nature portfolio

Peer Review File

Atomic resolution structure of human papillomavirus bound to heparin.

Corresponding Author: Dr Susan Hafenstein

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript by Langley et al. presents a technically impressive cryoEM analysis of HPV16 bound to heparin at 1.9 Å resolution. The authors employ spophisticated image processing methods to identify and characterize deviations from icosahedral symmetry. Using a specialized version of sub-particle classification (Icosahedral Subvolume Extraction and Correlated Classification; ISECC) to characterize asymmetry in heparin binding and propose a new model for how binding occurs at a subset of available capsomer interfaces. The image processing is sophisticated, and the biological questions are important. The idea that partial heparin occupancy could stabilize the capsid and reduce flexibility is intriguing, and the manuscript contributes new data to an area limited by low resolution structures. That said, the work largely refines and expands on ideas proposed in earlier lower-resolution studies, and the novelty of the mechanistic conclusions may be overstated. Key inferences, including the number of heparin binding sites, the degree of capsid stabilization, and the interpretation of weak density—are not fully supported or clearly presented. These concerns limit the strength of the conclusions as currently written.

Major concerns:

The observed asymmetry in binding is presented as surprising, but HPV capsomers break Caspar-Klug symmetry by using pentamers at both pentavalent and hexavalent positions. This structural feature could explain much of the asymmetry and should be discussed.

Similarly, the claim that three of five binding sites are occupied seems to be inferred from ISECC analysis, not directly observed. Is there any cryoEM density showing just three of five heparins bound that could be shown???

The interpretation that heparin reduces capsid flexibility is based on a small change (~0.04 to 0.02 Å) in sub-particle deviation. While conceptually reasonable, the magnitude is close to noise and should be presented more cautiously.

Minor concerns:

The manuscript refers to an asymmetric unit with six L1 subunits. Although this is standard practice for T=7d papillomaviruses, it may be helpful to clarify why only six quasi-equivalent subunits are modeled, despite the T=7 triangulation number, as this may confuse readers unfamiliar with the T=7d capsid architecture.

Phrases such as “MD confirms hydrogen bonds” overstate what simulations can show. Similarly, the idea that negatively charged molecules are poorly visualized by cryoEM is misleading; weak density more likely reflects flexibility, partial occupancy, or radiation damage.

Reviewer #3

(Remarks to the Author)
The paper by Langley et al . is a comprehensive cryoEM structural analysis of HPV16 "quasivirus" (packaging CRPV genome), with and without bound heparin. Through their well described methods of subparticle masking and averaging, they

obtained atomic resolution (~1.9 angstrom) of the majority of the pentavalent and hexalent L1 major capsid proteins and built a model of the capsid. Their main findings are the confirmation of the heparin binding site around the pentavalent L1 capsomeres, identification/validation of L1 residues that bind to heparin, local L1 conformational changes in the presence of heparin, lack of global L1 capsid changes upon heparin binding (contrary to published AFM work), prediction and modeling of hydrogen bonding between flexible L1 N-terminal regions, and asymmetric cross-capsomere disulfide bonding between L1 monomers from adjacent pentamers. Overall the paper is well written and the atomic resolution reveal some important details. However some concerns regarding the actual samples and the conclusion of asymmetric disulfides shoukld be addressed. Further, the paper could be improved by using the data thus far to test some functional consequences of mutation- see below for detail.

Major concerns:

- There is no data on the quality of the quasivirus sample. Are these virions completely matured? Is the reason that all disulfides are not observed and asymmetric disulfides are concluded just because the sample is not completely oxidized/matured? What fraction of the L1 is present as reduced monomers? Also, how many particles contain the CRPV genome (capsid/genome ratio)? Could the presence of other DNA or no DNA in some particles be contributing to sample heterogeneity?

- The paper's impact could be improved if the structural data obtained were used to make targeted mutations to directly test the importance of N-terminal H-bonds (the beta sheets in Fig6 B,C and supplemental figure 8) in capsid stability. Could these regions be mutated to disrupt those predicted H-bond B-sheet networks? SOmething like a proline substitution or insertion. Would particles still form? Would they be infections? Less stable? Some further extension beyond just the observed/predicted structure would boost the impact of this work.

Other concerns:

There is very little difference between "top" and "side" views in Fig 3C.

The histidine residues in Fig5 look to have extra density (similar to His36 in Supplemental Fig 4) but this is not mentioned or discussed. If the prior MS/MS data on L1 is available can you reanalyze, looking for methylation or other modifications on His36 (and the other one near Cys 428?)

The blue H-bonds in Fig 6C,D,E,G are difficult to see. Labeling of some residues in these figures, and/or drawing a schematic of these H-bond networks, would help orient the reader as to what exactly were looking at. The nomenclature of the L1 chains (A,B,C, etc) and the figure panels (A,B,C, etc) make it difficult for the reader to understand, consider direct labeling of the L1 monomers in the figure and/or changing the labeling of the subpanels to avoid confusion.

Reviewer #4

(Remarks to the Author)
Human Papillomaviruses (HPV) form a group of more than 200 related viruses.
Most HPV infections are asymptomatic and cleared by the immune system without causing any health issues. However, persistent infection with high-risk types can lead to cellular changes that may result in cancer over time. The first step of the HPV life cycle involves engagement of sulfated glycans (heparan sulfate proteoglycans) by the virus capsid. In this study, Langley et al. investigate the interactions of HPV with the receptor ligand heparin using cryo-electron microscopy. The work builds on previous results by the same group (Structure, 2017), in which a structure of HPV-16 in complex with heparin had been determined at intermediate resolution. The lower resolution limited insights into specific interactions and did not allow analysis of possible conformational changes triggered by receptor engagement.

In this paper, the structure of the complex (using HPV-16 quasiviruses) is presented at near-atomic resolution (1.9 Å), allowing for the first time detailed insights into specific contacts as well as changes in the capsid protein. The new results thus advance the field by providing an excellent platform for understanding the architecture of the HPV-16 capsid, and they also allow for a reinterpretation of some structural features that were observed in a 3.1 Å-cryoEM structure of the HPV capsid alone published in 2021 by Hafenstein and colleagues.

I am somewhat less impressed with the level of detail that is available for the interactions of HPV with heparin. Heparin is shown to bind to the HPV capsid in between the fivefold capsomers and each hexavalent capsomer. The electron densities for heparin are not well defined (maybe due to multiple conformations), and in my opinion the density is not detailed enough to assign specific interactions and so the insights into heparin binding remain rather limited. I felt that the title of the manuscript raised expectations that the actual data do not quite fulfill.

Scientifically, the work is sound, and the structures presented are of very good quality. Proper criteria to determine resolution etc, have been applied.

Specific comments:

Line 443: “A single heparin molecule isolated from PDB 5W10 was fit into heparin density (ref 32)”. There is a mistake here, entry 5W10 is for a GAF domain and has nothing to do with reference 32. Maybe the authors mean 5W1O?

Several structures of HPV capsids, L1 pentamers and complexes with heparin are already available, including crystal structures of HPV L1 pentamers bound to heparin. It would seem to be important to compare these previous findings with the current data, in order to allow the reader to fully appreciate what is new in this manuscript.

Reviewer #5

(Remarks to the Author)
The manuscript by Langley et al. reports the cryo-EM determination of a new structure of the human papillomavirus (HPV16) complexed with heparin. The impact of the report is underscored by the fact that HPV, the virtually sole etiological agent responsible for cervical cancer (and several others), requires heparan sulfate to infect human cells, and heparin is a close structural analog that was found to competitively inhibit infection. Therefore, the findings reported in the manuscript bring us closer to an atomistic-level understanding of stages of HPV infection.

Judging from my background in computational biophysics and molecular modeling, the current contribution of simulation models to the manuscript's conclusions is quite modest. While simulating the entire capsid at atomistic resolution is no small feat, as it requires significant expertise and resources, it rarely brings reliable insights due to the short timescales and a high dependence on the initial structure. The one result that seems statistically reliable - the enrichment of chloride ions at the heparin binding sites - could be reasonably achieved with a simpler calculation of electrostatic potential around the molecule (note: I'd suggest refraining from the use of terms "electropositive" and "electronegative" as they most commonly denote the properties of elements, not of molecular species).

Given the novelty of the data, the available methods, and the biological questions related to the system, I have the following comments or questions:

- Is there a sufficient reason not to use density-guided simulation protocols to refine the low-quality parts of the model, such as the L1 C-termini or the N-terminal interactions? In particular, a masked density could be combined with a smaller, constrained subsection of the entire capsid; this would constitute a more unbiased validation of the proposed structural features.

- In a similar vein, given how poorly the heparin-binding site is resolved in the cryo-EM density (Fig 3), it would be straightforward to explore stable binding modes of heparin and heparan sulfate in atomistic simulations. In particular, the difference between heparin and heparan sulfate is of high relevance; it would be expected that heparin binds more strongly due to the electrostatic contacts, but the less negatively charged heparan sulfate is the actual biological moiety required for infection, so one should not discard the possible differences between the two. Having them explored side-by-side in atomistic MD simulations would greatly complement the main experimental takeaways of the article.

- It is hard to judge the relevance of motifs shown in Fig 7 and SI Figs 8 and 11 through an ensemble average over 500 ns (much less than the expected correlation times of binding/folding) without seeing the dynamics of the individual motifs. For example, seeing the temporal dynamics of beta-sheet content over time at each motif would help understand if the motifs form and unfold dynamically over time, or, e.g., start in the formed state and fall apart over time.

- It is understandable that a model with enforced icosahedral symmetry will not capture any structural details of the dsDNA genome, but is there experimental evidence that an empty HPV capsid devoid of disulfide bonds (as I understand was the case in the model) is stable on its own without the nucleic acid component? In other words, regardless of the structural details, does the atomistic model represent an inherently stable moiety, and is therefore expected to remain stable in a simulation? This would be worth commenting on briefly.

To summarize, the manuscript under consideration presents an interesting finding, and I believe that a more purposeful use of simulation methods can greatly enhance the reliability, interpretability, and usefulness of the results, complementing the experimental methods exactly where they lack sufficient resolution.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors thoughtfully addressed all concerns raised by this reviewer. Congratulations on some beautiful work.

Reviewer #3

(Remarks to the Author)

The revised manuscript has included some additional discussion and analysis to address prior reviewers' concerns. They better define the observed heparin density and explain issues regarding the asymmetric disulfide previously claimed. The main findings remain largely confirmatory of the prior work so it is difficult to see how far this advances the field beyond just the higher resolution (which is very impressive). Disagreement with published AFM work regarding capsid expansion and softening remains confusing. That work describes the importance of Lys442/443 in capsid softening upon HS engagement as well as exposure of the L2 RG1 epitope upon furin cleavage during infection. This work seems to suggest no role for these L1 residues in HS engagement so this remains an unresolved issue. Prior suggestions to include some mutagenesis of the observed L1 N-terminal beta-sheets were largely ignored, citing older work where large deletions of this region prevented assembly. I still believe that some further extension beyond just the observed/predicted structure would boost the impact of this work.

Reviewer #4

(Remarks to the Author)
In revising the manuscript, the authors have adequately addressed my concerns. The electron density for heparin appears to be more clear now, and the inclusion and discussion of earlier structures adds to the manuscript.

I support publication.

Reviewer #5

(Remarks to the Author)
My (arguably not an insider's) view on computational virology is that it's an interesting field reaching a degree of maturiry, and certainly delivers interesting insight where the domain of applicability of atomistic simulations overlaps with the mechanistic question at hand, including several cases mentioned by the authors; it however remains restricted to inferring biological effects from limited near-equilibrium fluctuations, although (as highlighted in the revision) often benefits statistically from the sampling multiplication provided naturally by the symmetry of viral systems. Either way, the main driver behind my criticism here is the desire for MD to become a valued and information-rich source of biologically relevant information that complements experimental insight, and articles like this present a perfect opportunity to make such a case; I'm sure authors share this broad perspective.

In that vein though, given that the proposed contribution explicitly focuses on the role of heparin, and the ambiguous physiological relevance of the experimental model, I would think that addressing the question of heparin/heparan sulphate binding is a natural problem for MD simulations to focus on, even on an approximate level (certain insights can be gained without multiple months of sampling, although ultimately it's the authors' call how much they are willing to trade off accuracy for speed). If future work is planned to address this, it obviously weakens the current manuscript in terms of novel insight, although the methodological improvements on the cryo-EM side - refining the heparin maps - partially make up for this loss.

To the changes implemented: I think the new panels added to illustrate L1 N-terminal heterogeneity convey the authors' message better. I also appreciate the clarification on the presence of disulfide bonds in the in silico model. However, the individual time series in the new Fig. S10d seem most consistent with a slow (possibly multiexponential) decay of the beta-sheet character of the C-F N-terminal pairing, with a majority of initially formed beta-sheets dissociating within the first 50 to 300 ns. Yet this apparent trend is not observed in the time averages of h-bond formation (panel c of the same figure); does this mean that h-bonding and beta-sheet formation are decoupled? How can these two trends be reconciled?

I also think that the information density in Figure 6 is too low to justify its inclusion as a separate figure. If possible, the new interfaces in Figure 5 could include h-bonds either color- or thickness-coded to indicate h-bond probability, unless structural heterogeneity prevents a clear presentation of results in such a way; at the moment, panel 6d is completely superfluous, and the reader has to switch between labels in Fig 5d-g and contact matrices in Fig 6 anyway.

Finally, I would still appreciate a comment from the authors on the perceived value of using, e.g., density cross-correlation to either restrain the sampling, or judge the veracity, of the conformations explored by the flexible regions. In my naive view, if the cross-correlation remained high or increased, this would directly show that the MD ensemble improves on the initial refinement, but perhaps there are reasons why such analysis would be misguiding or uninformative (?).

POINT-BY-POINT RESPONSE

Review round: 1

Manuscript ID: NCOMMS-25-42398-T

Title: Atomic resolution structure of human papillomavirus bound to heparin

The delay in resubmission was needed to perform additional assays to address Reviewers comments and concerns (point-by-point follows this summary), which greatly enhanced the manuscript. We express sincere thanks to the Reviewers. Changes include:

- Improvement of the heparin density, which was accomplished with iterative rounds of 3D classification, and resulted in a 3.4 Å heparin map with distinct and continuous heparin density.
- Reassessment of discontinuous disulfide bond density. Further investigation informed us that the lack of disulfide bond density was likely due to electron beam damage, which is known to occur rapidly, during the first five frames.
- Comparison of the different conformations of the L1 N-termini.
- Clarification of methods, specifically how the occupied binding sites were quantified and a description of our ISECC software capabilities.
- Updated figures to include residue labels and improved color coding.
- A comparison of the heparin binding site identified in the context of the capsid (our work presented here) with the heparin binding proposed in previously published works.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This manuscript by Langley et al. presents a technically impressive cryoEM analysis of HPV16 bound to heparin at 1.9 Å resolution. The authors employ spophisticated image processing methods to identify and characterize deviations from icosahedral symmetry. Using a specialized version of sub-particle classification (Icosahedral Subvolume Extraction and Correlated Classification; ISECC) to characterize asymmetry in heparin binding and propose a new model for how binding occurs at a subset of available capsomer interfaces. The image processing is sophisticated, and the biological questions are important. The idea that partial heparin occupancy could stabilize the capsid and reduce flexibility is intriguing, and the manuscript contributes new data to an area limited by low resolution structures. That said, the work largely refines and expands on ideas proposed in earlier lower-resolution studies, and the novelty of the mechanistic conclusions may be overstated. Key inferences, including the number of heparin binding sites, the degree of capsid stabilization, and the interpretation of weak density—are not fully supported or clearly presented. These concerns limit the strength of the conclusions as currently written.

Our presentation has been clarified for how we determined the number of heparin binding sites and the degree of capsid stabilization using custom software. We have edited text to describe that the number of occupied heparin binding sites and the capsid flexibility were quantified (not inferred). Importantly ISECC is unique in this regard in allowing assessments of flexibility and original

location of the bound ligand, and we have edited to make this clear (Lines 282-290, 955-966). We also further classified and refined the heparin subparticles to improve the heparin density. The sections discussing heparin density have been edited.

Major concerns:

The observed asymmetry in binding is presented as surprising, but HPV capsomers break Caspar-Klug symmetry by using pentamers at both pentavalent and hexavalent positions. This structural feature could explain much of the asymmetry and should be discussed.

The text has been clarified and the contribution of HPV L1 quasi-equivalence to asymmetry of binding (pentavalent vs hexavalent) is discussed further, Lines: 67-77 and in discussion. Additionally, we have cited work that illustrates the different structure of the two canyons due to quasi-equivalence.

Similarly, the claim that three of five binding sites are occupied seems to be inferred from ISECC analysis, not directly observed. Is there any cryoEM density showing just three of five heparins bound that could be shown???

The occupancy of three sites is determined, which has been clarified. Using ISECC, the occupied binding sites can be mapped back to the capsid, showing exactly where each heparin-occupied subparticle originated. This procedure gives us the location of heparin bound relative to each other on each capsid. The occupied sites are thus quantified and not inferred. This analysis has been clarified in Results (line 270).

The interpretation that heparin reduces capsid flexibility is based on a small change (~0.04 to 0.02 Å) in sub-particle deviation. While conceptually reasonable, the magnitude is close to noise and should be presented more cautiously.

We have added text to describe in more detail and add clarity to how this analysis is done. This is not an Angstrom measurement, but a ratio with no units, which was misleading, and this has been clarified. (Lines 500-508)

Minor concerns:

The manuscript refers to an asymmetric unit with six L1 subunits. Although this is standard practice for T=7d papillomaviruses, it may be helpful to clarify why only six

quasi-equivalent subunits are modeled, despite the T=7 triangulation number, as this may confuse readers unfamiliar with the T=7d capsid architecture.

> More description of T=7d and an explanation for using six L1 has been added to the Introduction Lines 67, 80.

Phrases such as "MD confirms hydrogen bonds" overstate what simulations can show. Similarly, the idea that negatively charged molecules are poorly visualized by cryoEM is misleading; weak density more likely reflects flexibility, partial occupancy, or radiation damage.

> Both the overstatement and the misstatement have been corrected. We have updated the language to "MD supports the presence of hydrogen bonds."

Reviewer #3 (Remarks to the Author):

The paper by Langley et al . is a comprehensive cryoEM structural analysis of HPV16 "quasivirus" (packaging CRPV genome), with and without bound heparin. Through their well described methods of subparticle masking and averaging, they obtained atomic resolution (~1.9 angstrom) of the majority of the pentavalent and hexalent L1 major capsid proteins and built a model of the capsid. Their main findings are the confirmation of the heparin binding site around the pentavalent L1 capsomeres, identification/validation of L1 residues that bind to heparin, local L1 conformational changes in the presence of heparin, lack of global L1 capsid changes upon heparin binding (contrary to published AFM work), prediction and modeling of hydrogen bonding between flexible L1 N-terminal regions, and asymmetric cross-capsomere disulfide bonding between L1 monomers from adjacent pentamers. Overall the paper is well written and the atomic resolution reveal some important details. However some concerns regarding the actual samples and the conclusion of asymmetric disulfides shoukld be addressed. Further, the paper could be improved by using the data thus far to test some functional consequences of mutation- see below for detail.

Major concerns:

- There is no data on the quality of the quasivirus sample. Are these virions completely matured? Is the reason that all disulfides are not observed and asymmetric disulfides are concluded just because the sample is not completely oxidized/matured? What fraction of the L1 is present as reduced monomers? Also, how many particles contain the CRPV genome (capsid/genome ratio)? Could the presence of other DNA or no DNA in some particles be contributing to sample heterogeneity?

Overwhelmingly, the capsids are filled. At the beginning of the reconstruction process, 2D classification can sort particles with and without DNA, and in our case even after further classification, no empty capsids were found in our prep. Thus, the HPV capsids included in the reconstruction contain DNA. Furthermore, the HPV used in our reconstruction all came from preparations that were properly matured according to the established Buck et al. 2006 protocol (24 hour incubation at 37°C).

It has been established (Hattne et al. 2018 and Pieri et al. 2021) that disulfide bonds are highly sensitive to beam damage. Likely most bonds are destroyed in the first frames of data collection, thus the absence of disulfide bonds is a consequence of data collection. All this information has been added to the manuscript. (Lines 346-348, 546-550)

- The paper's impact could be improved if the structural data obtained were used to make targeted mutations to directly test the importance of N-terminal H-bonds (the beta sheets in Fig6 B,C and supplemental figure 8) in capsid stability. Could these regions be mutated to disrupt those predicted H-bond B-sheet networks? SOmething like a proline substitution or insertion. Would particles still form? Would they be infections? Less stable? Some further extension beyond just the observed/predicted structure would boost the impact of this work.

Such mutations have already been made and investigated. We have added the citations to the manuscript. N-terminal Mutations and the consequences have been reported before and more text has been included along with the citation. (Lines 551-553)

Other concerns:

There is very little difference between "top" and "side" views in Fig 3C.

Figure 3C has been updated with more distinct views.

The histidine residues in Fig5 look to have extra density (similar to His36 in Supplemental Fig 4) but this is not mentioned or discussed. If the prior MS/MS data on L1 is available can you reanalyze, looking for methylation or other modifications on His36 (and the other one near Cys 428?)

The residues previously shown in Fig5 are residues 174-176 (PCT) and 427-429 (ACQ). Proline 174 appears to be a histidine with extra density because the C alpha and the continuing protein chain were removed. The figure has been edited to include residue 173 to avoid this confusion. Histidines throughout the model were examined for extra density, and only histidine 36 has extra density.

The blue H-bonds in Fig 6C,D,E,G are difficult to see. Labeling of some residues in these figures, and/or drawing a schematic of these H-bond networks, would help orient the reader as to what exactly were looking at. The nomenclature of the L1 chains (A,B,C, etc) and the figure panels (A,B,C, etc) make it difficult for the reader to understand, consider direct labeling of the L1 monomers in the figure and/or changing the labeling of the subpanels to avoid confusion.

Figure 6 was reworked to include more zoomed-in views, residue labels, and more visible hydrogen bonds. The ribbon diagram was removed from the zoomed-in views to improve the view of the model built into the cryoEM density.

Reviewer #4 (Remarks to the Author):

Human Papillomaviruses (HPV) form a group of more than 200 related viruses. Most HPV infections are asymptomatic and cleared by the immune system without causing any health issues. However, persistent infection with high-risk types can lead to cellular changes that may result in cancer over time. The first step of the HPV life cycle involves engagement of sulfated glycans (heparan sulfate proteoglycans) by the virus capsid. In this study, Langley et al. investigate the interactions of HPV with the receptor ligand heparin using cryo-electron microscopy. The work builds on previous results by the same group (Structure, 2017), in which a structure of HPV-16 in complex with heparin had been determined at intermediate resolution. The lower resolution limited insights into specific interactions and did not allow analysis of possible conformational changes triggered by receptor engagement.

In this paper, the structure of the complex (using HPV-16 quasiviruses) is presented at near-atomic resolution (1.9 Å), allowing for the first time detailed insights into specific contacts as well as changes in the capsid protein. The new results thus advance the field by providing an excellent platform for understanding the architecture of the HPV-16 capsid, and they also allow for a reinterpretation of some structural features that were observed in a 3.1 Å-cryoEM structure of the HPV capsid alone published in 2021 by Hafenstein and colleagues.

I am somewhat less impressed with the level of detail that is available for the interactions of HPV with heparin. Heparin is shown to bind to the HPV capsid in between the fivefold capsomers and each hexavalent capsomer. The electron densities for heparin are not well defined (maybe due to multiple conformations), and in my opinion the density is not detailed enough to assign specific interactions and so the insights into heparin binding remain rather limited. I felt that the title of the manuscript raised expectations that the actual data do not quite fulfill.

To improve the heparin density, the heparin subparticles were reprocessed. Iterative rounds of 3D classification with blush regularization were used to sort particles with and without heparin density. This approach worked well, and the improved results are now presented in the manuscript. The heparin density is continuous, and the heparin model fits nicely into the heparin density. Contacts were identified in the new map; however, reevaluation of contacts in the updated map identifies the same residues identified previously.

Scientifically, the work is sound, and the structures presented are of very good quality. Proper criteria to determine resolution etc, have been applied.

Specific comments:

Line 443: "A single heparin molecule isolated from PDB 5W10 was fit into heparin density (ref 32)". There is a mistake here, entry 5W10 is for a GAF domain and has nothing to do with reference 32. Maybe the authors mean 5W1O?

The correct PDB entry ID is 5W1O, and this mistake has been corrected.

Several structures of HPV capsids, L1 pentamers and complexes with heparin are already available, including crystal structures of HPV L1 pentamers bound to heparin. It would seem to be important to compare these previous findings with the current data, in order to allow the reader to fully appreciate what is new in this manuscript.

Text and a figure have been added to address the Reviewer's concern. The six panel figure compares previous HPV + heparin structures, specifically an Xray crystallography structure from Dasgupta et al. 2011 (PDB 5W1O) and a 4.3Å cryoEM structure from Guan et al. 2017 (PDB 5KEQ). Dasgupta et al. 2011 solved the structure of HPV16 pentamers bound to heparin made of 8-10 monosaccharide units. Notably, this is different from the structure solved in this paper, as we used the entire icosahedral virus and incubated the virus with heparin composed of 40-51 subunits. The heparin binding sites seen in the Dasgupta et al structure were compared to our heparin binding site, and there was very little overlap. The cryoEM structure from Guan et al. 2017 was composed of icosahedral HPV16 and incubated with heparin composed of 10-30 subunits. The heparin binding sites are very similar, however the Guan et al. structure was limited by the resolution.

Please see lines 288-314 in Results and S. Fig 9.

Reviewer #5 (Remarks to the Author):

The manuscript by Langley et al. reports the cryo-EM determination of a new structure of the human papillomavirus (HPV16) complexed with heparin. The impact of the report is underscored by the fact that HPV, the virtually sole etiological agent responsible for cervical cancer (and several others), requires heparan sulfate to infect human cells, and heparin is a close structural analog that was found to competitively inhibit infection. Therefore, the findings reported in the manuscript bring us closer to an atomistic-level understanding of stages of HPV infection.

Judging from my background in computational biophysics and molecular modeling, the current contribution of simulation models to the manuscript's conclusions is quite modest. While simulating the entire capsid at atomistic resolution is no small feat, as it requires significant expertise and resources, it rarely brings reliable insights due to the short timescales and a high dependence on the initial structure. The one result that seems statistically reliable - the enrichment of chloride ions at the heparin binding sites - could be reasonably achieved with a simpler calculation of electrostatic potential around the molecule (note: I'd suggest refraining from the use of terms "electropositive" and "electronegative" as they most commonly denote the properties of elements, not of molecular species).

The authors genuinely appreciate feedback on the MD simulation component of the manuscript by an expert in computational biophysics. However, we strongly disagree with the statement that MD “rarely brings reliable insights due to the short timescales and a high dependence on the initial structure.” The literature is bursting with examples of all-atom MD simulations (describing systems ranging in size from small to massive) making essential contributions to scientific discoveries based on the hundreds of nanoseconds to microseconds timescale. This is because many functionally-relevant motions occur within this temporal regime. Based on examples from co-author Hadden-Perilla’s work alone, simulations of the intact hepatitis B capsid have been shown to faithfully reproduce the nanosecond-timescale dynamics reported by solid state NMR (10.3389/fmolb.2021.807577), and have been used to show that dilation of capsid pores is not a prerequisite for externalization of signaling peptides (10.1016/j.jbc.2023.105104), as well as to explain the structural basis behind point mutations that drive capsid assembly, confer drug resistance (10.1128/jvi.01082-18), and impair assembly (10.1021/acschembio.0c00277).

To address skepticism of readers who may not be familiar with the literature on intact capsid simulations and increase confidence in MD-derived data presented here, we include a new paragraph in the introduction that describes the reliability of state-of-the-art atomistic simulations and highlights their power as a complementary method to cryoEM, consistent with their role in the present work. We further include a reference to a recent review on the twenty-year history of intact capsid simulations and the many key contributions they have made to basic and applied science despite limited timescale and dependence on initial structure (10.1016/j.sbi.2025.103082).

While the reviewer is correct that calculation of electrostatic potential around the HPV capsid structure could have been used to characterize electrostatic nature of the surrounding space, this calculation would have to be repeated on millions of MD-derived

conformers to achieve a result similar to that produced by chloride occupancy analysis, owing to the significant conformational heterogeneity of the pentavalent-hexavalent canyon revealed by simulations. Rather than using electrostatics calculations to infer the behavior of nearby ions or small molecules like heparin around the capsid, the chloride occupancy analysis directly demonstrates the capsid's ability to influence chemical species within its environment, while simultaneously accounting for both protein flexibility and solvent dynamics. Ion behavior around capsids has been shown to represent a proxy for small-molecule interactions in other viruses, including predicting the gateways for nucleotide import (10.1038/ncomms15959,10.1371/journal.pbio.3001015) and locations of RNA-capsid contact (10.1021/ct3002128) in experimentally-validated studies. The results presented in this manuscript combining MD with cryoEM thus adds to this list of experimentally-valid work. We have added this commentary to the Discussion section for the benefit of the reader.

The authors use the term "electropositive" to refer to the electrostatic nature of the space surrounding the capsid-protein structure, not to elements or molecular species. The term "electropositive" by the common physics definition means electrically positive or having a positive electric charge. We agree that from a chemistry perspective, this term also refers to the tendency (of an element) to donate electron density or to lose electrons and form positive ions in chemical reactions. We believe that the physics definition, of common knowledge to the journal's audience, represents the most clear and precise description of the presented results, rendering the term appropriate in this context.

Given the novelty of the data, the available methods, and the biological questions related to the system, I have the following comments or questions:

- Is there a sufficient reason not to use density-guided simulation protocols to refine the low-quality parts of the model, such as the L1 C-termini or the N-terminal interactions? In particular, a masked density could be combined with a smaller, constrained subsection of the entire capsid; this would constitute a more unbiased validation of the proposed structural features.

We are uncertain what the Reviewer means by "low quality" of the L1 termini and do our best to address the concern.

Because of the capsid architecture there are multiple conformations for the N-terminus. For example, in chain D there is a stretch of L1 across the threefold that has weaker density, but it is continuous at lower contours. To understand these conformations, we have added new panels to Figure 5. Also in Figure 5, the panels that show the H-bonds were rendered in ribbon that obscured the quality of the map density. New panels have been added so that the quality of the density is obvious.

- In a similar vein, given how poorly the heparin-binding site is resolved in the cryo-EM density (Fig 3), it would be straightforward to explore stable binding modes of heparin and heparan sulfate in atomistic simulations. In particular, the difference between

heparin and heparan sulfate is of high relevance; it would be expected that heparin binds more strongly due to the electrostatic contacts, but the less negatively charged heparan sulfate is the actual biological moiety required for infection, so one should not discard the possible differences between the two. Having them explored side-by-side in atomistic MD simulations would greatly complement the main experimental takeaways of the article.

> The cryoEM map has been reprocessed using further 3D classification and blush-regularization. This approach improved the density which is sharper and higher quality. Please see new figure 3.
>
> The reviewer is correct that exploring the high-resolution details of heparin sulfate and heparin binding using MD simulations is of great interest and is indeed planned future work of the authors. While straightforward from a computational perspective, this endeavor represents a full-scale project that will take months simply to collect MD trajectories, even for a capsid fragment (the minimal required subunit, the starfish, is ~1 million atoms in explicit solvent). The time needed to carry out this work would significantly delay publication of the present manuscript, whose focus is to report a sub-2 Å reconstruction of heparin-bound HPV, which is an important advancement for structural biology of the capsid even if heparin itself is not resolved. The role of MD in this work is simply to supplement and support results obtained by cryoEM, and data-guided simulations of the heparin-bound starfish subunit and intact capsid will be presented in a follow-up study.

- It is hard to judge the relevance of motifs shown in Fig 7 and SI Figs 8 and 11 through an ensemble average over 500 ns (much less than the expected correlation times of binding/folding) without seeing the dynamics of the individual motifs. For example, seeing the temporal dynamics of beta-sheet content over time at each motif would help understand if the motifs form and unfold dynamically over time, or, e.g., start in the formed state and fall apart over time.

> The authors are grateful to the reviewer for catching an oversight in the manuscript. With regards to the amount of sampling, all MD analyses have been updated to reflect an extended amount of sampling corresponding to 710 ns of the intact capsid. However, all MD-derived data presented in this work is aggregated over 60 quasi-equivalent copies of the L1 chains (A, B, C, D, E, and F) present in the capsid, such that the sampling of these local dynamics represents as much as 60x710ns = 42.6 microseconds of cumulative sampling, an order of magnitude more than is commonly reported in state-of-the-art MD studies. We have clarified this detail in the Methods section and relevant figure captions.
>
> The hundreds of nanoseconds timescale is sufficient to track hydrogen bonding dynamics, as well as the interaction between two 9-residue regions of neighboring proteins, even in flexible systems, especially given essentially 60 replicates. The formation of beta sheets in this region do not involve complex folding, but the subtle adjustment of backbone phi/psi dihedrals to enable two extended regions of the proteins already in proximity to form hydrogen bonds. To address the reviewer's question regarding behavior over time, we have included an additional figure in the supplement

showing the time-evolution of secondary structure in these regions across all 60 quasi-equivalent chain copies.

- It is understandable that a model with enforced icosahedral symmetry will not capture any structural details of the dsDNA genome, but is there experimental evidence that an empty HPV capsid devoid of disulfide bonds (as I understand was the case in the model) is stable on its own without the nucleic acid component? In other words, regardless of the structural details, does the atomistic model represent an inherently stable moiety, and is therefore expected to remain stable in a simulation? This would be worth commenting on briefly.

The MD model of the HPV capsid used for analysis in the present work contains all 360 possible Cys428-Cys175 disulfide bonds, and we have added this fact to the manuscript text for improved clarity. The L1 capsid protein self-assembles in vitro to form stable particles, even in the absence of genome and the L2 protein (10.1038/s41579-021-00617-5), thus the MD model represents an inherently stable moiety and is indeed stable during MD simulation. Although equilibration and stability of the HPV capsid model is discussed in detail in the preprint that reports the MD simulation and biophysical characterization of the particle based on the same dataset, we have included a note on the model's appropriateness and stability to improve the reader's confidence in the results reported in the present manuscript, which are intended simply to complement the cryoEM reconstruction.

To summarize, the manuscript under consideration presents an interesting finding, and I believe that a more purposeful use of simulation methods can greatly enhance the reliability, interpretability, and usefulness of the results, complementing the experimental methods exactly where they lack sufficient resolution.

In the revised manuscript, MD data is used to complement and enhance interpretability of cryoEM results in three areas lacking sufficient resolution: N-terminal secondary structure and inter-chain hydrogen bonds, maximum C-terminal contribution to protein-DNA contacts, and side chains most likely to bind heparin.

Nature Communications manuscript NCOMMS-25-42398A

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors thoughtfully addressed all concerns raised by this reviewer. Congratulations on some beautiful work.

Reviewer #3 (Remarks to the Author):

The revised manuscript has included some additional discussion and analysis to address prior reviewers' concerns. They better define the observed heparin density and explain issues regarding the asymmetric disulfide previously claimed. The main findings remain largely confirmatory of the prior work, so it is difficult to see how far this advances the field beyond just the higher resolution (which is very impressive). Disagreement with published AFM work regarding capsid expansion and softening remains confusing. That work describes the importance of Lys442/443 in capsid softening upon HS engagement as well as exposure of the L2 RG1 epitope upon furin cleavage during infection. This work seems to suggest no role for these L1 residues in HS engagement, so this remains an unresolved issue.

Thank you. We agree that we should include the Lys442 as a potential contact since it is listed as a contact in the MD work. To illustrate the proximity of Lys442 to the heparin cryoEM density please see new panels added to supplementary figure 8. We have also added to the discussion the possible importance of Lys442/443 and cited the AFM paper again where the text has been added.

The L2 RG1 epitope located on the N-terminal residues 17-36 is not resolved, due to L2 flexibility and asymmetric incorporation. Thus, potential structural changes to L2 that occurred due to heparin binding could not be resolved using cryoEM.

We agree that there are seemingly confusing issues with the AFM work that describes increased stability, enlargement of the capsid, and “softening” of the capsid. Concerning stability, we show that the addition of heparin enhances stability (line 187-197, and supplemental figure 6.) The resolution mapping showed increased stability between capsomers and consistent stability of hexavalent capsomers. With HDX-MS, Feng et al. 2024 also found that heparin binding decreased flexibility in the N- and C-terminal connecting arms between capsomers. However, the apical loops of pentavalent capsomeres became more flexible after heparin incubation. These changes likely agree with the AFM result that probably would not pick up a loop flexibility change.

The measured enlargement by AFM indicates an increase in height from 52 to 54 nm (Fig. 2 of Feng et al 2024). We showed in Goetschius et al. 2021 that the HPV capsid varies in diameter from 56.8nm to 58.5nm, and based on the same analysis, HPV bound to heparin varies in diameter from 56.8nm to 58.4nm (Lines 199-201).

Concerning the AFM finding of "softening," which is defined by Feng et al. 2024 as making the virus easier to deform. We have used a focused refinement based analysis to measure flexibility of the capsid (Goetschius 2021); however, we cannot use cryoEM to test the "softening" of the capsid reported by AFM.

Although the AFM paper used the same length of heparin, they did use pseudovirus and VLP; whereas we used quasivirus throughout.

We have added more to the discussion to address the Reviewer's concerns on this points.

Prior suggestions to include some mutagenesis of the observed L1 N-terminal beta-sheets were largely ignored, citing older work where large deletions of this region prevented assembly. I still believe that some further extension beyond just the observed/predicted structure would boost the impact of this work.

We failed to report necessary details, which have now been added. All the predicted H-bonds that are proposed to be the most critical to the stability are main chain interactions. (Line 298). We now make it clear that most predicted hydrogen bonds occur between main chain atoms. Hydrogen bond frequencies have been added by incorporating MD and cryoEM results into the new improved Figure 5.

The deletion of only 10 residues disrupts assembly (Chen et al 2000). This detail should have been included and now has been added (Line 85). Figure 5 shows how critical the first nine amino acids are likely to be, as this deletion would disrupt all H-bonds in Figure 5B and four H-bonds in Figure 5D (Lines 442-443).

Reviewer #4 (Remarks to the Author):

In revising the manuscript, the authors have adequately addressed my concerns. The electron density for heparin appears to be more clear now, and the inclusion and discussion of earlier structures adds to the manuscript.

I support publication.

Reviewer #5 (Remarks to the Author):

My (arguably not an insider's) view on computational virology is that it's an interesting field reaching a degree of maturiry, and certainly delivers interesting insight where the domain of applicability of atomistic simulations overlaps with the mechanistic question at hand, including several cases mentioned by the authors; it however remains restricted to inferring biological effects from limited near-equilibrium fluctuations, although (as highlighted in the revision) often benefits statistically from the sampling multiplication provided naturally by the symmetry of viral systems. Either way, the main driver behind my criticism here is the desire for MD to become a valued and information-rich source of biologically relevant information that complements experimental insight, and articles like this present a perfect opportunity to make such a case; I'm sure authors share this broad perspective.

> We thank the reviewer for their thoughtful perspective on the role of MD simulations in computational virology and agree with the broader goal of establishing MD as a rigorous and informative complement to experimental approaches. We also appreciate the recognition that MD can provide valuable insight within its domain of applicability, including for large viral assemblies where symmetry enables enhanced statistical sampling.
>
> In the present work, however, our use of MD was intentionally focused and limited in scope. The objective of this study is the structural characterization of the HPV-heparin complex by cryoEM. Accordingly, the MD simulations were designed to complement the experimental data by providing atomistic insight into regions of the capsid that are not well resolved in the density (owing to their dynamics), rather than to directly model the heparin-bound complex or to interrogate binding mechanisms in detail.

In that vein though, given that the proposed contribution explicitly focuses on the role of heparin, and the ambiguous physiological relevance of the experimental model, I would think that addressing the question of heparin/heparan sulphate binding is a natural problem for MD simulations to focus on, even on an approximate level (certain insights can be gained without multiple months of sampling, although ultimately it's the authors' call how much they are willing to trade off accuracy for speed). If future work is planned to address this, it obviously weakens the current manuscript in terms of novel insight, although the methodological improvements on the cryo-EM side - refining the heparin maps - partially make up for this loss.

> As noted by the reviewer, addressing the molecular determinants of heparin/heparan sulfate binding by MD is indeed an important and compelling direction for future work. However, performing simulations of the HPV-heparin complex that are sufficiently converged and interpretable at the scale of the intact

capsid (16.8 million atoms in explicit solvent) represents a substantial undertaking that is far beyond the scope of the current study. We respectfully note that short-timescale simulations of the complex would be unlikely to provide robust or mechanistically meaningful insight into this system, given the size of the capsid, the intrinsic flexibility of the binding regions, and the heterogeneous nature of glycosaminoglycan interactions. Even the minimal capsid subunit required to study the heparin-HPV interactions (i.e. the starfish) represents a million-atom system. Additional simulations at this scale are not a trivial request. The reviewer's suggestion to trade accuracy for speed would, in practice, require coarse-grained modeling. However, the loss of atomic resolution inherent to such approaches precludes detailed characterization of the binding interaction, and it is therefore unclear that a coarse-grained model would offer greater mechanistic insight than the available low-resolution experimental data.

Importantly, even within this limited scope, the MD simulations provide independent, orthogonal support for key features inferred from the cryo-EM data, including residue-level interactions and dynamic behavior in flexible regions of the capsid. We believe this complementary use of MD strengthens the manuscript without shifting its focus away from the experimental advances, which constitute the primary contribution of the work. We agree that future studies integrating MD simulations of the HPV-heparin complex will be valuable for further elucidating binding mechanisms and functional implications, and we are actively pursuing this direction.

To the changes implemented: I think the new panels added to illustrate L1 N-terminal heterogeneity convey the authors' message better. I also appreciate the clarification on the presence of disulfide bonds in the in silico model. However, the individual time series in the new Fig. S10d seem most consistent with a slow (possibly multiexponential) decay of the beta-sheet character of the C-F N-terminal pairing, with a majority of initially formed beta-sheets dissociating within the first 50 to 300 ns. Yet this apparent trend is not observed in the time averages of h-bond formation (panel c of the same figure); does this mean that h-bonding and beta-sheet formation are decoupled? How can these two trends be reconciled?

We thank the reviewer for this insightful observation and for highlighting the apparent discrepancy between panels in Fig. S10. The original version of Fig. S10c included all hydrogen bonds between the corresponding residues, including those involving side chains. As only protein backbone interactions are relevant to defining β-sheets, the inclusion of side chain interactions obscured the direct relationship between hydrogen bonding and quaternary structure stability the reviewer was looking for. We note that the presence of hydrogen bonds between N-termini does not necessarily imply β-sheet formation, as the latter also requires a specific range of backbone dihedral angles to satisfy the β region of Ramachandran space.

To show a more clear characterization of the β-sheet, we have updated all panels of Fig. S10. Panel (a) now shows the longest run observed for the two-strand sheet, and panel (b) now shows the secondary structure of the two chains individually. Panel (c) now shows a timeseries for all 60 C-F copies, which accounts for both the dihedral angle and hydrogen bonding criteria to compute the fraction of residue pairs exhibiting β-sheet characteristics. Panel (d) shows this data averaged over all C-F copies. Based on this revised panel (d), the reviewer can see that indeed the average β-sheet fraction simultaneously present across C-F copies (which is ~0.8 in the starting structure) decays over the first 200 ns of simulation before stabilizing ~0.6. Since a fraction of 1.0 is all nine pairs meeting β-sheet criteria, ~0.8 is 7 pairs and ~6 is five pairs on average across C-F copies. The broad standard deviation shows that the fraction fluctuates significantly, as the inter-chain interaction is dynamic. However, these fluctuations are in the idealized backbone geometry plus hydrogen bonding criteria that define a β-sheet. As revised panels (e) and (f) show, C-F hydrogen bonding (both in terms of backbone residues involved in β-sheet formation and auxiliary contacts between side chains) remains persistent, such that all 60 copies of the C-F N-terminals remain continuously associated, thus stabilizing the inter-capsomer interaction as predicted by cryoEM.

I also think that the information density in Figure 6 is too low to justify its inclusion as a separate figure. If possible, the new interfaces in Figure 5 could include h-bonds either color- or thickness-coded to indicate h-bond probability, unless structural heterogeneity prevents a clear presentation of results in such a way; at the moment, panel 6d is completely superfluous, and the reader has to switch between labels in Fig 5d-g and contact matrices in Fig 6 anyway.

In light of the reviewer's excellent suggestion, we have revised Figure 5 to improve the connection between structural representations and quantitative analysis. The hydrogen bonds represented in the structure are now color coded based on the frequency with which they were observed during MD simulations. Figure 6 has been moved to the supplementary information. Worth noting is that Figure 6 includes details of all the suspected hydrogen bonds between L1 N-termini and their frequencies according to MD, beyond the examples highlighted in Figure 5, which only shows H-bonds observed from cryoEM.

Finally, I would still appreciate a comment from the authors on the perceived value of using, e.g., density cross-correlation to either restrain the sampling, or judge the veracity, of the conformations explored by the flexible regions. In my naive view, if the cross-correlation remained high or increased, this would directly show that the MD ensemble improves on the initial refinement, but perhaps there are reasons why such analysis would be misguiding or uninformative (?).

We thank the reviewer for this thoughtful suggestion and for highlighting approaches that more tightly integrate MD simulation with experimental density. We agree that density-guided simulations and cross-correlation analyses can be powerful in contexts where the goal is to refine or validate a structural model directly against cryoEM data. However, this is not the objective of the MD simulations presented here.

First, the MD simulations were deliberately designed to be unbiased, with the goal of probing the intrinsic conformational landscape of the HPV capsid. In this context, applying density restraints during simulation would introduce bias toward the experimental model and obscure precisely the type of conformational variability that the simulations are intended to reveal. It is this conformational variability that can provide insights into structural regions that have limited resolution in the cryoEM reconstruction, and it is in this context that MD data is used to complement cryoEM in the present work.

Next, the cryoEM reconstruction represents the heparin-bound capsid whereas the MD simulation represents the apo capsid. This comparison allowed us to assess the frequency with which structural features of the bound state are sampled in the apo ensemble, key to understanding whether heparin induces non-native conformations of the capsid that may be relevant to host cell adhesion and entry. Remarkably, we identified similarities between the heparin-bound capsid and apo ensemble, including in regions where cryoEM resolution was limited. The reported similarities are on a residue level (e.g. hydrogen bonding patterns, orientations of heparin-binding residues, and potential DNA contacts).

We believe the reviewer is inquiring about performing cross-correlation analysis between the cryo-EM density and a density calculated from MD conformers, explicitly for the flexible regions of the structure. Just as conformational heterogeneity leads to reduced experimental resolution for these flexible regions, a density calculated from thousands of heterogenous MD conformers will exhibit reduced resolution. We find the residue-level comparisons we present to be significantly more informative than cross-correlation of low-resolution densities; high cross-correlation between low-resolution densities does not imply that they represent the same underlying structural ensemble. Furthermore, because the cryo-EM and MD datasets represent different systems (heparin-bound versus apo capsids), the cryo-EM data cannot serve as a direct benchmark for assessing the validity of conformations sampled in the MD simulations.

Next, we are unsure what the reviewer is referring to when they say “show that the MD ensemble improves on the initial refinement.” We clarify that the intact capsid MD simulation was not used to refine the cryoEM structure presented in this work. As described in the manuscript text, the improved resolution of this HPV-heparin structure compared to previous reconstructions arises from advancements in instrumentation, image processing, and subparticle refinement techniques, highlighting the state-of-the-art in cryoEM.

Finally, we emphasize the agreement between two orthogonal approaches (cryo-EM and independently derived MD simulations) at the level of specific interactions, including in flexible or low-resolution regions, as a meaningful and stringent form of validation. We believe this residue-level convergence is more informative than global density comparisons, which would necessarily average over heterogeneous conformations and obscure the atomic detail needed to gain new insights into the dynamic nature of the capsid and its structural state upon host receptor binding.