

# CONSTANT PH MOLECULAR DYNAMICS OF PORCINE CIRCOVIRUS 2 CAPSID PROTEIN REVEALS A MECHANISM FOR CAPSID ASSEMBLY

Elvira Tarasova<sup>1,2\*</sup>, Noriaki Okimoto<sup>1,3</sup>, Shanshan Feng<sup>2</sup>, Dmitry Nerukh<sup>4</sup>, Reza Khayat<sup>2\*</sup>, Makoto Taiji<sup>1,3</sup>

<sup>1</sup>Laboratory for Computational Molecular Design, RIKEN Center for Biosystems Dynamics Research (BDR), 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan;

<sup>2</sup>Department of Chemistry and Biochemistry, The City College of New York, New York, New York, USA;

<sup>3</sup>Drug Discovery Molecular Simulation Platform Unit, RIKEN Center for Biosystems Dynamics Research (BDR), 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan;

<sup>4</sup>Department of Mathematics, Aston University, Birmingham, United Kingdom.

Address correspondence to E.T. (email: [etarasova@ccny.cuny.edu](mailto:etarasova@ccny.cuny.edu)) or R.K. (email: [rkhayat@ccny.cuny.edu](mailto:rkhayat@ccny.cuny.edu))

## ORCID

Reza Khayat (0000-0001-6601-2184).

**ABSTRACT:**

Spatiotemporal regulation of viral capsid assembly ensures selection of the viral genome for encapsidation. The porcine circovirus 2 is the smallest autonomously replicating pathogenic virus, yet it remains unknown how PCV2 capsid assembly is regulated to occur within the nucleus. We report that pure PCV2 capsid proteins, in the absence of nucleic acids, require acidic conditions to assemble into empty capsids in vitro. By employing Constant pH Replica Exchange Molecular Dynamics, we unveil the atomistic mechanism of pH-dependency for capsid assembly. The results show that the appropriate protonation configuration for a cluster of acidic amino acids is necessary to properly position the GH-loop for driving capsid assembly. We demonstrate that assembly is prohibited at neutral pH because deprotonation of these residues results in their electrostatic repulsion, shifting the GH-loop to a position incompetent with capsid assembly. We propose that encapsulation of nucleic acid overcomes this repulsion to properly position the GH-loop. Our findings provide the first atomic resolution mechanism of capsid assembly regulation. The findings are useful for development of therapeutics that inhibit PCV2 self-assembly.

## INTRODUCTION

Porcine circovirus (PCV) types 1-3 are 20nm T=1 icosahedral virion that possess circular single stranded DNA genomes and belong to the *Circovirus* genus of the *Circoviridae* family<sup>1, 2</sup>. These viruses are morphologically similar, but genetically, antigenically and pathogenically distinct. PCV1 is apathogenic and was first detected in porcine kidney cells (PK-15). PCV2 is highly pathogenic, infects nearly every organ of the host, and responsible for porcine circovirus-associated disease (PCVAD) or porcine circovirus disease (PCVD), which culminates in the immunosuppression of the host and death from secondary infections<sup>3-7</sup>. Naturally occurring PCV2 infections of animals in the wild include rodents, calves, foxes, minks, goats, racoon dogs, boars, and swine<sup>8-13</sup>. Controlled PCV2 infections under laboratory conditions include BALB/c mice, Kummung mice, and multiple human cell lines<sup>14-18</sup>. Infections in calves and rodents results in immunosuppression of the host<sup>8, 9, 12, 13, 19</sup>. PCV3 was recently shown to associate with multiple diseases in swine that include reproductive failure, dermatitis, and nephropathy syndrome<sup>20-22</sup>. Swine farms are billion-dollar industries in swine rearing countries, and PCV2 breakouts can be economically devastating<sup>16</sup>. The ability of PCV2 to overcome the species barrier, induce its pathogenic phenotype, and infect human cells in culture is suggestive of a pathogen that is on the precipitous of becoming zoonotic and potentially a threat to human health. Additionally, PCV2 is a deterrent for xenotransplantation, as swine are considered a viable source of organs for humans<sup>23-25</sup>. While a vaccination program for PCV2 has been established, PCV2's high mutation ( $1.2 \times 10^{-6}$  substitutions site<sup>-1</sup> year<sup>-1</sup>) and recombination rates are responsible for recent genotypes demonstrating resistance to vaccination<sup>26, 27</sup>. Consequently, alternative therapeutics that disrupt one or multiple stages of PCV2 life-cycle are needed. A promising strategy is to disrupt the self-assembly of viruses within the cell; however, this requires knowledge of the critical factors needed for the successful assembly of infectious virion. Studies describing PCV2 capsid assembly are limited<sup>19, 28, 29</sup>. Recent work demonstrated that capsid assembly occurs within the nucleus of cells expressing the capsid protein (CP)<sup>30</sup>. Knowledge of the PCV2 assembly mechanism will increase our understanding of virus capsid assembly and provide an avenue for developing therapeutics to inhibit PCV2 self-assembly.

The PCV2 CP (27.8 kDa) possesses a forty-two amino acid N-terminus where nearly half of the amino acids are arginines. Hereafter, we refer to this N-terminus as the arginine rich motif (ARM). The ARMs possess multiple nuclear localization signals (NLS) that are responsible for trafficking the virus to the nucleus during the entry process and/or genome packaging<sup>31</sup>. Other functions associated with the ARM include disruption of

liposomal membrane for cellular entry, and interaction with the negative charged phosphates of the genome <sup>32</sup>. Expression of the CP in *Trichoplusia ni* (insect) or HEK293 (human) cells results in the formation of nucleic acid containing capsids within the cells <sup>30, 33, 34</sup>. The capsid structures from these expression systems are nearly indistinguishable and demonstrate the CP to adopt the canonical viral jelly roll fold prevalent in many viral capsid proteins. However, the ARMs cannot be modeled in any of these studies <sup>30, 33, 35-37</sup>. This is most likely because the ARMs either do not adopt the icosahedral symmetry imposed during structure determination, or that the ARMs are unstructured.

PCV2 capsids assemble within the cell to enclose nucleic acid; however, PCV2 variants designed to eliminate this interaction are unable to assemble into capsids within the cell <sup>30</sup>. Others have shown that capsids can assemble *in vitro* to enclose nucleic acid -suggesting that the interaction between CP and nucleic acid is a requisite for capsid assembly <sup>38</sup>. We have been able to establish an *in vitro* protocol for assembling PCV2 capsids in the absence of nucleic acid. This allows us the unique opportunity to study capsid assembly in the absence of nucleic acid. By combining our biochemical assays with *in silico* studies, we demonstrate that ARM is not necessary for *in vitro* capsid assembly, and report that a protonation event of the CP is responsible for regulating capsid assembly. In particular, the CPs assemble into capsids at pH 5, and not at pH 3 or 7. This suggests that pH 5 somehow mimics cellular conditions necessary for capsid self-assembly, and the CP at this pH is likely to resemble the assembly competent structure prior to cellular capsid assembly. Concurrently, the *in silico* approach demonstrates the structural changes at pH 3 and 7 responsible for inhibiting assembly. We emphasize that it is not trivial to determine the atomic structure of monomeric CP at multiple pH values using experimental approaches. We have shown that full-atom molecular dynamics (MD) simulation of entire PCV2 capsid faithfully represents its structure and dynamics and, thus, can be used as a reliable source of information complementing and explaining experimental data <sup>39-42</sup>. Here, we use MD to study the conformational landscape of the CP at different pH values. The use of conventional MD for studying the effect of pH on structure is problematic as this approach maintains a single protonation state for each residue throughout simulation. Thus, the dependence between changes in protonation states and changes in the structure are lost, and potentially inaccurate conformations and dynamics are sampled. In the computational part of this work, we interpret the biochemical results to determine what happens to the monomeric CP structure at these pH values by performing recently developed Constant pH Molecular Dynamics simulation (CpHMD) with Replica Exchange (RE) techniques in pH-dimension (CpH REMD) <sup>43</sup>. Unlike the classical MD approach, CpHMD simulation allows for the change of

amino acid protonation states during the simulation. The RE technique provides enhanced sampling. Capsid assembly originates from monomeric CPs assembling into larger oligomers; thus, atomistic knowledge of the CP monomer is a requisite for understanding capsid assembly. We study the pH-dependent structural features of CP using CpH REMD.

We demonstrate that CpH REMD provides interpretation for the observed phenomena of pH driven *in vitro* capsid assembly. The CP conformation suited for capsid assembly is reliant on the position of the GH-loop, located at the 3-fold icosahedral axes of symmetry (*iaos*), and this position is regulated by a protonation/deprotonation event involving two pairs of acidic amino acids that we refer to as the dyads. Our results can also explain regulation of PCV2 self-assembly within the cell. We recently demonstrated that the GH-loop interacts with nucleic acid at the 3-fold *iaos* of nucleic acid containing VLP<sup>30</sup>. We hypothesize that cytoplasmic capsid assembly is prohibited by electrostatic repulsion of these dyads, and that interaction of the GH-loop with nucleic acid inside the nucleus at the 3-fold *iaos* may be the driving factor for overcoming this repulsion. To the best of our knowledge, this is first description of how PCV2 regulates its capsid assembly.

## RESULTS

### *In vitro* capsid assembly is driven by pH

The PCV2 CP (GenBank: AAO39759.1) can be expressed and purified from *E. coli* to homogeneity as monomeric subunits<sup>32,33</sup>. These samples are devoid of nucleic acid as determined by a 260 nm / 280 nm ratio. Incubation of CP in low pH conditions for several weeks produces capsids; however, deletion or modification of the ARM results in robust and rapid capsid assembly<sup>32</sup>. This may be because nucleic acid is needed to neutralize the positive charge associated with the ARM. Therefore, we chose to substitute and exclude the ARM from our assembly studies. Here we discuss two variants: 1) CP\*, which replaces the ARM with a hexa-histidine affinity tag followed by a twenty-six amino acid linker (His<sup>6</sup>-linker); and 2) CP<sup>ΔARM</sup>, which excludes the first forty-two amino acids (Fig. 1). The capsid assembly buffer was determined by optimizing the buffer composition used to produce crystals of PCV2 capsids for X-ray crystallography<sup>33</sup>. Reducing the pH of the assembly buffer from 8.0 to 3.0 produces the greatest number of capsids at pH 5 (Fig. 1b). Assembled capsids are not observed at pH values greater than 6.0 despite increasing the CP or PEG concentration. This suggests that protonation of one or more titratable residues (Asp, Glu or His) may be responsible for regulation of capsid assembly. To further

explore this possibility, we used molecular dynamics to study the conformations of CP\* at different pH values. We note that the ability of CP<sup>ΔARM</sup> to assemble with a similar pH profile as the CP\* suggests that the N-terminus sequence in CP\* does not affect capsid assembly (Fig. 1).

### ***Conformational sampling of CP\* with CpH-REMD simulation***

To understand the relationship between capsid assembly and pH, we investigated the representative CP\* conformations at pH 3, 5, and 7 using CpHMD simulations in combination with pH-REMD to enhance conformational sampling. We set 10 replicas, 1  $\mu$ s each, spanning the pH range from 3.0 to 7.5 at 0.5 pH-unit increments. There are 32 amino acids (Asp, Glu, His) in the CP sequence that are titratable in this pH range. These residues were the target of titration during the MD simulation.

Coordinates for the N-terminus are not present in any of the experimental structures of PCV2<sup>30, 32-34, 38</sup>; therefore, we modeled it using MODELLER<sup>44</sup>. The convergence of simulation was monitored by observing the evolution of free energy surfaces (FES) as a function of variables Rg vs. SASA and Rg vs. RMSD (Supplementary Fig. 1-11). The comparison of the Rg vs. SASA FES suggests the CP to adopt a more compact shape at pH 7 having smaller Rg and SASA. RMSD was calculated from the alignment of each conformation to the capsid CP structure (PDB entry 3R0R). Initial alignments produced large ( $> 20$  Å) RMSD and Rg values ( $< 29$  Å). We hypothesized that a flexible N-terminus may be responsible for the large Rg, SASA and RMSD values. Exclusion of the His<sup>6</sup>-linker reduced the RMSD ( $< 7.5$  Å) and Rg ( $< 22$  Å) values significantly (Supplementary Fig. 4 – 6). Analysis of FES indicated that our CpH-REMD simulation converged within 1  $\mu$ s. We note that because our biochemical experiments demonstrate CP\* and CP<sup>ΔARM</sup> to exhibit comparable assembly profiles, we conclude that the pH-dependent conformational changes regulating self-assembly must involve the segment CP\*<sup>42-231</sup>. We focus on the section without the His<sup>6</sup>-linker (CP\*<sup>42-231</sup>).

### ***Free energy surfaces at different pH values demonstrate different CP conformations***

Figure 2 shows the conformational landscapes for each pH by the free energy surfaces in Rg and RMSD values. Highly populated regions (blue) corresponding to minimum energy of CP conformations were identified during the 1  $\mu$ s simulation. The distribution of these regions is unique for each pH value. A minimum is observed at Rg value of 19.7 Å and RMSD value of 4.6 Å for pH 3. A minimum for pH 5 is observed at Rg 20.0 Å and

RMSD 4.0 Å. Two minima are observed at pH 7 with Rg 20.0 Å and RMSD of 4.4 Å (min 1), and Rg 19.5 Å RMSD 5.0 Å (min 2).

The existence of two minima suggests the presence of at least two predominant conformations at pH 7. Remarkably, the RMSD value of the minimum at pH 5 is lower than at pH 3 or 7, meaning that the conformation of CP\*<sup>42-231</sup> at pH 5 may be more similar to the conformation observed in the assembled capsid.

### ***pH changes the conformation of CD and GH-loops***

Representative CP conformations for each pH were identified using structural clustering of the extracted structures from the minima at each pH. To determine the effect of pH on the CP structure, we compared the representative conformation of each cluster (centroid) to the capsid CP structure (Fig. 3a)<sup>33, 37</sup>. Four regions deviate significantly from the capsid CP structure: the N-terminus (His<sup>6</sup>-linker), the CD-loop (residues 80-92), the GH-loop (residues 166-194), and the C-terminus (residues 225-231) (Figs. 3a, b and 4). There are thirteen titratable residues in the N-terminus (7 His and 6 Asp). Our pKa calculations demonstrated that the N-terminus His and Asp residues are fully and partially protonated at pH 3, respectively (Table 1). Increasing of pH from 3 results in partial deprotonation of these His residues and complete deprotonation of these Asp residues at pH 5, and complete deprotonation of both at pH 7. In the assembled capsid, the N-terminus helps define the CP:CP interface at the 2-fold *iaos* (Fig. 3a, 4a). At pH 3 and 5, the N-terminus is positioned near the 2-fold *iaos* (Figs. 4b, c). On the other hand, at pH 7 it is positioned near the face that defines the inner surface of the capsid, due to the ionic interactions between the negatively charged Asp in the N-terminus and the positively charged Arg and Lys at the capsid inner surface (Fig. 4d, e).

In the assembled capsid, the CD-loop contacts an equivalent loop in the neighboring subunit (Fig. 3a). Our MD results demonstrate the CD-loop to adopt distinct conformations at each pH (Fig. 4); however, the lack of titratable residues in the loop suggests that the observed variability may be intrinsic to the loop. Biochemical studies have demonstrated that substitution and truncation of the CD-loop do not abrogate capsid assembly, demonstrating that CD-loop is likely not a deciding factor in capsid assembly<sup>28</sup>. The GH-loop positions are more interesting: the loop adopts the same position as in assembled capsid at pH 5, however, the loop is closer to the EF-loop at pH 3, or twists and moves closer to the CD-loop at pH 7 (Figs. 4f-j). From the analysis of the experimental capsid structure, the GH-loop contributes a third (1000 Å<sup>2</sup>) of the CP surface that is buried upon capsid assembly<sup>45</sup>. We explore the significance of the GH-loop further below. In the assembled capsid, the C-terminus also helps define the CP:CP interface at the 2-fold *iaos*. Our MD results demonstrate the C-terminus

to adopt distinct positions at each pH value (Fig. 4a-e, and supplementary Figs. 16, 18-21). There is a single titratable residue (Asp228) in the C-terminus, and pKa calculations suggest that it becomes deprotonated at pH values greater than 5 (see below). Zhan et al. recently published that only the main chain atoms of amino acids 225-231 are necessary for capsid assembly, as deletion but not substitution of this region abolished capsid assembly<sup>29</sup>. It is likely that the C-terminus adopts the necessary position and conformation when two subunits interact at the 2-fold *iaos*.

### ***Two aspartic dyads located in the EF- and GH-loops regulate capsid assembly***

In the assembled capsid, two acidic residues in the EF-loop (Asp126 and Asp127) and two in the GH-loop (Asp168 and Asp172) are located in close proximity (Fig. 4f-j). Their protonation state may be responsible for the observed conformational variability of the GH-loop, so they were titrated in the simulation. To quantify the position of the GH-loop relative to the EF-loop at the three pH values, we calculated distances between the amino acids of these two loops for the extracted conformations (Fig. 2). Figure 5 represents the distribution of distances between the centers of mass for residues 168 and 126 (hereafter denoted as 168:126) (Figs. 5a, c, e, g) and residues 172:126 (Figs. 5b, d, f, h).

We compared the observed distance distribution at each pH with the distances in the experimentally determined capsid structure. At pH 3, the monomodal distribution between residues 168:126 (10-11.2 Å) was similar to the distance observed in the capsid structure (10.7 Å); however, distances between residues 172:126 were distributed monomodally (8.1-8.7 Å) and were considerably greater than that of the capsid structure (5.1 Å) (Figs. 5a, b). At pH 5, the multimodal distribution between residues 168:126 (8.5-10.5 Å and 10.9-14.5 Å) and 172:126 (5.1-6.7 Å and 8.1-9.3 Å) was similar to the distances observed in the capsid structure (Figs. 5c, d). At pH 7 minima 1, the multimodal distribution between residues 168:126 (10.3-13.1 Å and 16.0-20.0 Å) was similar to the distance observed in the capsid structure; however, the multimodal distribution between residues 172:126 (8.2-14.4 Å and 18.7-21.3 Å) is significantly greater than that observed in the capsid structure (Figs. 5e, f). These values reflect the observed twisting of the GH-loop in Fig.5d, where Asp168 adopts the same position as in the assembled capsid, while Asp172 is pushed away from the EF-loop. At pH 7 minima 2, the bimodal distributions between amino acids 168:126 and 172:126 are considerably greater than that of the capsid structure (Figs. 5g, h). The increased distances between these dyads at pH 7 versus 5 demonstrates that the GH-loop moves away from the EF-loop at neutral pH.

The most important factor responsible for this movement could be repulsive forces between the two deprotonated dyads. To address this possibility, we calculated the pKa values of titratable amino acids from the conformations extracted from each minimum (Fig. 2, Table 1). Based on the calculated pKa values shown in Table 1, Asp126 and Asp127 (EF-loop) are predominantly protonated at pH 3, partially deprotonated at pH 5, and predominantly deprotonated at pH 7 for both minima. Asp168 (GH-loop) is predominantly protonated at pH 3 and deprotonated at pH 5 and pH 7 (for both minima). Asp172 is protonated at pH 3, protonated at pH 5, and deprotonated at pH 7 (for both minima). Protonation of the four aspartic acids (two dyads) at pH 3 allows the EF- and GH-loops to be kept in close proximity (Fig. 4g). At pH 5, Asp126, Asp127 and Asp168 are deprotonated while Asp172 is protonated. The deprotonation event of these three aspartic acids results in electrostatic repulsion pushing the GH-loop away from the EF-loop (Fig. 4h) to the position observed in the capsid structure (Fig. 4f). At pH 7, all four aspartic acids are deprotonated, and additional electrostatic repulsion pushes GH-loop further from EF-loop (Figs. 4i, j). For nearly all representative conformations in two minima at pH 7 (Fig. 2c), we observed twisting of the GH-loop (Figs. 4i, j).

## DISCUSSION

The ability of virion to regulate the assembly of their CP to generate infectious virion may provide them greater evolutionary fitness over virion whose CP assemble uncontrollably into aggregates, empty, or cellular nucleic acid filled capsids within the infected cell. Some of these regulatory mechanisms may be inherent to the viral CP or capsid. Understanding the mechanisms responsible for this regulation are requisite to understanding viral life cycles, and the development of therapeutics to disturb formation of infectious virion. It has been shown that expression of PCV2 CP in mammalian cells results in capsid assembly solely within the nucleus<sup>30</sup>. This suggests that PCV2 is able to regulate capsid formation by inhibiting assembly within the cytoplasm, and triggering it within the nucleus where PCV2 genome replication occurs. In this work we begin to study how PCV2 performs this regulation by using biochemical and computational approaches.

Within the cell, CP assembles near neutral pH to enclose nucleic acid. In vitro, CP assembles near neutral pH only in the presence of nucleic acid; thus, the interaction between CP and nucleic acid is a requisite for capsid assembly<sup>38</sup>. The CP:nucleic acid is likely driven by the high Arg content of the ARM. To diminish the CP:nucleic interaction, and to focus on CP:CP interaction during capsid assembly, we chose to substitute and

exclude the ARM from our assembly studies. Here we demonstrate that PCV2 CP devoid of nucleic acid assemble *in vitro* only under acidic conditions (pH 5, Fig. 1). Earlier structural studies demonstrated these capsids to be indistinguishable from capsids assembled within insect and mammalian cells<sup>30, 32, 33, 37</sup>. Consequently, in the absence of nucleic acid, pH 5 can substitute as the driving factor for CP assembly. To understand how pH 5 influences capsid assembly, we simulated CP at different pH values using Constant pH MD simulation with Replica Exchange technique in pH-dimension. The CpH-REMD simulations demonstrated that the GH-loop adopts the position observed in the capsid only at pH 5 (Fig.4f, h). Calculation of pKa values using the CP conformations generated by CpH-REMD demonstrated that the GH-loop adopts the proper position at pH 5 because Asp126/Asp127 (EF-loop) and Asp168 (GH-loop) are deprotonated while Asp172 (GH-loop) is protonated (Fig.4h). However, these aspartic acids are deprotonated at pH 7, and electrostatic repulsion between them pushes the GH-loop away from the EF-loop to suppresses capsid assembly (Fig. 4i, j). Alignment of all GenBank PCV2 CP sequences demonstrated that at least three acidic amino acids (Asp/Glu) are present in the four positions, and that Asp172 is absolutely conserved<sup>30</sup>. This is in support of our hypothesis that PCV2 utilizes Asp172 for regulating capsid assembly. In the context of infection, the electrostatic repulsion described by the CpH-REMD study remains to be addressed because the pH of the cytoplasm and nucleus are anticipated to be close to neutral.

An electrostatic potential map of PCV2 capsid demonstrated its inner surface to be positively charged<sup>40</sup>. Amino acids responsible for these charges are likely to interact with the negative charged phosphates of nucleic acid packaged within the capsid<sup>41</sup>. Indeed, it has been shown that amino acid substitutions that interfere with CP-nucleic acid phosphate interaction abolish capsid assembly<sup>30</sup>. Given that Khayat et al. have demonstrated the GH-loop (Lys179) to interact with the backbone phosphates of nucleic acid and the importance of this loop for capsid assembly (as demonstrated by these studies), we propose that the interaction between the GH-loop and nucleic acid may be responsible for properly positioning the GH-loop for capsid assembly<sup>30</sup>. While others have demonstrated capsid assembly of *E. coli* expressed CP at pH 7, it remains to be determined if their material contained nucleic acid that may have facilitated the assembly process<sup>29, 46</sup>. Our results also demonstrate that the presence of CP ARM is not necessary for *in vitro* capsid assembly at pH 5.

Reports of PCV2 vaccine failure outbreaks and genotype shifts suggest that the development of antiviral therapeutics is necessary<sup>19, 33, 47</sup>. Our results show that specific positioning of the GH-loop is important for PCV2 capsid assembly, and distraction from this position may inhibit assembly. Therefore, compounds capable of

controlling the position of the GH-loop may serve as potential candidates for inhibiting PCV2 capsid assembly. We propose that two families of compounds may achieve this task. One family would position the GH-loop such that assembly is not possible. The alternative family would promote the CP conformation that is capable of assembling into capsids within the cytoplasm, where the PCV2 genome is absent. Such assemblies are noninfectious particles. These compounds may be effective inhibitors by promoting capsid assembly in the absence of nucleic acid. Thus, our observed CP conformations have physiological importance.

In conclusion, we believe that our finding is of fundamental and practical importance, as it provides an understanding of how a virus regulates its self-assembly and provides new opportunity for development of antiviral therapeutics. To the best of our knowledge, this is first explanation for the PCV2 assembly mechanism.

## **MATERIALS AND METHODS**

### **Purification of PCV2 capsid protein**

The preparation of the CP\* protein (Fig. 1) is as previously described<sup>33</sup>. The CP lacking the ARM (CP<sup>ΔARM</sup>) expression construct was generated by genetically fusing a His6 small-ubiquitin related modifier (SUMO) at the N-terminus of the CP (amino acids 42-233) (pET28a plasmid, Novagen) using the services of GenScript. Both constructs were expressed in the *Escherichia coli* Rosetta DE3 strain (Novagen). Cells were grown in Terrific Broth (TB) media at 37 °C until mid-log phase in the presence of appropriate antibiotic. Protein expression was induced with 0.25 mM Isopropyl β-D- 1-thiogalactopyranoside (IPTG) for 16 h at 20 °C. The cells were harvested by centrifugation (30 min, 4,000 rpm, 4 °C). The pellet was lysed in 50 mM N- cyclohexyl-3-aminopropanesulfonic acid (CAPS) pH 10.5, 50 mM sodium chloride (NaCl), 0.2 mM phenylmethanesulfonyl fluoride (PMSF), 0.5 mM tris (2-carboxyethyl) phosphine (TCEP) and 25 units of benzonase (MilliporeSigma). The lysate was centrifuged (30 min, 16,000 rpm, 4 °C) and the supernatant was subjected to Ni-NTA gravity flow chromatography. The protein was eluted with the lysis buffer supplemented with 1 M imidazole pH 10.5. The elution was treated with 25 units of benzonase while dialyzing in lysis buffer overnight. The protein was further purified using tandem cation-anion ion-exchange chromatography (HiTrap SP HP and HiTrap Q FF column, GE healthcare). Protein was bound to the HiTrapQ FF column and eluted around 200 mM NaCl. 250 mM L-Arg was added to the eluent and the protein was concentrated to 20 mg ml<sup>-1</sup>. Protein concentration was determined using the method of Gill and von Hippel<sup>48</sup>, aliquoted into 50 μl, flash frozen in N<sub>2</sub>(l), and stored at -80 °C.

## **SUMO-specific protease expression and purification**

The laboratory of Dr. Patrick J. Loll provided the expression construct for the SUMO-specific protease (dtuD1). The protein was expressed and purified following the publications protocol <sup>49</sup>.

## **In vitro capsid assembly**

The CP<sup>ΔARM</sup> protein was incubated with dtuD1 to release the free CP<sup>ΔARM</sup>. The composition of the digested material was confirmed using SDS-PAGE analysis. To determine the capsid assembly condition, 3 mg ml<sup>-1</sup> capsid protein was mixed with different assembly buffers in a 1:1 ratio and incubated at 4 °C for 48 hours. The first group of assembly buffer was 12% polyethylene glycol (PEG) 3350, 5% iso-propanol, 0.6 M Ammonia Citrate (pH 3.0, pH 4.0, pH 4.5, pH 5.0, pH 6.0, pH 7.0 and pH 8.0). The second group of assembly buffer was 0.6 M Ammonia Citrate (pH 5.0), 5% iso-propanol, 3%, 6%, 9%, 12% and 20% PEG 3350.

Negative stained grids were prepared by applying 5 μl of the sample at 0.1 mg ml<sup>-1</sup> to a 400-mesh carbon coated grid for 3 seconds. The specimen was wicked away and stained with 2% uranyl formate (UF). Three equivalent grids were prepared and imaged for each condition. A JEOL 1230 electron microscope at the New York Structural Biology Center (NYSBC) was used for imaging the specimen at a magnification of 60,000x. Imaging was performed using the Legimon package <sup>50</sup>.

## **Molecular dynamics simulation**

All simulations were performed using AMBER 18 package <sup>51</sup>. The starting structure for simulations was derived from the crystal structure of a CP\* subunit of the PCV2 capsid (PDB entry 3R0R) <sup>33</sup>. The unresolved first 32 amino acids of the structure were reconstructed with topology modelling in MODELLER <sup>44</sup> and added to the CP structure. The initial solvated system of the simulation was prepared with the leap module of AmberTools. The protein was solvated with TIP3P water model in a cubic box of the size providing 10 Å between protein and edges. To neutralize the system and mimic 0.15 M ionic strength, 128 Cl<sup>-</sup> and 115 Na<sup>+</sup> ions were added. The total number of atoms in the system was 129762. The force field parameters of AMBER FF10 <sup>51</sup> was assigned for the protein. The SHAKE algorithm was applied to bonds involving hydrogens bonds <sup>52</sup>. The long-range Coulomb interactions were treated with the particle mesh Ewald (PME) method <sup>53</sup> and the direct space cutoff distance was set to 8 Å. The van der Waals interactions were calculated using 8 Å cutoff.

Energy minimization of the solvated system was performed with steepest descent followed by conjugate gradient algorithm for 5000 steps. The energy-minimized system was then heated for 400 ps from 10 K to 300 K using Langevin dynamics (collision frequency of 5 ps<sup>-1</sup>). Then, 4 ns simulation of the system was performed in NPT ensemble. The temperature and pressure were kept constant at 300 K and 1 atm by Langevin thermostat and Berendsen barostat respectively. Furthermore, an additional 4 ns simulation was performed in NVT ensemble with the Langevin thermostat maintaining temperature 300 K.

For constant pH MD simulation, we used discrete protonation state model proposed by Mongan and developed by Swails for MD propagated in explicit solvent<sup>43</sup>. After each 200 fs of MD simulation performed in explicit solvent, the protonation change event was attempted for all titratable residues in GB implicit solvent<sup>54</sup>, solvent relaxation time was set to 200 fs. In CpHMD simulation, all Asp, Glu and His residues (a total of 32 ionizable residues) were titrated. We performed 10 replicas in pH-dimension spanning the pH range 3.0 – 7.5 at 0.5 pH-unit intervals, each replica was simulated for 1  $\mu$ s in NVT ensemble, resulting in a total of 10  $\mu$ s of simulation. Attempts to exchange between replicas were set to every 200 fs of the simulation. The CpHMD simulations were performed using PMEMD with GPU acceleration<sup>55-59</sup>. Exchange plots for each replica presented in SI (Supplementary Fig 13).

### Data Analysis of the simulation

Trajectory reconstructions for data analysis were performed for pH 3, 5 and 7 using Cpptraj program in AMBERTools. Structural analysis at each pH was performed on structures extracted from free energy minima at each pH (Fig. 2). Hierarchical agglomerative approach was applied for clustering the extracted conformations, all-vs-all C $\alpha$  atoms, with 2 Å cutoff<sup>60</sup>. Processing of protonation states was performed with cphstats in AMBER<sup>55-59</sup> and pKa values were calculated using Henderson–Hasselbalch equation based on the ratio of time points with a protonated vs. deprotonated amino acid with inhouse written program<sup>54</sup>. Cpptraj in AMBERTools was used for distance calculations between the specified residues center of mass, solvent accessible surface area (SASA), radius of gyration (Rg), and C $\alpha$  root mean square deviation (RMSD)<sup>60</sup>. Buried surface area (BSA) values were calculated using PDBePISA server<sup>61</sup>. Contour plots for population maps, distances histograms, and ribbon diagrams were made in Origin 2018b (OriginLab), MOE (Chemical Computing Group, Inc) and UCSF Chimera<sup>62</sup>.

## **ACKNOWLEDGEMENTS**

The calculations were performed on the HOKUSAI Big Waterfall system at RIKEN; E.T. acknowledges a JSPS International Research Fellowship, grant number PE17023. We would like to thank Dr. Yoshinori Hirano, Dr. Gentaro Morimoto for their technical support, and Yumie Ohyama for administrative support. Funds responsible for supporting these studies were provided by NIH National Institute of General medical Sciences and National institute of Allergy and Infectious Diseases (5SC1AI114843) to R.K., by Grant Number 5G12MD007603-30 from the National Institute on Minority Health and Health Disparities to R.K. Electron microscopy was performed at the Simons Electron Microscopy Center and National Resource for Automated Molecular Microscopy located at the New York Structural Biology Center, supported by grants from the Simons Foundation (349247), NYSTAR, and the NIH National Institute of General Medical Sciences (GM103310).

## **CONFLICT OF INTEREST**

The authors declare that no conflict of interest exists.

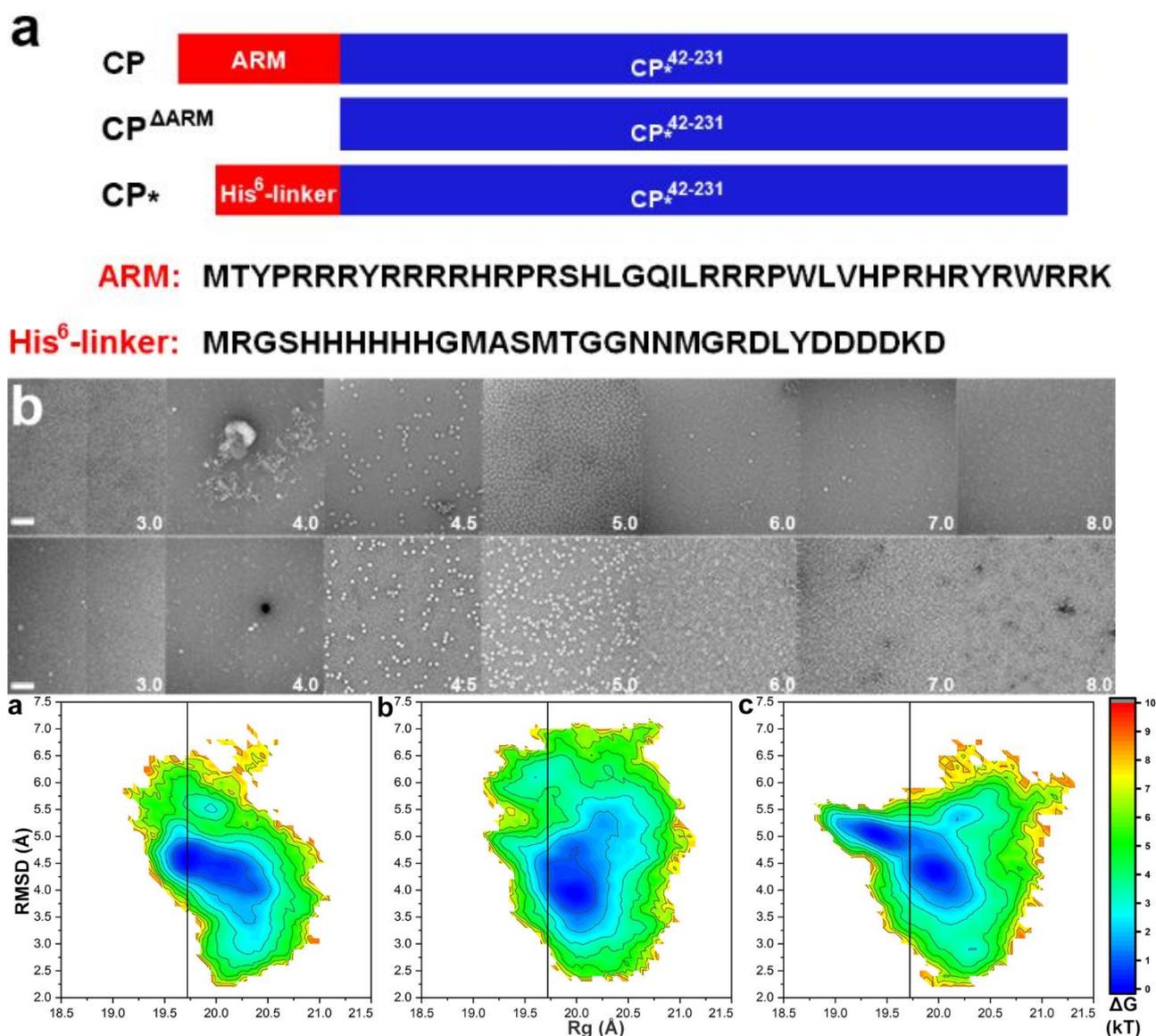
## Reference

1. D. Todd, *Avian Pathol*, 2000, **29**, 373-394.
2. K. Rosario, M. Breitbart, B. Harrach, J. Segales, E. Delwart, P. Biagini and A. Varsani, *Arch Virol*, 2017, **162**, 1447-1463.
3. A. Mankertz, J. Mankertz, K. Wolf and H. J. Buhk, *J Gen Virol*, 1998, **79 ( Pt 2)**, 381-384.
4. T. Opriessnig, X.-J. Meng and P. G. Halbur, *Journal of Veterinary Diagnostic Investigation*, 2007, **19**, 591-615.
5. S. Ramamoorthy and X. J. Meng, *Anim Health Res Rev*, 2009, **10**, 1-20.
6. J. Segalés, *Virus Res*, 2012, **164**, 10-19.
7. Z. B. Deng, A. W. Yuan, W. Luo, N. D. Wang, Q. L. Gong, X. L. Yu and L. Q. Xue, *Acta Vet Hung*, 2013, **61**, 234-243.
8. L. Li, T. Shan, O. B. Soji, M. M. Alam, T. H. Kunz, S. Z. Zaidi and E. Delwart, *J Gen Virol*, 2011, **92**, 768-772.
9. M. Y. Halami, M. Freick, A. A. Shehata, H. Muller and T. W. Vahlenkamp, *Vet Microbiol*, 2014, **173**, 125-131.
10. Z. Wang, Y. Ma, H. Khalil, R. Wang, T. Lu, W. Zhao, Y. Zhang, J. Chen and T. Chen, *Int J Nanomedicine*, 2016, **11**, 4025-4036.
11. X. Wang, W. Li, X. Xu, W. Wang, K. He and H. Fan, *Gene*, 2018, **651**, 57-61.
12. T. Song, J. Hao, R. Zhang, M. Tang, W. Li, W. Hui, Q. Fu, C. Wang, S. Xin, S. Zhang, P. Rui, H. Ren and Z. Ma, *BMC Vet Res*, 2019, **15**, 107.
13. T. Song, S. Zhang, J. Hao, S. Xin, W. Hui, M. Tang, W. Li, R. Tian, X. Liu, P. Rui, H. Ren, C. Wang, Q. Fu and Z. Ma, *Transbound Emerg Dis*, 2019, **66**, 1-6.
14. M. Kiupel, G. W. Stevenson, J. Choi, K. S. Latimer, C. L. Kanitz and S. K. Mittal, *Vet Pathol*, 2001, **38**, 74-82.
15. K. Hattermann, C. Roedner, C. Schmitt, T. Finsterbusch, T. Steinfeldt and A. Mankertz, *Xenotransplantation*, 2004, **11**, 284-294.
16. A. L. Pinheiro, L. H. Bulos, T. S. Onofre, M. de Paula Gabardo, O. V. de Carvalho, M. C. Fausto, R. M. Guedes, M. R. de Almeida and A. Silva Junior, *Res Vet Sci*, 2013, **94**, 764-768.
17. X. Liu, T. Ouyang, T. Ma, H. Ouyang, D. Pang and L. Ren, *BMC Veterinary Research*, 2018, **14**, 137.
18. X. Liu, T. Ouyang, H. Ouyang, X. Liu, G. Niu, W. Huo, W. Yin, D. Pang and L. Ren, *Sci Rep*, 2019, **9**, 5638.
19. N. Wang, Y. Zhan, A. Wang, L. Zhang, R. Khayat and Y. Yang, *J Gen Virol*, 2016, **97**, 3331-3344.
20. I. Tischer, R. Rasch and G. Tochtermann, *Zentralbl Bakteriol Orig A*, 1974, **226**, 153-167.
21. I. Tischer, H. Gelderblom, W. Vettermann and M. A. Koch, *Nature*, 1982, **295**, 64-66.
22. R. Palinski, P. Piñeyro, P. Shang, F. Yuan, R. Guo, Y. Fang, E. Byers and B. M. Hause, *Journal of Virology*, 2017, **91**, e01879-01816.
23. B. Ekser, P. Rigotti, B. Gridelli and D. K. Cooper, *Transpl Immunol*, 2009, **21**, 87-92.
24. R. N. Pierson, 3rd, A. Dorling, D. Ayares, M. A. Rees, J. D. Seebach, J. A. Fishman, B. J. Hering and D. K. Cooper, *Xenotransplantation*, 2009, **16**, 263-280.
25. J. Denner and A. Mankertz, *Viruses*, 2017, **9**.
26. S. Duffy, L. A. Shackelton and E. C. Holmes, *Nat Rev Genet*, 2008, **9**, 267-276.
27. C. T. Xiao, P. G. Halbur and T. Opriessnig, *J Gen Virol*, 2015, **96**, 1830-1841.
28. D. Wang, S. Zhang, Y. Zou, W. Yu, Y. Jiang, Y. Zhan, N. Wang, Y. Dong and Y. Yang, *Front Cell Infect Microbiol*, 2018, **8**, 232.
29. Y. Zhan, W. Yu, X. Cai, X. Lei, H. Lei, A. Wang, Y. Sun, N. Wang, Z. Deng and Y. Yang, *J Virol*, 2020, DOI: 10.1128/JVI.00042-20.
30. R. Khayat, K. Wen, A. Alimova, B. Gavrilov, A. Katz, J. M. Galarza and P. Gottlieb, *Virology*, 2019, **537**, 186-197.
31. J. Shuai, W. Wei, L. Jiang, X. I. Li, N. Chen and W. Fang, *Acta Biochimica et Biophysica Sinica*, 2008, **40**, 71-77.
32. S. Dhindwal, S. Feng and R. Khayat, *J Mol Biol*, 2019, **431**, 3261-3274.
33. R. Khayat, N. Brunn, J. A. Speir, J. M. Hardham, R. G. Ankenbauer, A. Schneemann and J. E. Johnson, *J Virol*, 2011, **85**, 7856-7862.
34. Z. Liu, F. Guo, F. Wang, T. C. Li and W. Jiang, *Structure*, 2016, **24**, 319-328.
35. S. C. Harrison, A. J. Olson, C. E. Schutt, F. K. Winkler and G. Bricgone, *Nature*, 1978, **279**, 368-373.
36. R. A. Crowther, J. A. Berriman, W. L. Curran, G. M. Allan and D. Todd, *J Virol*, 2003, **77**, 13036-13041.
37. S. Dhindwal, B. Avila, S. Feng and R. Khayat, *J Virol*, 2019, **93**.
38. X. Mo, X. Li, B. Yin, J. Deng, K. Tian and A. Yuan, *PLoS Pathog*, 2019, **15**, e1007562.
39. E. Tarasova, V. Farafonov, R. Khayat, N. Okimoto, T. S. Komatsu, M. Taiji and D. Nerukh, *J Phys Chem Lett*, 2017, **8**, 779-784.

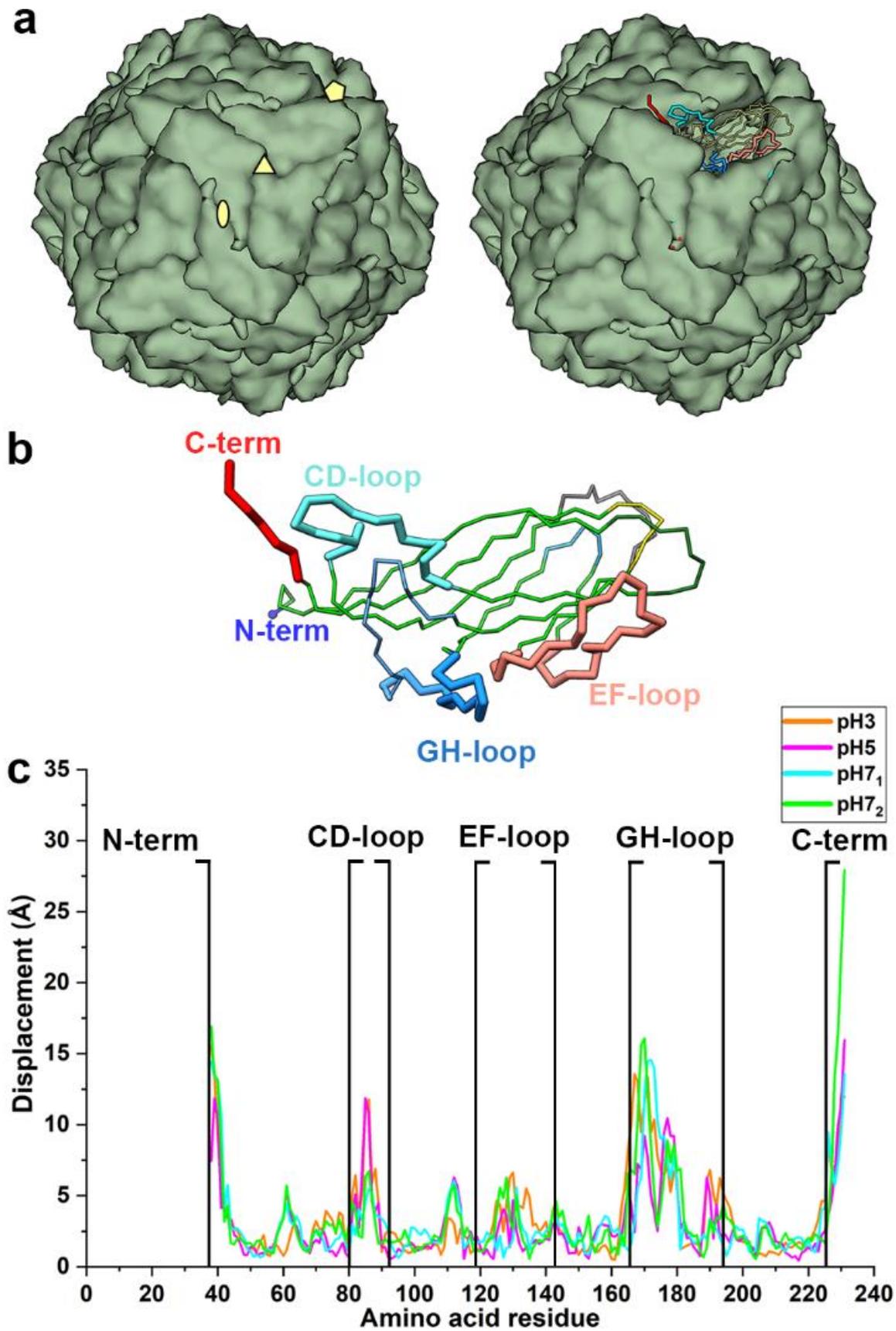
40. E. Tarasova, I. Korotkin, V. Farafonov, S. Karabasov and D. Nerukh, *Journal of Molecular Liquids*, 2017, **245**, 109-114.
41. E. Tarasova, V. Farafonov, M. Taiji and D. Nerukh, *Journal of Molecular Liquids*, 2018, **265**, 585-591.
42. E. Tarasova and D. Nerukh, *The Journal of Physical Chemistry Letters*, 2018, **9**, 5805-5809.
43. J. M. Swails, D. M. York and A. E. Roitberg, *J Chem Theory Comput*, 2014, **10**, 1341-1352.
44. A. Sali and T. L. Blundell, *J Mol Biol*, 1993, **234**, 779-815.
45. E. Krissinel and K. Henrick, *J Mol Biol*, 2007, **372**, 774-797.
46. P.-C. Wu, W.-L. Lin, C.-M. Wu, J.-N. Chi, M.-S. Chien and C. Huang, *Applied Microbiology and Biotechnology*, 2012, **95**, 1501-1507.
47. G. Franzo, M. Cortey, J. Segales, J. Hughes and M. Drigo, *Mol Phylogenet Evol*, 2016, **100**, 269-280.
48. S. C. Gill and P. H. von Hippel, *Anal Biochem*, 1989, **182**, 319-326.
49. S. D. Weeks, M. Drinker and P. J. Loll, *Protein Expr Purif*, 2007, **53**, 40-50.
50. B. Carragher, N. Kisseberth, D. Kriegman, R. A. Milligan, C. S. Potter, J. Pulokas and A. Reilein, *Journal of Structural Biology*, 2000, **132**, 33-45.
51. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712-725.
52. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *Journal of Computational Physics*, 1977, **23**, 327-341.
53. T. Darden, D. York and L. Pedersen, *The Journal of Chemical Physics*, 1993, **98**, 10089-10092.
54. J. Mongan, D. A. Case and J. A. McCammon, *Journal of Computational Chemistry*, 2004, **25**, 2038-2048.
55. A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *Journal of Chemical Theory and Computation*, 2012, **8**, 1542-1555.
56. S. Le Grand, A. W. Götz and R. C. Walker, *Computer Physics Communications*, 2013, **184**, 374-380.
57. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J Chem Theory Comput*, 2013, **9**, 3878-3888.
58. D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, I. Cheatham, T.E. , V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. G. Smith, R. SalomonFerrer, J. M. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *Journal*, 2018.
59. D. J. Mermelstein, J. A. McCammon and R. C. Walker, *J Mol Recognit*, 2019, **32**, e2765.
60. D. R. Roe and T. E. Cheatham, *Journal of Chemical Theory and Computation*, 2013, **9**, 3084-3095.
61. E. Krissinel and K. Henrick, *Acta Crystallographica Section D*, 2004, **60**, 2256-2268.
62. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, **25**, 1605-1612.

## FIGURE LEGENDS AND TABLES

**Fig. 1.** In vitro PCV2 capsid assembly. (a) Schematic representations of full-length CP, CP $\Delta$ ARM, and CP\* describing differences between proteins. The amino acid sequences of the N-termini are written below. (b) Transmission electron micrographs of negative stained in vitro assembled PCV2 capsid. Top, effect of pH (white text) in assembly of CP\* capsid. Bottom, effect of pH (white text) in assembly of CP $\Delta$ ARM capsid.

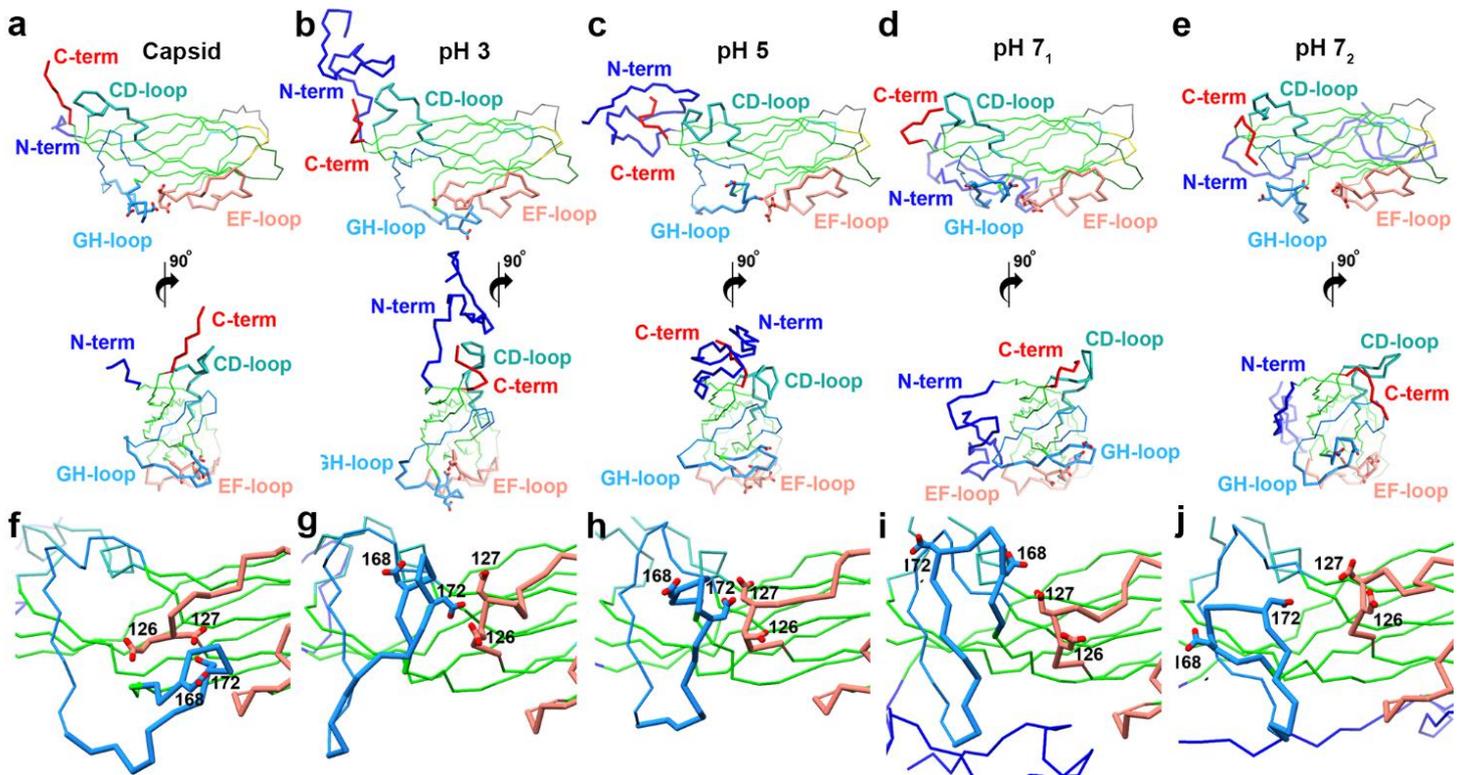


**Fig. 2.** Conformational landscape sampled by the capsid protein CP\*<sup>42-231</sup>. Free energy surface as function of parameters RMSD (CP\*<sup>42-231</sup> C $\alpha$  atoms) vs. Rg (CP\*<sup>42-231</sup> C $\alpha$  atoms) from the simulation at (a) pH 3, (b) pH 5, (c) pH 7. RMSD was calculated by comparing conformations generated from 1  $\mu$ sec of simulation to the reference CP in the assembled capsid (PDB entry 3R0R). Vertical black lines signify Rg value of PDB entry 3R0R.

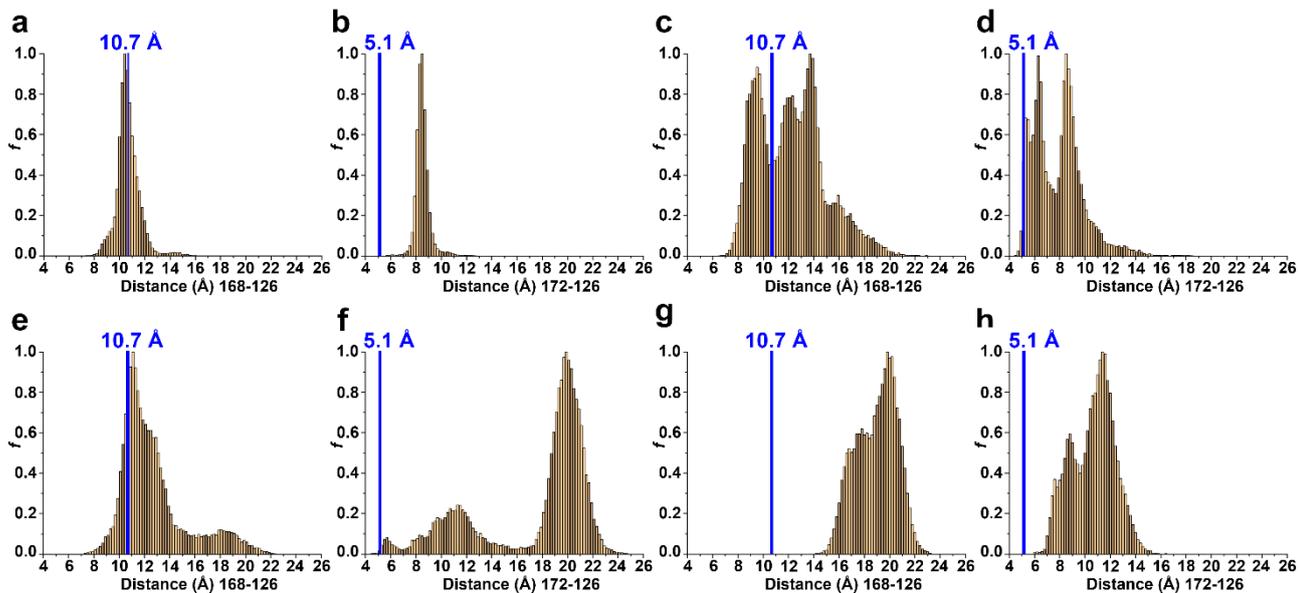


**Fig. 3.** Conformational variability of CP at pH 3, 5 and 7. (a) left: PCV2 capsid with CP shown as green blobs. The icosahedral axes of symmetry (*iaos*) are shown as pentagon (5-fold), triangle (3-fold), and ellipse (2-fold). Right: one of the blobs (subunits) is shown as a  $\text{Ca}$  tube representation with the important loops (N-term, CD, EF, GH).

and GH as thick tubes). (b) PCV2 CP C $\alpha$  tube representation with important loops (thick tubes) labelled. (c) distances between equivalent C $\alpha$  atoms of the representative conformations at pH 3, 5, 7 after alignment to capsid structure (PDB entry 3R0R). Panels a and b were made using the structure of the capsid. In the context of the assembled capsid, the N-terminus is in the interior of the capsid near the 2-fold *iaos*, CD-loop is located on the surface of the capsid near the 2-fold *iaos*, GH-loop is buried near the 3-fold *iaos*, and the C-terminus is located on the capsid surface near the 2-fold *iaos*.



**Fig. 4.** Representative conformations from the first cluster at each pH value. Regions with significant deviation from the capsid structure are highlighted according to Fig. 3b. a-e top versus bottom) C $\alpha$  trace of two orientations related by an anticlockwise 90° rotation about the y-axis. a) CP extracted from assembled capsids, b) representative of most populated cluster at pH 3, c) representative of most populated cluster at pH 5, d) representative of most populated cluster from first peak at pH 7, e) representative of most populated cluster from second peak at pH 7, f-j) close-up of EF- (pink) and GH-loops (blue) showing their acidic dyads f) CP extracted from assembled capsids, g) representative of most populated cluster at pH 3, h) representative of most populated cluster at pH 5, i) representative of most populated cluster from first peak at pH 7, j) representative of most populated cluster from second peak at pH 7. RMSD values for the representative structures: b – 4,6 Å; c – 3,9 Å; d – 4.3 Å; e – 5.1 Å.



**Fig. 5.** Distribution of distances between centers of mass for residues 126 and 168 (a, c, e, g) and residues 126 and 172 (b, d, f, h); at (a, b) pH 3; (c, d) pH5; (e, f) pH 7 (minimum 1); and (g, h) pH7 (minimum 2). The blue vertical line on each graph represents the distance between these residues in the capsid structure (10.7 Å between 168 and 126; 5.1 Å between 172 and 126). The distribution was calculated for all structures extracted from free energy surface minima at each pH

	pH 3	pH 5	pH 7 <sub>1</sub>	pH 7 <sub>2</sub>		pH 3	pH 5	pH 7 <sub>1</sub>	pH 7 <sub>2</sub>
HIS 10	5.1	5.6	6.0	5.5	GLU 104	3.3	3.0	3.3	3.4
HIS 11	5.3	5.5	6.1	5.5	ASP 115	3.9	4.3	5.3	4.8
HIS 12	4.7	5.4	5.8	5.5	<b>ASP 126</b>	<b>3.5</b>	<b>4.0</b>	<b>4.6</b>	<b>4.2</b>
HIS 13	5.1	5.5	5.9	5.7	<b>ASP 127</b>	<b>3.6</b>	<b>4.4</b>	<b>5.0</b>	<b>5.1</b>
HIS 14	4.9	5.5	6.3	6.4	ASP 139	3.6	4.4	4.9	4.1
HIS 15	5.0	5.7	6.0	5.8	HIS 148	4.9	5.2	5.4	5.2
ASP 29	3.0	3.9	3.9	3.8	HIS 157	6.6	6.3	6.9	7.4
ASP 32	2.8	3.9	4.5	5.2	<b>ASP 168</b>	<b>3.7</b>	<b>3.9</b>	<b>4.6</b>	<b>4.5</b>
ASP 33	2.9	3.6	3.9	4.3	<b>ASP 172</b>	<b>6.1</b>	<b>5.8</b>	<b>5.6</b>	<b>4.2</b>
ASP 34	3.0	4.1	5.0	3.4	ASP 194	2.0	2.8	3.1	3.3
ASP 35	3.5	4.1	4.6	2.6	HIS 195	6.3	6.5	6.8	7.3
ASP 37	3.3	4.0	4.3	3.6	GLU 203	4.3	5.7	5.9	6.1
HIS 38	5.6	6.1	6.3	6.2	ASP 208	2.8	3.4	3.5	3.9
ASP 70	4.9	5.1	5.2	5.6	GLU 210	4.6	5.1	5.1	5.0
ASP 78	2.3	3.2	3.8	3.8	GLU 223	4.3	4.9	5.3	5.3
GLU 94	2.9	3.7	4.4	4.0	ASP 228	4.4	5.0	5.3	4.9

**Table 1.** Titratable amino acid calculated pKa values.—pKa values were calculated using the conformations extracted from the minima of pH 3, 5 and 7. Blue and red indicate protonated and deprotonated, respectively. Amino acids Asp126, Asp127, Asp168, and Asp172 are in bold.