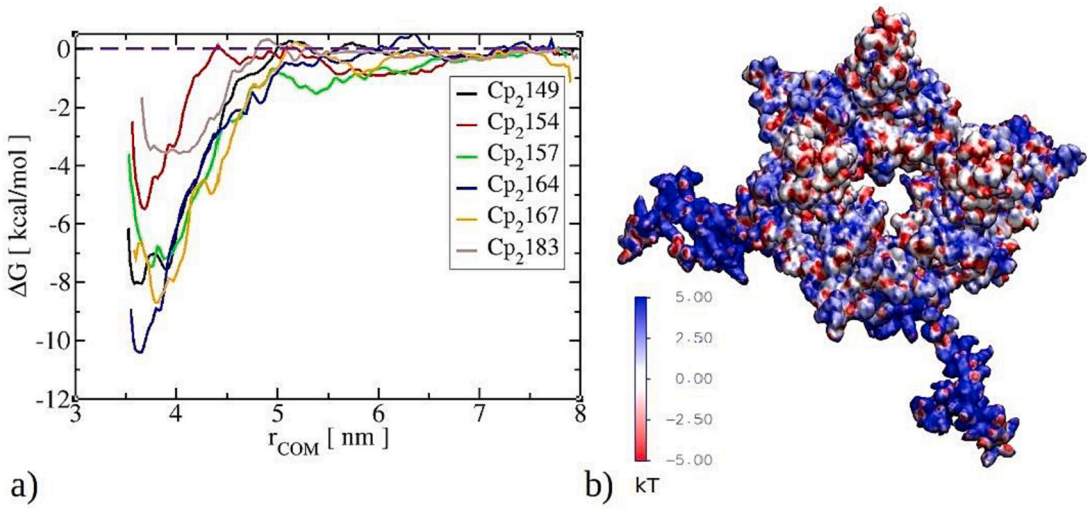


Fig. 3. a) Free energy of Cp_2 binding in the Cp_2 trimer obtained by the US technique using the SIRAH CG force field at 300 K and atmospheric pressure. Cp_2 comprising Cp of different lengths: Cp149, Cp154, Cp157, Cp164, Cp167, Cp183, were considered. b) Electrostatic map of $Cp_2$183 trimer. The electrostatic potential map (in kT, calculated using Adaptive Poisson-Boltzmann solver), is visualized on the three-dimensional surface of the Cp_2 trimer. Negatively charged regions, indicated by negative isocontours, are represented in red, while positively charged regions are depicted in blue.

binding energy value of Cp149 derived from our simulations, we compared it with values documented in the literature. Ceres et al. estimated the effective subunit-subunit energy of VLP149 to be −3.1 kcal/mol at 0.15 M NaCl and up to −4 kcal/mol at 0.7 M NaCl by applying a thermodynamic model to analyze the capsid assembly kinetics [21]. Similarly, Chevreuil et al. estimated the Cp_2 interaction to be −6 kbT (−3.56 kcal/mol) by fitting a theoretical model to capsid melting curve [24]. Asor et al. also reported values derived from experimental data, ranging from −4.15 kcal/mol at 50 mM NaCl to −5.1 kcal/mol at 500 mM NaCl [28]. The reported binding energies are rather modest, however previous studies have shown that weak subunit-subunit interactions promote efficient capsid growth by enabling faster dimer exchange, lowering the kinetic traps and enabling correction of misassembled structures [20,21]. At the same time, Depta et al. reported an interaction potential between Cp_2 of HBV with the minimum value of −32 kJ/mol (equivalent to 7.64 kcal/mol) using MD with the Martini force field [74]. Here, as well as in the US simulations performed in this paper, the Cp_2 has contacts with two other dimers within the trimer of dimers (see Fig. 1), so the effective subunit-subunit contact should be equal to the half of the binding energy. Therefore, both theoretical models closely resemble experimentally reported values. The slight deviations from experiments may come from two main reasons: i) experimental values for Cp_2 interactions refer to the effective averaged interaction between all 120 subunits participating in the VLP assembly, ii) allostery effects, impacting assembly of intermediates [7,22]. Allostery effects have not been considered explicitly in the theoretical calculations. Finally, we have to point out that despite the relatively high charge of $Cp_2$149 (−14e, Fig. 1b), their PPIs are of modest attractive character. This is caused by the fact that the negatively charged molecular fragments are mostly located at the tips of the dimer spikes and on the upper side of the triangular pore in the trimer of dimers (see Fig. 2b). These regions are effectively shielded by salt counterions.

While VLP149 were intensively studied in the literature and the experimental values of $Cp_2$149 interactions have been reported, there is no data available for proteins with longer C-ter. The US simulations, performed in the present paper, permit the evaluation of the binding energy of other Cp, demonstrating that PPIs between Cp_2 with various C-ter are significantly perturbed by the addition of positively charged arginine amino acids (see Fig. 2a). Upon the elongation of the C-ter and partial Cp charge neutralization in $Cp_2$154 and $Cp_2$157 (see Fig. 1b), the binding energy does not become more attractive. In contrast, the binding energy of $Cp_2$154 in the trimer lowers (in the absolute value) to ca. -5.50 ± 0.59 kcal/mol since new repulsion interactions induced by the positive charges in the C-ter modulate the overall binding. When C-ter is elongated to Cp157, the charged amino acids are moved more towards the capsid interior, allowing attractive PPI on the protein interface, improving the binding up to −7.34 ± 0.51 kcal/mol. The binding energy is similar as for $Cp_2$149, however the COM distance of the interaction is slightly shifted, i.e. from 3.6 nm in $Cp_2$149 trimer to 3.8 nm in $Cp_2$157 trimer.

The Cp164 dimer is the least charged from all considered proteins. The decrease in the electrostatic repulsion interaction allows this dimer to bond in the trimer for the best with the binding energy of −10.50 ± 0.58 kcal/mol. For the processability of the VLPs, comprising these Cp, such behavior means rather high capsid stability, therefore, more severe experimental conditions for the VLP purification via disassembly [42]. The low charge of Cp164 (+1 charge, close to the isoelectric point) clearly contributes to the highest stability of the trimer of dimers. This is evident as Cp167, with additional positive charges at the C-terminus, exhibits more repulsive energy compared to Cp164 (i.e., −8.70 ± 0.79 kcal/mol). The naturally occurring $Cp_2$183 has the lowest binding energy in its trimer out of all six Cp_2 considered, i.e. the binding energy of the Cp_2 in the trimer of $Cp_2$183 is −3.50 ± 0.44 kcal/mol. The free energy is still negative, indicating that the overall PPIs are weakly attractive, although not sufficient to induce efficient self-assembly into VLPs due to highly charged long and flexible C-ter. This agrees with experimental observations, which reported that higher ionic strengths (>0.25 M NaCl) are needed to screen electrostatic interactions and initiate efficient VLP183 assembly, either without or with nucleic acids, giving empty or nucleic acids encapsulated capsids [13]. In naturally occurring capsids, such electrostatic screening takes place via binding with nucleic acids. However, it was reported that over 90 % full-length Cp183 capsids are empty under in vivo conditions [75,76], and the solubility of Cp183 dimers is low [13]. This highlights the dynamic and adaptable nature of Cp proteins. Less attractive PPIs for $Cp_2$183 should be better for their processing in the VLP purification, but this is not the case even in high salt conditions due to strong interactions with DNA or RNA from the host expression system, requiring optimizations in the disassembly and VLP purification from the host nucleic acids [42]. We will discuss this process in more detail in the next sections.

Finally, to delve into PPIs within Cp_2 trimers, we analyzed Lennard-Jones, Coulomb and solvation interaction energy changes during MD

simulations, which preceded pulling and extensive US simulations (see Fig. S2). The elongation of C-terminus leads to stronger LJ interactions as protein size increases, and this is observed before trimer disassembly. At the same time, Coulomb interactions vary with C-terminus length, demonstrating substantial repulsion in Cp_2149 and Cp_2183 trimers, and less repulsion for other trimers. Despite the importance of the Coulomb energy, it may not fully reflect electrostatic-based PPIs in solution conditions, where water molecules and salt ions screen these interactions. Thus, solvation energy was also calculated to provide a more comprehensive understanding. It incorporates the screening effects of water and salt ions, thereby enhancing the insight into attractive binding for Cp proteins with larger charges (see Fig. S2b). These results suggest that processing Cp types may rely on changes in ionic strength, modulating electrostatic interactions, and affecting solvation energy values. Still, one should consider that SIRAH, as a CG force field, possesses limitations in handling interactions that demand atomic-level precision, e.g. when water molecules in a specific way modulate solvation. In addition, coarse-graining may limit the accuracy of the reproduction of native contacts in protein-protein interfaces. Furthermore, the electrostatic binding energy, obtained as the sum of Coulomb and solvation energy

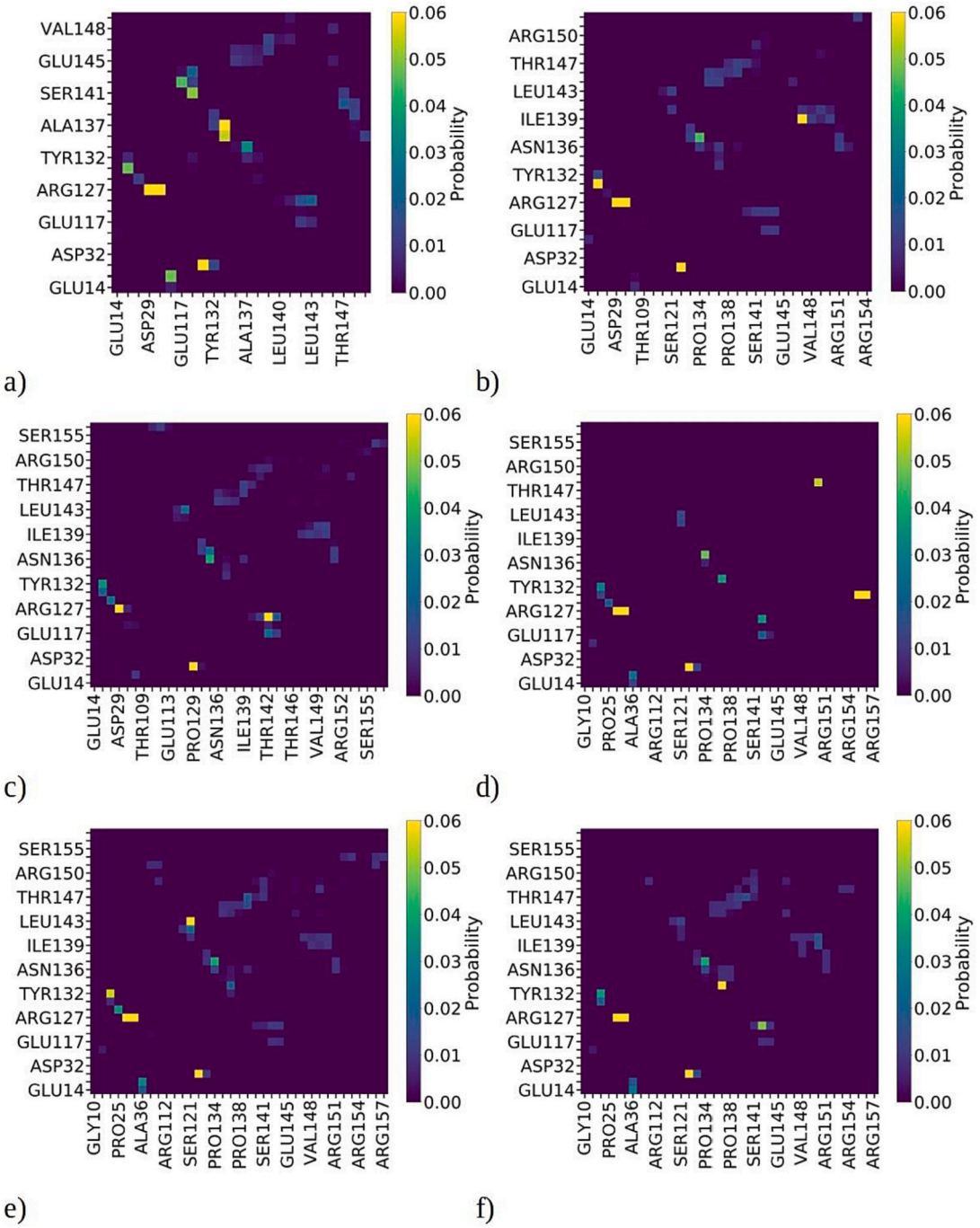


Fig. 4. Probability of maintaining native residue-residue contacts during Cp_2 dimer disassembly from trimers of dimers for a) Cp_2149, b) Cp_2154, c) Cp_2157, d) Cp_2164, e) Cp_2167 and f) Cp_2183. Residues that remain in contact for longer durations show an increased number of counts in contact maps, accompanied by a corresponding higher probability. The data, normalized by the total number of counts, were generated by combining results from three independent US runs for each case. During data analysis statistical significance of each trajectory was taken into account by reweighting the effect of applied bias. Due to the visual data presentation, only every third residue is marked on axes.

(considering only 100 snapshots), is consistently repulsive in all cases, but it does not include correct sampling and entropy effects (included in US simulations explained above) and was calculated only over 25 ns MD for demonstration purposes. However, it supports the observations known from experiments, i.e. electrostatic interactions play a crucial role in HBV VLPs capsid assembly and disassembly [18].

3.2. Residue contacts during disassembly of Cp_2 trimer

To understand the key residues participating in the PPIs during assembly or disassembly of the trimer, we analyzed the evolution of residue-residue contacts between Cp_2 dimers in the trimer during the simulated disassembly process (i.e. US simulations). For that, we have

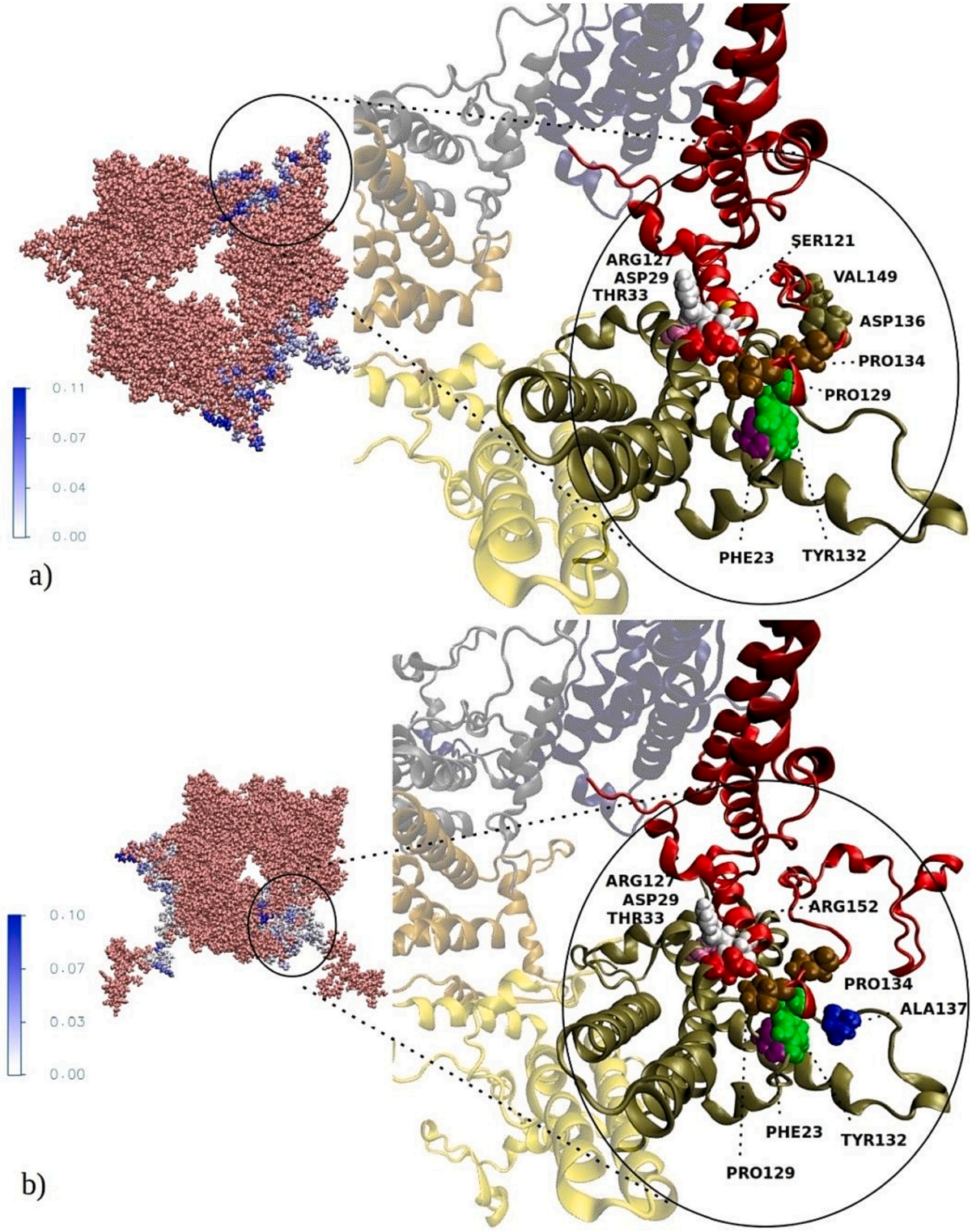


Fig. 5. Illustration of residues, which maintain contact between Cp dimers in the trimer of dimers with the highest probability for: a) Cp_2149 and b) Cp_2183. On the left side, the whole trimer of dimers is depicted, with residues of high contact probability colored in blue. On the right side, relevant interacting parts (i.e. hotspots) are magnified to provide more details.

pinpointed the residues from different Cp_2 dimers that were in close contact within the initial structure at the start of the simulations. Subsequently, we closely observed the evolution of these contacts, monitoring whether they maintained their proximity (residues staying close) or if they broke apart (residues moving apart) as the Cp_2 dimers separate from the trimer structure. Residues were considered to break contact if the distance between their COMs exceeded 6 Å. The residue pairs that have the largest number of counts (instances, where they remain in close contact, i.e. below 6 Å) are considered as hotspot interactions. Therefore, these pairs represent the residues that interact for the longest time (i.e. exhibiting larger number of counts) during disassembly and, therefore, play a crucial role in maintaining the overall stability of the trimer and, probably, the capsid. Residues were considered to maintain contact if the distance between their COMs remained below 6 Å.

Fig. 4 shows contact maps and the probability of preserving native contacts among individual residues within the Cp_2 trimer, which is composed of six Cp variants under investigation. These contacts represent the native interactions present in the initial structure, which includes the trimer of dimers, and they are also found within the entire capsid. The position of residues that maintain a high and low number of contacts in trimers out of $Cp_2$149 and $Cp_2$183 are illustrated in Fig. 5 in blue and salmon red, respectively. The inset pictures show some of the most frequently occurring residues, which participate in native and non-native contacts as explained below. For Cp149, the top five residue pairs that maintained contacts for the longest duration during simulations are: ARG127-ASP29, PHE23-PRO129, ARG127-THR33, ALA137-PRO134, and ASN136-PRO134 (see Table S2 in the Supporting Information, SI). The top-ranked residue pair, ARG127-ASP29 (see Fig. 5a), consists of the positively charged arginine and negatively charged aspartic acid, which are held together by electrostatic interactions. These two residues are located in a hydrophobic and uncharged environment of the protein, further enhancing the strength of the electrostatic interaction between them. ARG127 also maintains a strong contact with THR33, which may be reinforced by its interaction with ASP29. The remaining residue pairs within the top 5 (PHE23-PRO129, ALA137-PRO134, ASN136-PRO134), as well as those within the top 10 (see Table S2), are hydrophobic and uncharged, with the exception of SER141 and GLU117. This aligns with previous studies, which have reported that capsid binding is primarily driven by hydrophobic forces resulting from the burial of apolar residues at subunit-subunit interfaces [7]. Moreover, an experimental study by Rat et al. demonstrated that the deletion of residues 124–135 of Cp had a dramatic impact on HBV capsid assembly, leading to the formation of misassembled oligomers and monomers [77]. Indeed, the majority of the top 10 residue pairs observed in our simulations are located within this region.

When it comes to non-native contacts formed during disassembly, the residues that establish the largest number of contacts are TYR132, ASN136, THR142, and VAL149 (see Fig. S3). This observation aligns with the findings of the all-atom MD study of Cp149 capsid disassembly conducted by Ghaemi et al. [45], which determined that the residues most readily breaking from native contacts are TYR132, ASN136, ALA137, and VAL149. Finally, experimental studies showed mutation of TYR132 to alanine causes a deficiency in capsid assembly, although it can undergo co-assembly if mixed with wild-type proteins [32,78,79].

When considering Cp variants with longer C-terminal regions, the primary trend aligns with that of Cp149. For instance, Cp154 shares the same top 3 residue pairs as Cp149, with the additional residue pairs being PRO129-PHE23 and ILE139-THR147 (see Table S2). ARG127-ASP29 is ranked as the first for Cp157, the second for Cp164 and Cp167, and the fourth for Cp183. This ranking suggests that the strongest electrostatic repulsion introduced by the C-terminus of Cp183 makes it more susceptible to breaking this contact. Furthermore, the occurrence of the PHE23-PRO129 pair is consistent across all six cases, ranking it as the first for Cp183, the second for Cp149, Cp154, and Cp157, and the fourth for Cp164. We acknowledge that the differences observed are likely a result of our study focusing on the trimer of dimers, which may not capture the full statistical behavior of the entire capsid. When it comes to residues establishing non-native contacts, TYR132, THR146, and THR147 are observed the most often for Cp154, while TYR132, PRO135, ARG150, and ARG151 are observed for Cp157, and similar patterns are observed for Cp164, Cp167, and Cp183 (see Fig. S3). It's worth noting that residues from the nucleic acid binding region, such as Arg150 and Arg154 for Cp157, Arg150 and Arg151 for Cp164, Arg156 for Cp167, and Arg150 and Arg151 for Cp183, participate in non-native contacts too. Despite the electrostatic repulsion introduced by the arginines between dimers, their extensive contacts with other residues not only defy the repulsive forces but also confer resistance to structural disruption, ultimately contributing to local stabilization.

3.3. Binding of DNA to Cp_2 trimers

Purification and further loading of capsids, as well as the ability to effectively deliver and release cargo at the appropriate site within the organism, depends not only on the PPIs between Cp with different length of C-ter, but also on the presence of nucleic acids. Therefore, we aimed to study how the stability of trimers of Cp_2 changes when nucleic acids are attached to them. Since the DNA molecule may interact with the Cp_2 during assembly in a diverse way, we initially performed the docking simulations of DNA to proteins. The exemplary protein-DNA complex is visualized in Fig. 6a.

Per each of the Cp type, 100 of the most favorable protein-DNA docked complexes (i.e. ranked by the highest docking score) were further simulated using CG-MD simulations. Fig. 5b-f shows the calculated binding energies for each of the 100 MD simulations of 50 ns for each Cp with different C-ter length. The majority of the trimer-DNA complexes made out of $Cp_2$154 exhibit a repulsive character of the binding energy ranging from +1 to +36 kcal/mol (Fig. 6b). Approximately half of the complexes fall within the range of +20 kcal/mol and only six of the complexes (out of 100) exhibit low attractive binding energy in the range of −5 kcal/mol. The average binding energy over 100 MD simulations in this case is +14.88 ± 8.38 kcal/mol. This observation clearly indicates the low affinity of this Cp (with the short C-ter) to bind nucleic acids. Similar conclusions can also be made for the protein-DNA interactions occurring in $Cp_2$157 trimer-DNA complexes, depicted in Fig. 6c. However, compared to Cp154, there is a more pronounced tendency for less repulsive binding in the complexes, i.e. in the range between +10 to +20 kcal/mol (i.e. the average binding energy of 12.16 ± 8.11 kcal/mol). Here, only twelve complexes (12 %) exhibit attractive binding energy within the range of −3 to −10 kcal/mol. Standard errors in the energies calculated are mostly below 5 % (see Fig. S4), however higher deviations in energies occur for the $Cp_2$157 case.

With the addition of more positively charged residues to the C-terminus domain (i.e. the Cp157 dimer changes the charge from −4 to +2 when Cp164 dimer is considered, see Fig. 1b), a higher proportion of trimer-DNA complexes exhibit an attractive binding energy. For the $Cp_2$164 trimer-DNA complex (see Fig. 6d), twenty-four complexes exhibit an attractive binding energy ranging from −0.5 to −20 kcal/mol. Although complexes with repulsive binding energy are still present, the average binding energy of attractive complexes is −7.87 ± 5.78 kcal/mol, clearly indicating the more preferred binding than is seen for shorter Cp. The highest attractive binding energy was obtained for $Cp_2$167 and $Cp_2$183 trimer-DNA complexes, as shown in Fig. 6e,f. Here, the average binding energy is of −9.75 ± 6.67 kcal/mol and −12.41 ± 8.44 kcal/mol, considering the attractive and repulsive nature of the interactions. Forty-one and sixty-nine complexes exhibit an attractive binding energy within up to −28 kcal/mol and −37 kcal/mol for $Cp_2$167 trimer-DNA and $Cp_2$183 trimer-DNA, respectively. Since the Cp in these cases are highly positive charged (+8 and +16 for $Cp_2$167 and $Cp_2$183) and the charge is localized on the unstructured, flexible C-ter, they tend to bind nucleic acids efficiently, what is also commonly known in the literature [12]. The strongest binding energy has been found for the

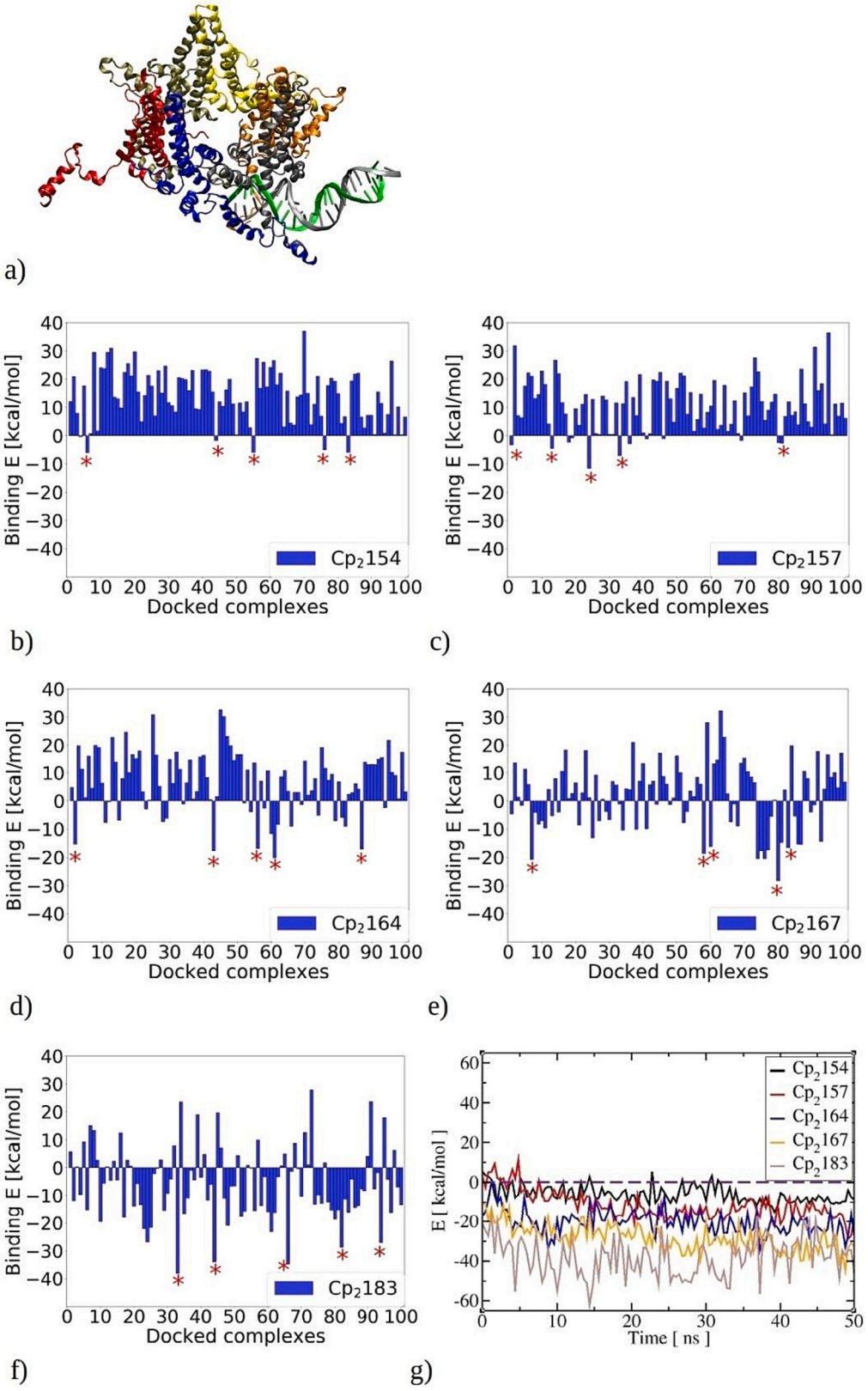


Fig. 6. a) Example of the docked Cp_2 trimer-DNA complex. Cp183 was used for the visualization with each Cp monomer marked with a different color. The DNA molecule is represented with green and gray double strands. Binding energy between the Cp_2 trimer and the DNA molecule for 100 CG-MD simulations starting from 100 DNA docked structures with b) Cp_2154, c) Cp_2157, d) Cp_2164, e) Cp_2167, and f) Cp_2183 trimers calculated by MM-GBSA method. g) The change of the binding energy throughout the MD simulation of the most attractive trimer-DNA complex for each of the Cp_2. MD runs marked with the red star on panels b-f indicate protein-DNA complexes used for US calculations, described in Section 3.4.

trimers comprising full-length Cp_2183 proteins. Here, the attractive binding energy reaches −37 kcal/mol.

To demonstrate the change of the binding energy over the MD simulation, we plot the time evolution of the interaction energy for the most attractive trimer-DNA complexes in Fig. 6g. It affirms the enhanced binding capabilities of Cp with a longer C-ter. However, we see that the binding energy fluctuates significantly over the MD simulation, indicating the flexibility of C-ter and sampling of different conformations. Since the C-ter in Cp183 is the longest and this chain is unstructured, the dynamical changes in the binding energy are the strongest in this case. Finally, we see that the overall binding energy is heavily dependent on the DNA binding pose, therefore, the consideration of 100 separate MD simulations has been applied. It is still an approximation and limits the present study, however due to the computational cost, the necessary trends were captured and they agree with the experimental observations. Indeed, the uncertainty in the calculated binding energy remains

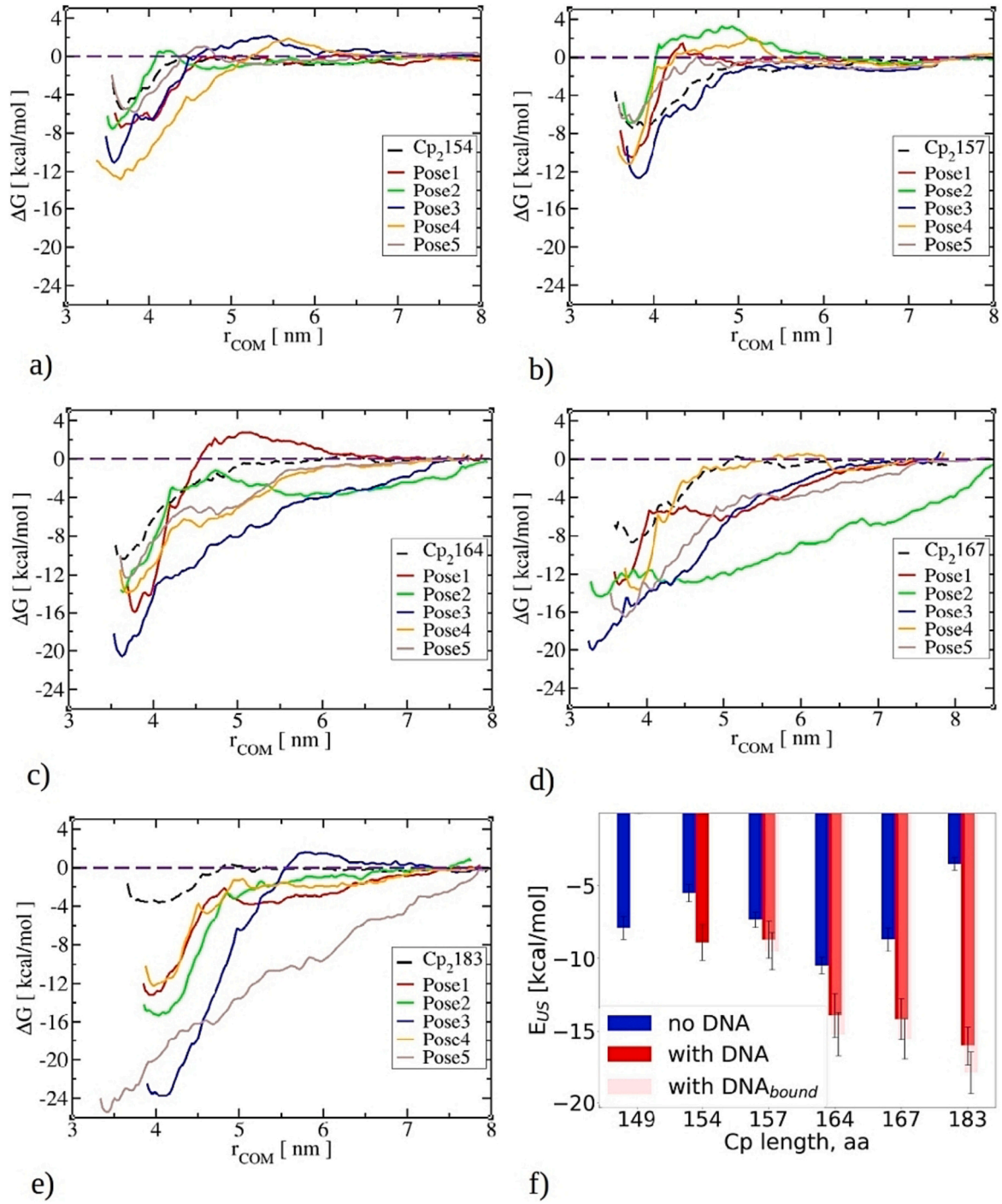


Fig. 7. Disassembly of Cp_2 from Cp_2 trimer in the presence of DNA. Free energy of Cp_2 binding in a trimer obtained by umbrella sampling simulations at 300 K for: a) Cp_2154, b) Cp_2157, c) Cp_2164, d) Cp_2167, and e) Cp_2183. US simulations starting from different poses (see Fig. S5) of DNA attached to the trimer are shown. US simulations of Cp_2 disassembly from Cp_2 trimer without the DNA molecule bonded are marked by black dashed lines. f) The summary of the DNA impact on the PPIs between different Cp in the VLP: the average binding energy between proteins without (in blue) and with (in red) bonded DNA to the trimer. The binding energy between proteins with DNA attached was calculated in two ways: including poses, where no DNA binding was present during pulling (see Fig. S6a, S7a, S8a, S9a) with the values marked in red in panel f and as an average of all five poses, including those, where DNA binds proteins during pulling, i.e. marked as with DNA_{bound}.

noticeable (see Fig. S3). In certain instances, particularly where the binding energy approaches zero due to the near-cancellation of van der Waals and electrostatic interactions, the uncertainty expands to a range between 10 % and 25 %. This phenomenon is largely attributable to the fluctuation of potential terms throughout MD trajectories, especially when they nearly nullify each other. It is important to acknowledge the inherent challenges in simulating protein-DNA complexes, particularly with CG models, where typical MD simulations possess sampling limitations. In summary, the results demonstrate that the addition of the arginine-rich C-terminal region significantly enhances the affinity of Cp for nucleic acids binding, starting from Cp154 and Cp167, which exhibit primarily electrostatic repulsion, and culminating with Cp183, where most complexes become attractive.

3.4. Impact of DNA on interactions between Cp_2 subunits

To estimate the impact of DNA on the PPIs within the assembled Cp_2 trimer, i.e. during the purification of VLPs from the host genome [42], US simulations of Cp_2 disassembly from the trimer-DNA complex have been performed. The simulations were conducted in a manner similar to the trimer disassembly presented above, but here the DNA molecule was attached in the vicinity of the trimer. The poses selected from MD runs (marked with the red star in Fig. 6) are visualized in Fig. S5. Cp149 was not considered for these simulations since it does not possess the C-ter nucleic acids binding unit and it was reported not to bind nucleic acids [80].

Fig. 7 illustrates the calculated free energy profiles for Cp_2 disassembly from the corresponding trimers in the presence of DNA. To provide a point of comparison, we also included the free energy of Cp_2 disassembly from trimers of Cp_2 dimers without DNA attached in the vicinity (as presented in Fig. 3). They are marked by the black dashed line. In the case of trimers comprising Cp154 dimers, the binding energy in the absence of DNA is −5.36 kcal/mol, as determined by the minimum from the PMF curve. However, when the DNA molecule is bonded in the vicinity of the trimer, the binding energy shifts to the values ranging from −5.89 ± 0.98 kcal/mol to −12.70 ± 1.68 kcal/mol, depending on the starting DNA binding pose from docking. The average binding energy over five US simulations with different DNA binding poses is −8.91 ± 1.23 kcal/mol, which is 1.51 times more attractive than the energy in the absence of DNA. Similarly, is observed for trimers of Cp157 dimers: the average binding energy in the presence of DNA is approximately −9.51 ± 1.27 kcal/mol, while PPIs without DNA were of −7.34 ± 0.51 kcal/mol. Here, the stabilization factor is 1.29, as the binding energy in the absence of DNA is slightly more attractive compared to Cp154 proteins due to lower charge of Cp dimers (see Fig. 1b) and less repulsive interactions. Moreover, the average binding energy of $Cp_2$154 and $Cp_2$157 in the presence of DNA is 1.4 kcal/mol more attractive than that of trimers of $Cp_2$149, which lack nucleic acid binding regions. This means that in the presence of nucleic acids, Cp154 and Cp157 should assemble more easily. Similar observation was reported in capsid disassembly experiments conducted by Valentic et al. [42]. In their experiments, Cp154 and Cp157 capsids exhibited slightly lower dimer yields compared to Cp149 capsids, suggesting that Cp154 and Cp157 capsids are more stable [42].

The stabilizing effect of DNA on the PPIs in the trimer is particularly significant for trimers containing longer C-ter that bind nucleic acid, such as those formed from $Cp_2$164, $Cp_2$167, and $Cp_2$183. In the absence of DNA, $Cp_2$164, $Cp_2$167 and $Cp_2$183 exhibit binding energies of −10.5 ± 0.58 kcal/mol, −8.70 ± 0.79 kcal/mol and −3.50 ± 0.44 kcal/mol, respectively (see Fig. 3 or plots in dashed black in Fig. 7c,d,e). However, in the presence of DNA, the average binding energies increase. This may be partially connected to the fact that for some poses the DNA binds the protein-protein interface, as seen from the visual inspection of US trajectories (see Fig. S6-S9). The average value of binding energy still increases, i.e. to −13.94 ± 1.51 kcal/mol, −14.27 ± 1.39 kcal/mol and −16.07 ± 1.34 kcal/mol considering poses, when DNA is differently located. It is the average value calculated using pose 1, 2, 4, 5 for $Cp_2$164 and $Cp_2$167, pose 1, 2, 3, 4 for $Cp_2$183, i.e. poses where DNA completely detaches from dimer. This results in stabilization factors of 1.33 and 1.64 for $Cp_2$164 and $Cp_2$167, respectively, in the presence of DNA. Uncertainties of these protein-DNA binding energies are below 2 kcal/mol, i.e. between 6 and 13 % (see Table S1). The strongest stabilizing effect is seen in $Cp_2$183 trimers, with an attractive energy 4.59 times higher in the presence of DNA compared to its absence. The presence of the DNA molecule screens the repulsive electrostatic interactions caused by the positively charged C-terminus, thus, enhancing the stability of the trimer. In addition, proteins with additional positively charged residues at the C-terminus establish stronger DNA interactions, leading to higher energies necessary to disassemble the complex.

We have to note that the DNA binding poses have an effect on binding energy, as consistent with the rationale explained in Section 3.3 and Fig. 6. For four core protein lengths (Pose 3 of Cp157, Pose 3 of Cp164, Pose 3 of Cp167, and Pose 5 of Cp183), the most attractive binding energy was consistently observed when the C-terminal region of Cp_2 maintained direct contact with DNA, even at center of COM distances of 9 nm (see Fig. S6b, Fig. S7b, Fig. S8b, and Fig. S9b). Therefore, these PMF curves exhibit lower minima and different shape compared to all other curves when the C-terminus of Cp_2 did not maintain contact with DNA. This indicates a possible distinct mechanism of disassembly as a function of the C-terminus. While we aimed to separate dimers farthest from the DNA, in some instances, direct contact could not be avoided, suggesting that dimers are more prone to disassemble in regions with lower nucleic acid density. Since such local heterogeneities in nucleic acid density in the capsid may occur, which may impact the binding energies measured experimentally, the full capsid structure remains challenging to simulate using the precise and computationally demanding coarse-grained representation employed in the present study. More simplified CG methods can be used; however, such analyses go beyond our current objectives.

Several experimental studies have reported that Cp183 capsids loaded with nucleic acids are more stable than Cp149 capsids, which aligns with the observations in our study [11,42]. Interactions with DNA are stronger and electrostatic repulsion between Cp dimers is screened, therefore, they are known to be less soluble [13], necessitating cumbersome processing conditions. Indeed, study of Valentic et al. reported lower Cp183 dimer yield upon capsid disassembly to purify it from host nucleic acids compared to Cp149 [42]. In the same study, Cp164 and Cp167 capsids were demonstrated to result in the lowest yield of dimers, suggesting their capsids were the most stable. Higher strength of PPIs for these Cp is also visible from the US calculations performed in the present study (see Fig. 7f), which is now also explained using theoretical calculations. Jiang et al. proposed a thermodynamic model for HBV capsids, revealing that Cp164 capsids exhibit greater stability compared to Cp183 capsids, each with an optimal amount of loaded nucleic acids [56]. Their study examined the stability of each capsid in relation to the amount of loaded nucleic acids, identifying the optimal nucleic acid amount for each C-terminal length and its effect on capsid stability. Comparable PPI strength between $Cp_2$164 and $Cp_2$183 in the presence of DNA is also visible from our calculations. However, full-length dimers tend to bind DNA better than Cp164. This difference may be connected to the fact that we have not simulated optimal loading of capsids with DNA, but demonstrated results based on one DNA molecule bonded to different trimers of Cp dimers. Indeed, Sominskaya et al. reported that HBV capsids encapsulate varying amounts of nucleic acids depending on the core protein length [12].

The efficiency of VLP (dis)assembly depends on the interplay of interactions between Cp and Cp with the nucleic acid. From the calculations performed in the present study, we observe that ideal Cp used for VLP production for binding therapeutic DNA or RNA, enabling reasonable capsid loading, but still enabling ease of processing, would be Cp154 or Cp157. Both of these constructs possess attractive PPIs between dimers (which are more stable than the full-length Cp183) and

should not bind DNA too strongly to permit easier purification. At the same time, in comparison to Cp149 they bind DNA molecules, therefore, might be optimal for HBV VLP production for gene therapy. Since molecular bases of binding between HBV core proteins (with different C-ter) and nucleic acids have not been reported in the literature and experimental trends could not be explained in detail, our study quantifies the strength and interplay of protein-protein and protein-DNA interactions as a function of the C-ter lengths. Using one of the most complete force fields for CG and multiscale simulations of complex biological systems, we could reveal hotspot contacts and protein binding that modulate trimer stability and its disassembly in the presence of DNA.

4. Conclusions

In this study, we conducted a comprehensive investigation into the stability of hepatitis B capsid fragments, specifically trimers of Cp_2 dimers, chosen as representative structures for understanding capsid disassembly trends in the presence and absence of nucleic acid. Our focus was on six core proteins, each with varying lengths of the natural nucleic acid binding region (149aa to 183aa), to elucidate the changes in protein-protein interaction strength as a function of C-terminus lengths. To achieve this, we employed extensive molecular dynamics and umbrella sampling simulations utilizing the SIRAH coarse-grained force field.

Initially, we scrutinized protein-protein interactions between Cp_2 in the respective trimers by analyzing binding energy through umbrella sampling simulations. The calculated binding free energies for Cp_2149, Cp_2154, Cp_2157, Cp_2164, Cp_2167, and Cp_2183 were −7.9 kcal/mol, −5.5 kcal/mol, −7.3 kcal/mol, −10.5 kcal/mol, −8.7 kcal/mol, and −3.5 kcal/mol, respectively. All with a relative uncertainty below 1 kcal/mol. It clearly indicates that flexible C-terminus, located mostly in the interior of the capsid, modulates interactions between proteins, significantly impacting capsid assembly and stability. Notably, the binding energy of Cp_2149 agreed well with previously reported values from both experimental and theoretical studies. Binding energy values for other constructs are obtained in the present work for the first time. The observed perturbations in binding energies of different Cp were primarily attributed to electrostatic effects, stemming from the increase in the total charge of Cp_2 as C-terminus length increased (from −14 to +16). In addition, we monitored the breaking of contacts between residues during disassembly of the Cp_2 trimer, identifying hotspot contacts that modulate trimer stability. Relevant residue contacts were revealed to occur between ARG127-ASP33, ARG127-THR33, and PHE23-PRO129 over all Cp types considered. They are located in helical regions at the central dimer-dimer interface. Therefore, our findings suggest that the flexible C-terminus does not significantly influence the core assembly domain, with hydrophobic interactions being the driving force, but electrostatic interactions between other C-ter. This is consistent with previous experimental studies.

To account for the impact of DNA on PPIs during Cp_2 trimer disassembly, molecular dynamics simulations of 100 Cp_2 trimer-DNA complexes, followed by umbrella sampling simulations, have been performed. Our results indicated that the affinity to DNA increased with the length of the C-terminus. Specifically, Cp_2154 trimer-DNA and Cp_2157 trimer-DNA complexes exhibited mostly repulsive nature of DNA-bonded complexes, while more pronounced attractive character and higher affinity to bind DNA increased up to Cp_2183 trimer-DNA. Obtained results revealed that the presence of DNA stabilized trimers, i.e. PPIs between assembled Cp, possessing a nucleic acid binding region, i.e. the average binding energy between Cp dimers increases to −9.21 kcal/mol, −9.39 kcal/mol, −15.03 kcal/mol, −14.19 kcal/mol and −17.92 kcal/mol for Cp154, Cp157, Cp164, Cp167 and Cp183, respectively, with a relative uncertainty of 5–17 %, dependent on the protein-DNA complex. This stabilization effect was most pronounced for Cp183, where binding energy became nearly five times more attractive in the presence of DNA. The DNA's stabilizing effect was attributed to attenuating electrostatic repulsion among positively charged C-termini and establishing direct contact with the disassembly dimer. These results correlate to the lower dimer yields of capsid disassembly, reported in experiments, as more attractive binding energy makes disassembly more difficult. With the present study, we have demonstrated the delicate balance between protein-protein and protein-nucleic acid interactions in the stability of virus-like particles. We have explained that excessively stable structures may pose experimental challenges, while insufficient binding energy may hinder their formation. This nuanced understanding is crucial for the rational design and practical application of VLPs in various fields.

CRediT authorship contribution statement

Srdjan Pusara: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Wolfgang Wenzel:** Writing – review & editing, Resources, Methodology, Funding acquisition, Conceptualization. **Mariana Kozlowska:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

The datasets generated during and/or analyzed during the current study are available via the NOMAD repository under https://doi.org/10.17172/NOMAD/2024.02.23-1.

Acknowledgements

This research has been funded by Deutsche Forschungsgemeinschaft (DFG) SPP priority programme DiSPBiotech (SPP1934, project number WE1863/30-2). W. W. and M. K. acknowledge funding by the Carl Zeiss Stiftung within the Center for Synthetic Genomics. We acknowledge support by the KIT-Publication Fund of the Karlsruhe Institute of Technology.

During the preparation of this work the authors used chatGPT 3.5 in order to improve clarity and readability of sentences. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2024.130365.

References

[1] C. Qian, X. Liu, Q. Xu, Z. Wang, J. Chen, T. Li, Q. Zheng, H. Yu, Y. Gu, S. Li, N. Xia, Vaccines 8 (2020) 139.
[2] A. Zeltins, Mol. Biotechnol. 53 (2013) 92–107.
[3] M.J. Rohovie, M. Nagasawa, J.R. Swartz, Bioeng. Transl. Med. 2 (2017) 43–57.
[4] P. Pumpens, E. Grens, Mol. Biol. 50 (2016) 489–509.
[5] C. Trépo, H.L.Y. Chan, A. Lok, Lancet 384 (2014) 2053–2063.
[6] B. Rehermann, M. Nascimbeni, Nat. Rev. Immunol. 5 (2005) 215–229.
[7] A. Zlotnick, B. Venkatakrishnan, Z. Tan, E. Lewellyn, W. Turner, S. Francis, Antivir. Res. 121 (2015) 82–93.
[8] M. Seifer, D.N. Standring, J. Virol. 68 (1994) 5548–5555.
[9] M. Nassal, J. Virol. 66 (1992) 4107–4116.
[10] A. Gallina, F. Bonelli, L. Zentilin, G. Rindi, M. Muttini, G. Milanesi, J. Virol. 63 (1989) 4645–4652.
[11] F. Birnbaum, M. Nassal, J. Virol. 64 (1990) 3319–3330.
[12] I. Sominskaya, D. Skrastina, I. Petrovskis, A. Dishlers, I. Berza, M. Mihailova, J. Jansons, I. Akopjana, I. Stahovska, D. Dreilina, V. Ose, P. Pumpens, PLoS One 8 (2013) e75938.

[13] J.Z. Porterfield, M.S. Dhason, D.D. Loeb, M. Nassal, S.J. Stray, A. Zlotnick, J. Virol. 84 (2010) 7174–7184.
[14] M. Newman, P.K. Chua, F.-M. Tang, P.-Y. Su, C. Shih, J. Virol. 83 (2009) 10616–10626.
[15] L. Selzer, R. Kant, J.C.-Y. Wang, B. Bothner, A. Zlotnick, J. Biol. Chem. 290 (2015) 28584–28593.
[16] C. Kim, C.J. Schlicksup, C. Pérez-Segura, J.A. Hadden-Perilla, J.C.-Y. Wang, A. Zlotnick, J. Biol. Chem. 299 (2023) 105104.
[17] Z.D. Harms, L. Selzer, A. Zlotnick, S.C. Jacobson, ACS Nano 9 (2015) 9087–9096.
[18] W.K. Kegel, P. van der Schoot, Biophys. J. 86 (2004) 3905–3913.
[19] C.A. Lutomski, N.A. Lyktey, Z. Zhao, E.E. Pierson, A. Zlotnick, M.F. Jarrold, J. Am. Chem. Soc. 139 (2017) 16932–16938.
[20] A. Zlotnick, Virology 315 (2003) 269–274.
[21] P. Ceres, A. Zlotnick, Biochemistry 41 (2002) 11525–11531.
[22] A. Zlotnick, S. Mukhopadhyay, Trends Microbiol. 19 (2011) 14–23.
[23] J. Zhou, A. Zlotnick, S.C. Jacobson, Anal. Chem. 94 (2022) 985–992.
[24] M. Chevreuil, L. Lecoq, S. Wang, L. Gargowitsch, N. Nhiri, E. Jacquet, T. Zinn, S. Fieulaine, S. Bressanelli, G. Tresset, J. Phys. Chem. B 124 (2020) 9987–9995.
[25] C. Uetrecht, N.R. Watts, S.J. Stahl, P.T. Wingfield, A.C. Steven, A.J.R. Heck, Phys. Chem. Chem. Phys. 12 (2010) 13368.
[26] L. Selzer, A. Zlotnick, Cold Spring Harb. Perspect. Med. 5 (2015) a021394.
[27] C.A. Lutomski, N.A. Lyktey, E.E. Pierson, Z. Zhao, A. Zlotnick, M.F. Jarrold, J. Am. Chem. Soc. 140 (2018) 5784–5790.
[28] R. Asor, L. Selzer, C.J. Schlicksup, Z. Zhao, A. Zlotnick, U. Raviv, ACS Nano 13 (2019) 7610–7626.
[29] R. Asor, C.J. Schlicksup, Z. Zhao, A. Zlotnick, U. Raviv, J. Am. Chem. Soc. 142 (2020) 7868–7882.
[30] P. Ceres, S.J. Stray, A. Zlotnick, J. Virol. 78 (2004) 9538–9543.
[31] Z. Zhao, J.C.-Y. Wang, C.P. Segura, J.A. Hadden-Perilla, A. Zlotnick, ACS Chem. Biol. 15 (2020) 3124–3132.
[32] A. Patterson, Z. Zhao, E. Waymire, A. Zlotnick, B. Bothner, ACS Chem. Biol. 15 (2020) 2273–2280.
[33] S. Qazi, C.J. Schlicksup, J. Rittichier, M.S. VanNieuwenhze, A. Zlotnick, ACS Chem. Biol. 13 (2018) 2114–2120.
[34] C.J. Schlicksup, P. Laughlin, S. Dunkelbarger, J.C.-Y. Wang, A. Zlotnick, ACS Chem. Biol. 15 (2020) 1708–1717.
[35] S.P. Katen, S.R. Chirapu, M.G. Finn, A. Zlotnick, ACS Chem. Biol. 5 (2010) 1125–1136.
[36] Z. Tan, K. Pionek, N. Unchwaniwala, M.L. Maguire, D.D. Loeb, A. Zlotnick, J. Virol. 89 (2015) 3275–3284.
[37] R.C. Oliver, W. Potrzebowski, S.M. Najibi, M.N. Pedersen, L. Arleth, N. Mahmoudi, I. André, ACS Nano 14 (2020) 10226–10238.
[38] P.K. Chua, F.-M. Tang, J.-Y. Huang, C.-S. Suen, C. Shih, J. Virol. 84 (2010) 2340–2351.
[39] I. Petrovskis, I. Lieknina, A. Dislers, J. Jansons, J. Bogans, I. Akopjana, J. Zakova, I. Sominskaya, Microorganisms 9 (2021) 283.
[40] Y. Zhang, Y. Liu, B. Zhang, S. Yin, X. Li, D. Zhao, W. Wang, J. Bi, Z. Su, Protein Expr. Purif. 178 (2021) 105747.
[41] A. Strods, V. Ose, J. Bogans, I. Cielens, G. Kalnins, I. Radovica, A. Kazaks, P. Pumpens, R. Renhofa, Sci. Rep. 5 (2015) 11639.
[42] A. Valentic, J. Müller, J. Hubbuch, Front. Bioeng. Biotechnol. 10 (2022) 929243.
[43] A. Zlotnick, J.M. Johnson, P.W. Wingfield, S.J. Stahl, D. Endres, Biochemistry 38 (1999) 14644–14652.
[44] J.A. Hadden, J.R. Perilla, C.J. Schlicksup, B. Venkatakrishnan, A. Zlotnick, K. Schulten, eLife 7 (2018) e32478.
[45] Z. Ghaemi, M. Gruebele, E. Tajkhorshid, Proc. Natl. Acad. Sci. 118 (2021) e2102530118.
[46] A. Arkhipov, P.L. Freddolino, K. Schulten, Structure 14 (2006) 1767–1777.
[47] C. Pérez-Segura, B.C. Goh, J.A. Hadden-Perilla, Viruses 13 (2021) 564.
[48] J. Zang, M. Liu, H. Liu, L. Ding, Phys. Chem. Chem. Phys. 24 (2022) 23209–23225.
[49] H. Liu, S. Okazaki, W. Shinoda, J. Chem. Inf. Model. 59 (2019) 5104–5110.
[50] J. Tu, J.J. Li, Z.J. Shan, H.L. Zhai, Antivir. Res. 137 (2017) 151–164.
[51] K. Fujimoto, M. Fukai, R. Urano, W. Shinoda, T. Ishikawa, K. Omagari, Y. Tanaka, A. Nakagawa, S. Okazaki, Pure Appl. Chem. 92 (2020) 1585–1594.
[52] G. Watanabe, S. Sato, M. Iwadate, H. Umeyama, M. Hayakawa, Y. Murakami, S. Yoneda, Chem. Pharm. Bull. (Tokyo) 64 (2016) 1393–1396.
[53] A. Pavlova, L. Bassit, B.D. Cox, M. Korablyov, C. Chipot, D. Patel, D.L. Lynch, F. Amblard, R.F. Schinazi, J.C. Gumbart, J. Med. Chem. 65 (2022) 4854–4864.
[54] J. Kim, J. Wu, J. Chem. Phys. 140 (2014) 235101.
[55] D. Meng, R.P. Hjelm, J. Hu, J. Wu, Biophys. J. 101 (2011) 2476–2484.
[56] T. Jiang, Z.-G. Wang, J. Wu, Biophys. J. 96 (2009) 3065–3073.
[57] J. Kim, J. Wu, Biophys. J. 109 (2015) 1689–1697.
[58] B. Böttcher, M. Nassal, J. Mol. Biol. 430 (2018) 4941–4954.
[59] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, Nucleic Acids Res. 46 (2018) W296–W303.
[60] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, J. Chem. Theory Comput. 4 (2008) 435–447.
[61] L. Darré, M.R. Machado, A.F. Brandner, H.C. González, S. Ferreira, S. Pantano, J. Chem. Theory Comput. 11 (2015) 723–739.
[62] P.D. Dans, A. Zeida, M.R. Machado, S. Pantano, J. Chem. Theory Comput. 6 (2010) 1711–1725.
[63] P.D. Dans, L. Darré, M.R. Machado, A. Zeida, A.F. Brandner, S. Pantano, in: J. C. Setubal, N.F. Almeida (Eds.), Advances in Bioinformatics and Computational Biology vol. 8213, Springer International Publishing, Cham, 2013, pp. 71–81.
[64] T.J. Dolinsky, P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, N.A. Baker, Nucleic Acids Res. 35 (2007) W522–W525.
[65] M.R. Machado, S. Pantano, Bioinformatics 32 (2016) 1568–1570.
[66] L. Darré, M.R. Machado, P.D. Dans, F.E. Herrera, S. Pantano, J. Chem. Theory Comput. 6 (2010) 3793–3807.
[67] G. Bussi, D. Donadio, M. Parrinello, J. Chem. Phys. 126 (2007) 014101.
[68] M. Parinello, A. Rahman, J. Appl. Phys. 52 (1981) 7182–7190.
[69] T. Darden, D. York, L. Pedersen, J. Chem. Phys. 98 (1993) 10089–10092.
[70] J.S. Hub, B.L. de Groot, D. van der Spoel, J. Chem. Theory Comput. 6 (2010) 3713–3720.
[71] Y. Yan, D. Zhang, P. Zhou, B. Li, S.-Y. Huang, Nucleic Acids Res. 45 (2017) W365–W373.
[72] N. Homeyer, H. Gohlke, Mol. Inform. 31 (2012) 114–122.
[73] M.S. Valdés-Tresanco, M.E. Valdés-Tresanco, P.A. Valiente, E. Moreno, J. Chem. Theory Comput. 17 (2021) 6281–6291.
[74] P.N. Depta, M. Dosta, W. Wenzel, M. Kozlowska, S. Heinrich, Int. J. Mol. Sci. 23 (2022) 14699.
[75] X. Ning, D. Nguyen, L. Mentzer, C. Adams, H. Lee, R. Ashley, S. Hafenstein, J. Hu, PLoS Pathog. 7 (2011) e1002255.
[76] J. Hu, K. Liu, Viruses 9 (2017) 56.
[77] V. Rat, X. Pinson, F. Seigneuret, S. Durand, C. Herrscher, R. Lemoine, J. Burlaud-Gaillard, P.-Y. Raynal, C. Hourioux, P. Roingeard, M. Tramier, H. De Rocquigny, J. Mol. Biol. 432 (2020) 3802–3819.
[78] K. Klumpp, A.M. Lam, C. Lukacs, R. Vogel, S. Ren, C. Espiritu, R. Baydo, K. Atkins, J. Abendroth, G. Liao, A. Efimov, G. Hartman, O.A. Flores, Proc. Natl. Acad. Sci. 112 (2015) 15196–15201.
[79] C.R. Bourne, S.P. Katen, M.R. Fulz, C. Packianathan, A. Zlotnick, Biochemistry 48 (2009) 1736–1742.
[80] N. Hillebrandt, P. Vormittag, N. Bluthardt, A. Dietrich, J. Hubbuch, Front. Bioeng. Biotechnol. 8 (2020) 489.