Mechanism of HIV-1 Capsid Rupture and Uncoating by Reverse Transcription

Kuntal Ghosh, Manish Gupta and Gregory A. Voth*

Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Illinois 60637, USA

Keywords: HIV-1, reverse transcription, capsid uncoating, molecular dynamics, coarse-graining

*Corresponding Author: gavoth@uchicago.edu

Abstract

One of the key events in the HIV-1 life cycle is reverse transcription, during which single-stranded viral RNA (ssRNA) is converted into double-stranded DNA (dsDNA). This process occurs inside the mature virus capsid and, once it reaches a critical threshold, drives capsid rupture. This uncoating is essential for infection because it releases viral genetic material into the host cell nucleus. Despite its importance, many mechanistic details of this process remain poorly understood. To address this gap, we develop a multiscale computational method for simulating reverse transcription inside the capsid, termed Coarse-Grained Kinetic Monte Carlo (CG-KMC). CG-KMC stochastically adds deoxynucleotide triphosphates (dNTPs) to the coarse-grained RNA model, enabling stepwise growth of DNA inside the HIV-1 capsid. We implement this method within an integrative coarse-grained framework that combines a "bottom-up" capsid model with a "top-down" representation of the viral RNA/DNA genome. Our simulations phenomenologically capture and predict diverse capsid rupture pathways during reverse transcription. The resulting ruptured structures closely match previously identified cryo-ET images. We further perform an extensive analysis of the rupture process, examining its mechanistic and kinetic aspects as well as the role of capsid–DNA interactions. Our findings

illuminate how different capsid–DNA conditions give rise to distinct rupture pathways, which differ from ruptures due to simple outward pressure expansion models from within the capsid.

Introduction

The HIV-1 life cycle comprises a sequence of tightly regulated steps, from particle assembly and protein biosynthesis to the integration of the viral genome into the host DNA for infection (*1*). Among these, the process of reverse transcription (RT) – the conversion of flexible viral single-stranded RNA (ssRNA) genome into stiff double-stranded DNA (dsDNA) within the mature capsid – is particularly critical yet still not fully understood (*2*). Experimental studies (*3-5*) have probed RT; however, a unified consensus on its timing and the sequence of accompanying events has yet to emerge. It is also fairly well established that RT within the viral capsid is intricately linked to the capsid uncoating (*6-11*), involving its partial or complete disassembly. However, the precise timeline of RT relative to capsid destabilization and nuclear entry remains unresolved. As RT proceeds, the encapsulated ssRNA gradually transforms into dsDNA, exerting considerable but not homogeneous mechanical force on the capsid interior wall that contributes to capsid rupture.

A mature HIV-1 capsid is composed of capsid (CA) protein dimers that self-assemble into a range of morphologies (*12, 13*). In its canonical mature form, the capsid adopts a closed fullerene-like, cone-shaped shell built from approximately 250 hexamers and precisely 12 pentamers (*14, 15*). This shell encloses the viral ssRNA genome together with several essential enzymes, including protease, reverse transcriptase, and integrase (*16, 17*). It was originally thought that RT is initiated and completed in the cytoplasm (*2, 4*), even before the capsid reaches the nuclear pore complex (NPC), which characterizes the boundary between the nucleus and the cytoplasm. More recent studies, however, suggest that RT proceeds as the capsid is docking at or traversing the NPC, and may even be completed inside the nucleus (*6, 18*). These observations underscore a critical gap: although the biochemical stages of RT are well characterized, their temporal coordination with capsid structural transitions during and after nuclear entry remains poorly defined. Furthermore, the detailed mechanistic pathways leading to the capsid uncoating process are not well understood.

To address these challenges, we developed and deployed in this work large-scale coarse-grained (CG) models of the RT process inside a mature HIV-1 capsid. Coarse-graining (*19, 20*) involves constructing a simplified representation of a much larger atomistically resolved system, with the goal of retaining its primary physical features and effective interactions at a fraction of the computational cost. All-atom and very large scale molecular dynamics (AA MD) has been used previously (*15, 21, 22*) to examine the HIV-1 capsid shell. One of these papers (*22*) presented a more detailed model that revealed that when the CA lattice-bound anion IP6 and a model for the genome package were included in the AA MD simulation model, the capsid shell developed distinct inhomogeneous strain (lattice compression and expansion) patterns. This result was also suggestive of non-uniform capsid rupture behavior as the internal RT process proceeds.

On the other hand, simulating the complete RT process within a capsid with AA MD is computationally intractable. However, in the past few years CG-MD has had considerable success in modeling HIV-1 capsid maturation and host interactions (*13, 23-27*), as well as simulating the entry of the capsid into the nuclear pore complex (NPC) (*28, 29*). Building on these developments, in this work, we develop an integrative CG model of a "bottom-up" HIV-1 capsid (derived from the AA interactions) with a top-down genome model (encompassing both RNA and DNA). We develop and implement a stochastic Kinetic Monte Carlo (*30, 31*) (KMC) method at the CG level, which we denote as Coarse-grained Kinetic Monte Carlo (CG-KMC), to model RT and subsequent capsid rupture and uncoating. We show that our model accurately captures capsid rupture as RT progresses, in striking agreement with previously reported experimental evidence as imaged by cryo-electron tomography (*5*). We also investigate and characterize the impact of the varying capsid-DNA interactions on the rupture pathways.

Results

Capsids stay intact during initial stages of reverse transcription

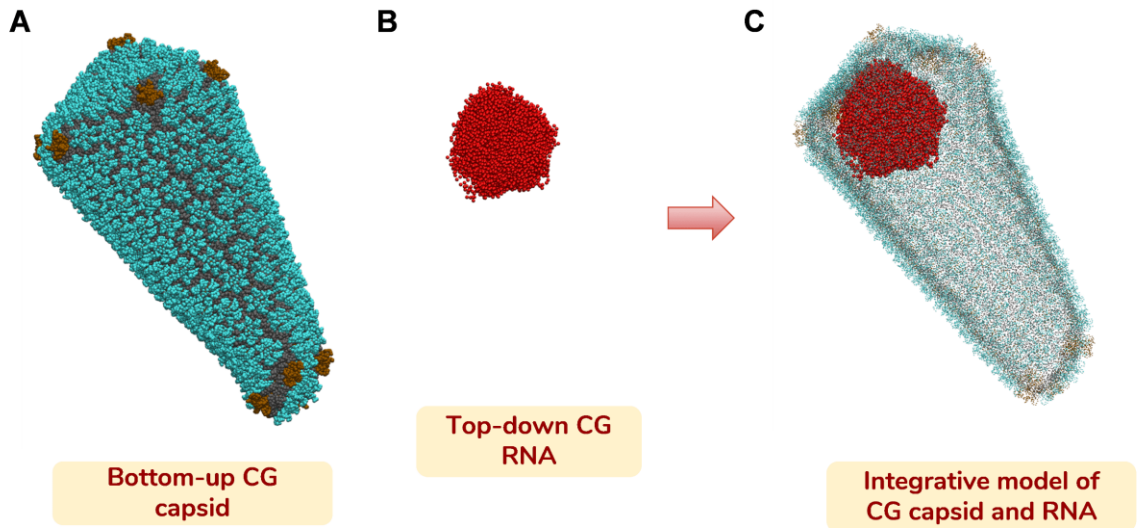

Figure 1: Overview of the integrative CG model: (A) A CG capsid model constructed systematically using a "bottom-up" approach from underlying atomistic trajectories (NTD of hexamers in cyan, NTD of pentamers in ochre, CTD of both in gray) (B) A highly condensed "ball" of RNA (red), designed in a "top-down" fashion (C) The RNA is embedded inside the capsid to generate the initial configuration of the capsid-RNA complex, subsequently used for CG-KMC simulation. Pentamers are colored brown in the CG capsid.

A condensed "ball" of RNA (two single-stranded copies) was embedded inside the cone-shaped HIV-1 capsid (**Figure 1**). In concordance with experiments (*32*), capsid uncoating does not begin instantaneously in the presence of dNTPs. Several critical processes involving enzymes such as reverse transcriptase (*33-35*) (incorporated implicitly) facilitate the growth of RT products from single-stranded copies of RNA. Stages I and II (described in the Methods section) depict the growth of the RNA-DNA hybrid, formed stochastically with the addition of one-site CG dNTPs to the RNA beads, and the subsequent cleavage of the hybrid. The reader is directed to the Methods section and the SI for a discussion of the polypurine tracts (PPT) (*36*). As is evident in **Figure 2**, growth of the RNA-DNA hybrid takes place inside the capsid, without leading to distortions in the capsid lattice (*37*). It is also important to note that both the initial ssRNA and the RNA-DNA hybrid strands are modeled to be sufficiently flexible, thus allowing them to "coil up" as they keep growing as a result of the CG-KMC algorithm. Even with respect to Stage III in our model (described in the Methods section) which generates dsDNA, there is no rupture at the beginning. This can be corroborated well by the experimental notion that the growth of dsDNA leads to rupture only after a certain time interval (~ 6-10 hours) (*5, 32*), and

after attaining a certain length of the genome (~ 3.5 kb or about one-third of the genome size) (*32*).

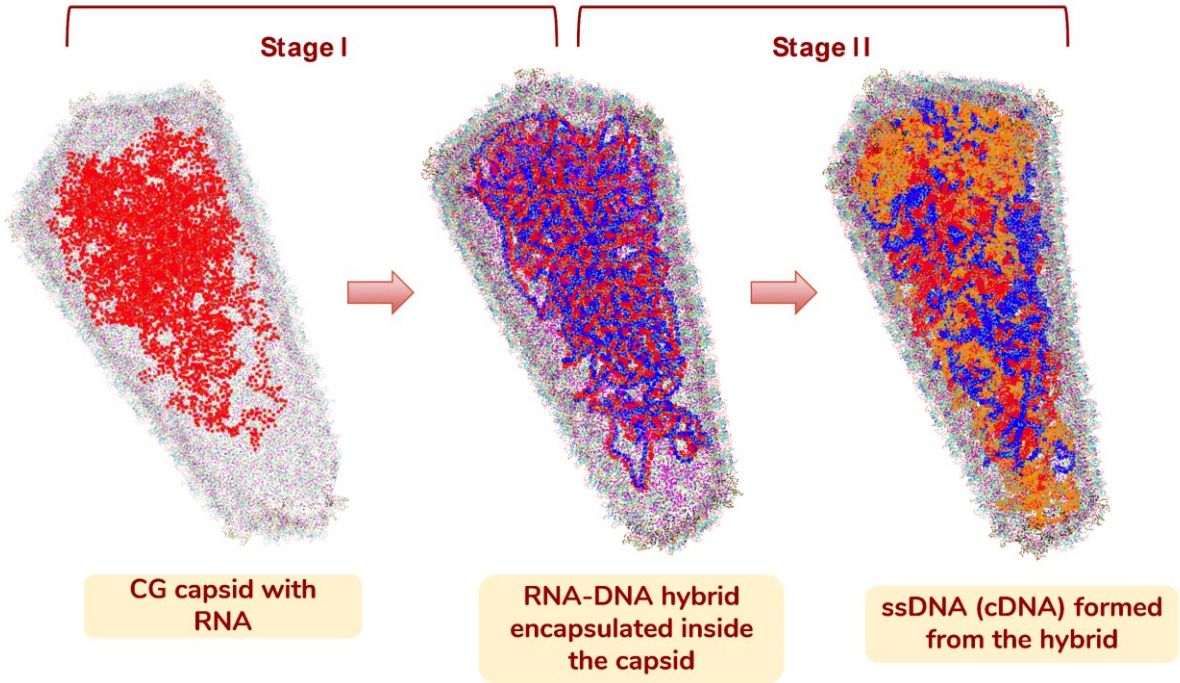

Figure 2: Stages I and II of RT simulated by CG-KMC depicted here: (A) The initial configuration of the capsid with a condensed "ball" of RNA (red) gradually occupies the available volume inside the capsid after a few CG-MD steps. (B) In Stage I, dNTPs stochastically and sequentially bind to the RNA beads using CG-KMC, forming the RNA-DNA hybrid (blue-red complex), thus implicitly incorporating the effect of reverse transcriptase enzyme. (C) Stage II depicts the subsequent degradation of the RNA-DNA hybrid, in a type of "cleavage", which yields a single-stranded complementary DNA (cDNA) (orange).

Capsid starts uncoating as the double-stranded DNA grows

In Stage III of RT, CG-KMC stochastically keeps adding the dNTPs to the single-stranded cDNA, mimicking the growth of dsDNA. As mentioned before and experimentally established, dsDNA is much stiffer (i.e, has a much higher persistence length) than the flexible ssRNA. As dNTPs are sequentially incorporated into the cDNA, a corresponding strand assembles on the template, with sequential bead–bead connections introduced to model the underlying base-pair interactions. To ensure rigidity and stiffness of the DNA, harmonic interactions were used for modeling the base pair interactions. As the dsDNA grows, it progressively occupies the capsid interior, repeatedly impinging on the inner surface of the capsid lattice. This, however, does not lead to rupture in the initial stages. This behavior in the model can be interpreted as the implicit effect of the nucleocapsid (NC) protein, which is experimentally known for condensing the

dsDNA (*8, 32, 38*). However, the amount of NC protein present cannot condense the full length of the dsDNA, at which point the internal mechanical pressure imparted by the dsDNA on the walls of the capsid increases, eventually leading to rupture. At certain genome lengths, the dsDNA appears to exert substantial pressure on the capsid walls, as indicated by a marked reduction in nucleic acid strand mobility. Each CG-MD trajectory samples a distinct conformational state of the cDNA as it undergoes RT, thereby generating heterogeneous patterns of mechanical stress imparted by the dsDNA on the capsid interior, as opposed to a uniform outward pressure. Through this ensemble of trajectories, our model captures the inherently stochastic nature of reverse transcription and capsid uncoating with both efficiency and at least semi-quantitative fidelity.

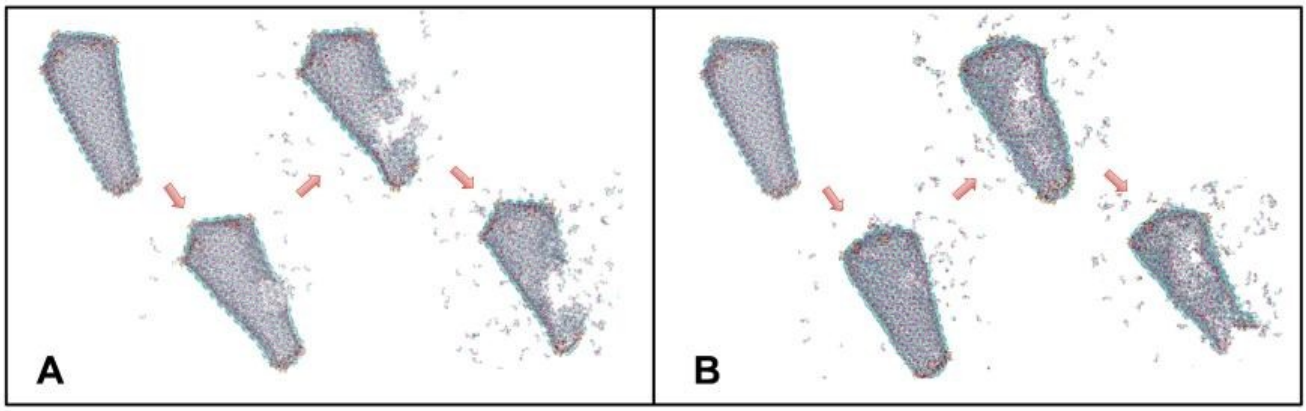

Figure 3: Two representative pathways of capsid uncoating are illustrated: (A) Rupturing, predominantly marked by detachment of hexamers from the midsection of the capsid lattice, progresses from the formation of small local defects to extensive lattice disassembly (B) An alternative pathway initiates with local defects near the wider end of the capsid and propagates longitudinally through the midsection toward the narrow tip, ultimately leading to significant lattice disruption.

The pathway depicted in **Figure 3A** shows initial stages of uncoating along the walls of the capsid. Conversely, the pathway depicted in **Figure 3B** reveals stages of uncoating proceeding from the wider and even narrower ends of the capsid. Rupturing at either end of the cone-shaped capsid is typically characterized by detachment of pentamers located in the high-curvature regions. **Figures S1A-B** of the Supporting Information depict gradual detachment of pentamers from the wider end of the cone-capsid lattice, whereas **Figures S1C-D** depict gradual uncoating near the narrow end, both directly at the tip and through longitudinal propagation of cracks initiated at the midsection of the capsid. In each of these pathways, uncoating is initiated by the formation of small local patches, succeeded by the progression of uncoating along the

cracks, leading to much larger patches. Eventually the capsid disintegrates significantly, thus facilitating release of the viral genome. Two representative pathways of capsid rupture, including the dynamic growth of dsDNA from ssRNA, are shown in **Movie S1**.

We also computed the Green-Lagrange strain patterns (*39*) (**Figure S2)** to highlight the regions of the capsid walls experiencing increased strain, as an indication of the tentative uncoating domains. Here, we show the evolution of strain patterns at CG resolution for an illustrative case. As RT progresses, strain patterns appear in local patches on the capsid walls, encompassing the hexamers. As this local patch expands along the walls, initial stages of uncoating can be seen, with the most strained CG beads getting detached from the lattice. As this progresses, the beads experiencing the most strain get further detached from the walls, eventually leading to partial disassembly. While our group's previous study (*22*) examined the strain patterns that could be associated with capsid uncoating, it did not provide a systematic framework for inducing uncoating driven by RT. By contrast, the present work has an explicit dynamic description of RT inside the capsid, and the predicted rupture patterns are in agreement with experimentally determined lattice separation (correlated with the strain) patterns (*22*).

We also explored capsid rupture patterns as a consequence of expanding an idealized "virtual object" inside the capsid. The reader is directed to the SI for a detailed discussion. This approach does not include an explicit description of the viral genome inside the capsid. **Figure S4** depicts capsid rupture pathways obtained by isotropically expanding a single virtual sphere and two virtual spheres stacked on top of each other within the capsid. As the rupture patterns show, some configurations are not very realistic, and unlike the RT-driven capsid rupture, the virtual object expansion approach appears more deterministic. The growth of the same geometric object leads to identical rupture pathways across multiple replicas. Thus, to obtain a distribution of rupture pathways which CG-KMC naturally produces, one must tune the geometric parameters of each virtual object. This underscores the necessity of using an explicit description of the dsDNA growth inside the capsid, which can effectively drive anisotropic capsid rupture, as done in this work. We emphasize again that the capsid model utilized here is one derived bottom-up from realistic atomic-scale interactions via a rigorous coarse-graining approach based on statistical thermodynamics.

Impact of capsid-DNA interactions on uncoating pathways

Capsid-DNA interactions must play a critical role in determining the nature of the disassembly pathway. However, not much is known about the exact nature of interactions governing the initial rupturing of the capsid walls due to the growing dsDNA strands. To this end, we systematically carried out an analysis probing into the role of the CA-DNA interactions, and how it governs the nature of capsid rupture. Modified Lennard-Jones (LJ) interactions (Eq 1) and Gaussian interaction potentials (Eq 2) were used to model the CA-DNA interactions (see Methods). Owing to the nature of that aspect of our model, we explored different attraction parameters in the modified LJ interactions. It must be pointed out that the same set of parameters were used for the CG simulations before the critical genome size is reached, at which point, the attractive potential is switched on (which we call an experimentally motivated "switch" , used to mimic the release of the condensing effect of NC). It is this switch that triggers capsid rupture once the growing dsDNA reaches the experimentally determined critical length—approximately one-third of the full genome (*32*). We subsequently carried out multiple sets of CG-KMC simulations starting from the initial capsid and RNA model, with three different attractive parameters given by ε = 2.0, 3.5 and 5.0 kcal/mol (in Eq 1).

Figures 4-6 illustrate snapshots of representative uncoating pathways for each of these values, with a visual representation of the growth of dsDNA inside the capsid interior from the cDNA, and subsequent rupture. We also report the fraction of the genome corresponding to a given configuration, denoted as χ , defined as the number of CG beads that have undergone RT divided by the total number of CG beads in the genome (3000 in this case). As RT progresses, dsDNA punctures the wider end of the capsid lattice, prominently looping out (**Figure 4**). Such observations can also be made for ε = 3.5 (**Figure 5**) and ε = 5.0 (**Figure 6**); however, rupturing is much more pronounced for the latter. As these dsDNA protrusions puncture the capsid walls, strands of cDNA can sometimes be seen leaking out. For a more comprehensive analysis, we look at multiple rupturing pathways as a function of the CA-DNA interactions in **Figures 7-9** (discussed again in a later section on kinetics). In brief, a higher attractive parameter often leads to formation of more local patches and aggressive rupturing, effectively "crumbling" the capsid lattice. The partially disassembled capsid structures, as predicted by our CG models, agree well

with tomographic slices of ruptured cores observed 8-10 hours under standard endogenous RT conditions (*5*).

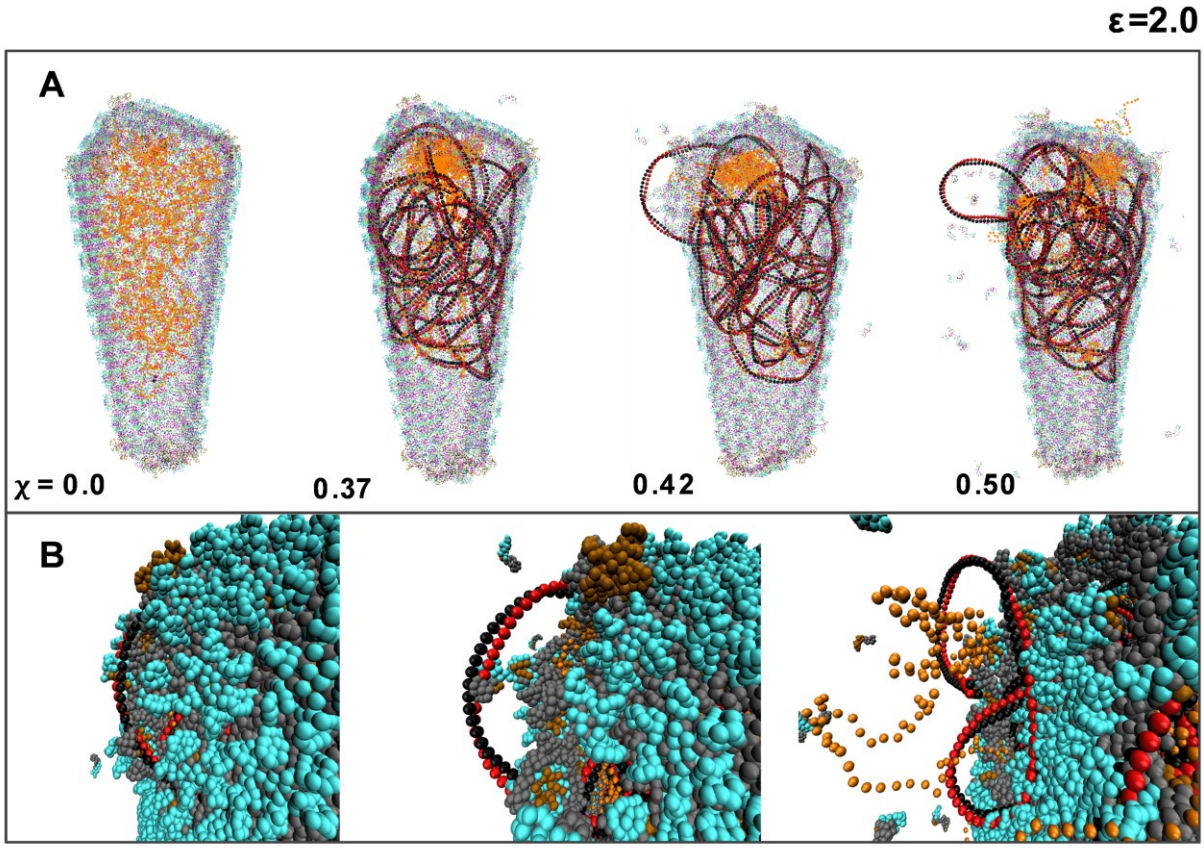

Figure 4: For ε = 2.0 kcal/mol, the growth of dsDNA is depicted: (A) Snapshots of the capsid enclosing the genome are shown. Initially, the genome is the cDNA (orange). CG-KMC is carried out on the cDNA to form dsDNA (black–red hybrid). As dsDNA grows, rupturing occurs near the wider end of the capsid, and the dsDNA eventually loops out prominently. (B) A closer view of the uncoating process. Clear signs of dsDNA protrusion can be seen, accompanied by detachment of both capsid pentamers and hexamers from the lattice. As RT progresses, the capsid walls begin to disassemble, releasing loops of dsDNA. Strands of cDNA also appear to loop out. Only one representative pathway for ε = 2.0 kcal/mol is shown. Here, χ represents the fraction of the total genome size, and the approximate CG timestamp corresponding to a configuration can be given as $\chi \times 3000 \times 10^4$ CG steps.

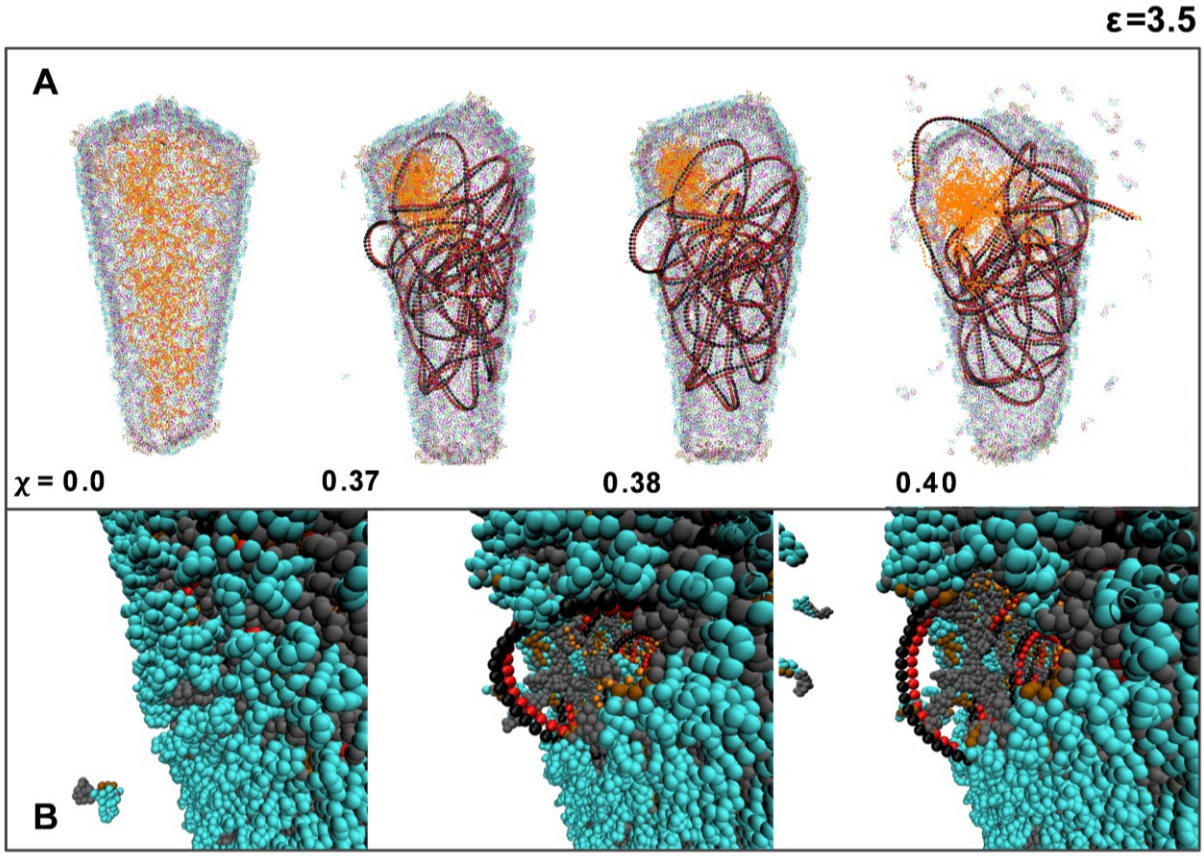

Figure 5: For ε = 3.5 kcal/mol, the growth of dsDNA is depicted: (A) Snapshots of the capsid enclosing the genome are shown. Initially, the genome is the cDNA (orange). CG-KMC is carried out on the cDNA to form dsDNA (black–red hybrid). As dsDNA grows, rupturing occurs near the midsection of the capsid, and the dsDNA eventually loops out prominently. (B) A closer view of the uncoating process. Clear signs of dsDNA protrusion can be seen, accompanied by detachment of capsid hexamers from the lattice. As RT progresses, the capsid walls begin to crumble, releasing loops of dsDNA. Only one representative pathway for ε = 3.5 kcal/mol is shown. Here, χ represents the fraction of the total genome size, and the approximate CG timestamp corresponding to a configuration can be given as $\chi \times 3000 \times 10^4$ CG steps.

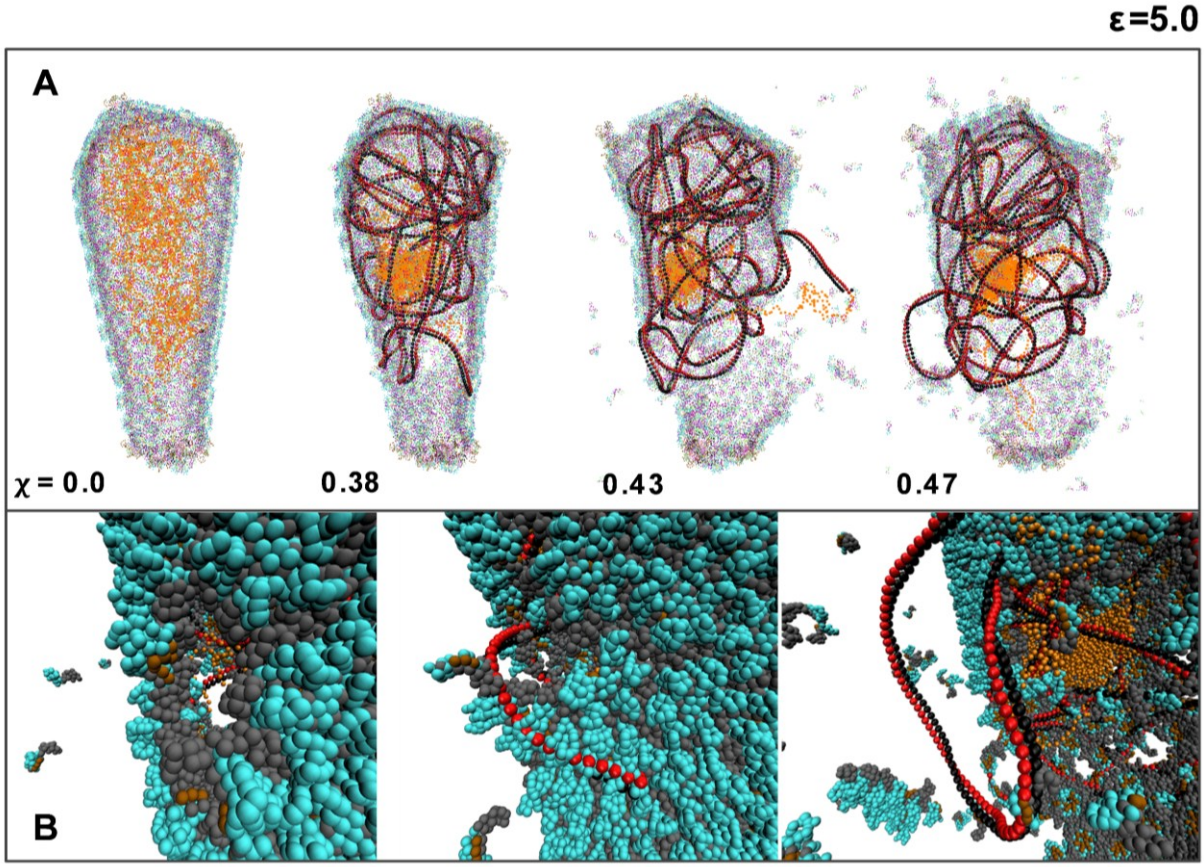

Figure 6: For ε = 5.0 kcal/mol, the growth of dsDNA is shown: (A) Snapshots of the capsid enclosing the genome are presented. Initially, the genome consists of cDNA (orange). CG-KMC is carried out on the cDNA to form dsDNA (black–red hybrid). As dsDNA grows, severe rupturing occurs near the midsection of the capsid, ultimately leading to partial disassembly, with clear dsDNA protrusions visible. (B) A closer view of the uncoating process shows distinct dsDNA protrusion accompanied by pronounced uncoating as the strands loop outward. Uncoating is markedly more substantial than for ε = 2.0 and 3.5 kcal/mol. Only one representative pathway for ε = 5.0 kcal/mol is shown. Here, χ represents the fraction of the total genome size, and the approximate CG timestamp corresponding to a configuration can be given as $\chi \times 3000 \times 10^4$ CG steps.

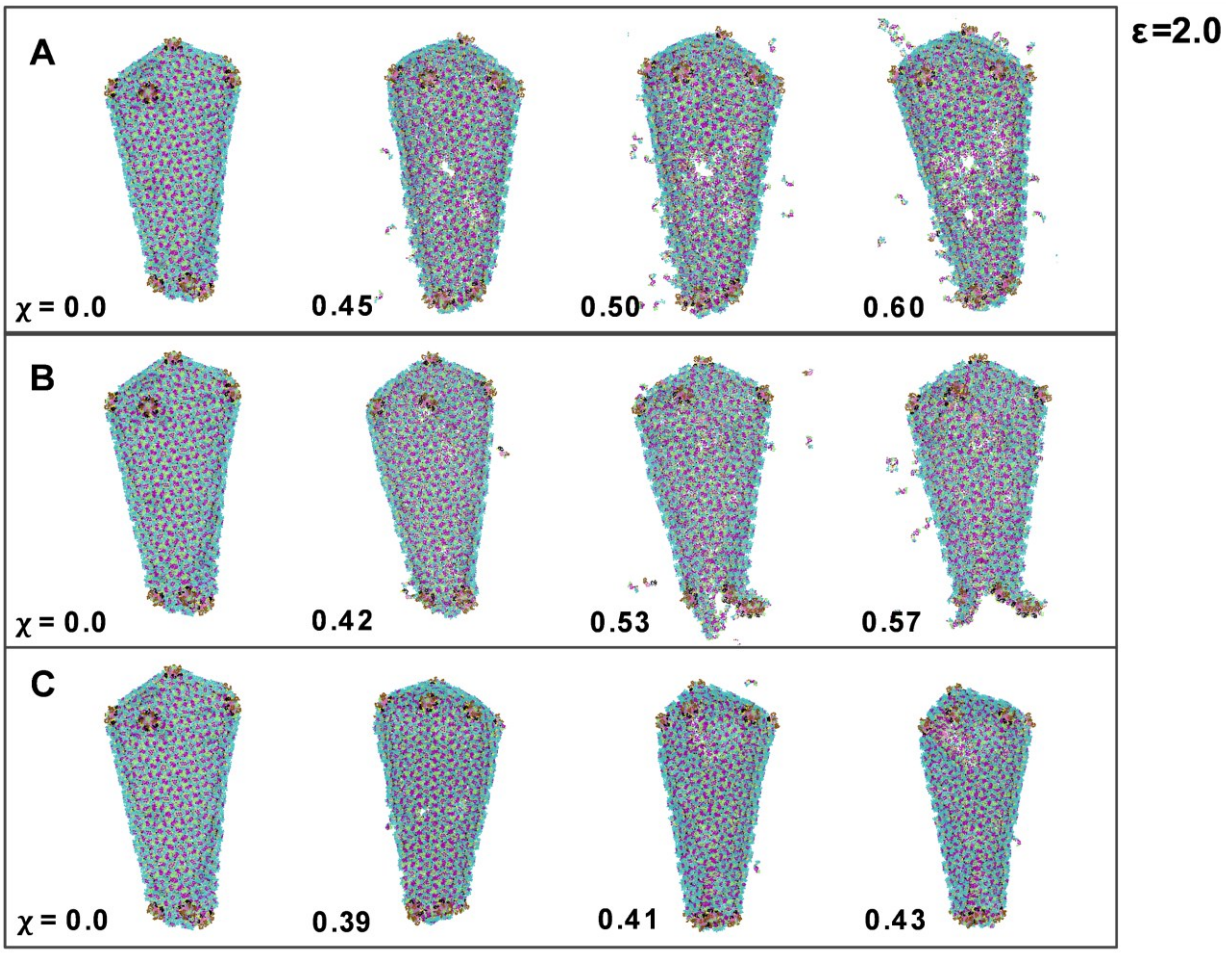

Figure 7: Three representative capsid uncoating pathways for ε = 2.0 kcal/mol: (A) Small cracks first emerge in the midsection of the capsid lattice as RT proceeds, gradually expanding into larger openings. (B) Cracks initiate at the narrow end of the cone-shaped capsid and are marked by the detachment of both pentamers and hexamers. (C) Very slow, gradual crack formation occurs near the wider end of the capsid, with rupture propagating along these cracks toward the midsection of the lattice. Here, χ represents the fraction of the total genome size, and the approximate CG timestamp corresponding to a configuration can be given as $\chi \times 3000 \times 10^4$ CG steps.

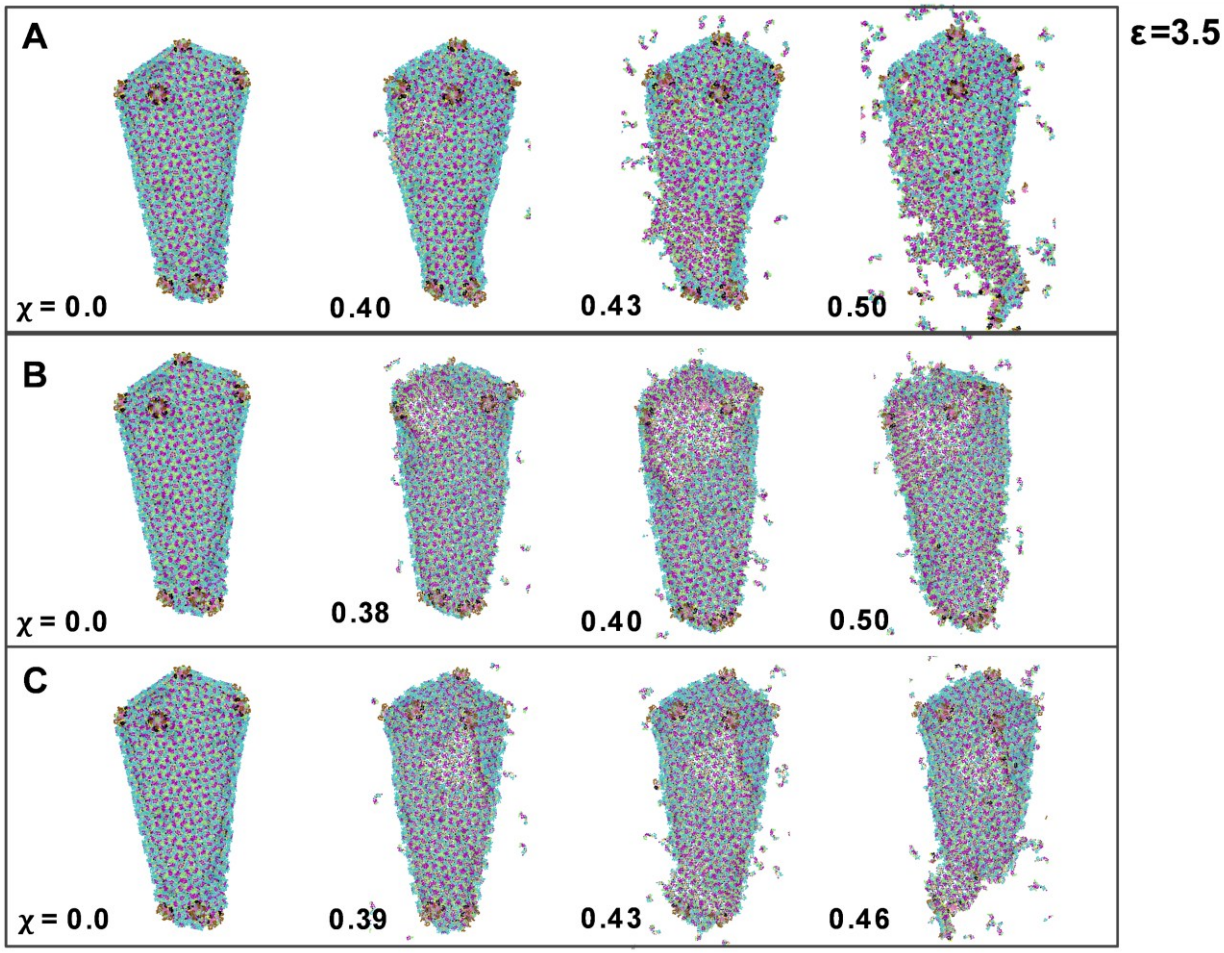

Figure 8: Three representative capsid uncoating pathways for ε = 3.5 kcal/mol: (A) Cracks first appear in the midsection of the capsid lattice as RT progresses, gradually widening into larger openings that lead to partial disassembly. (B) Openings that arise near the wider end of the capsid steadily expand and are accompanied by the detachment of both pentamers and hexamers. (C) Cracks originating in the midsection propagate toward the tip of the capsid, ultimately culminating in pronounced disassembly at the tip. Here, χ represents the fraction of the total genome size, and the approximate CG timestamp corresponding to a configuration can be given as $\chi \times 3000 \times 10^4$ CG steps.

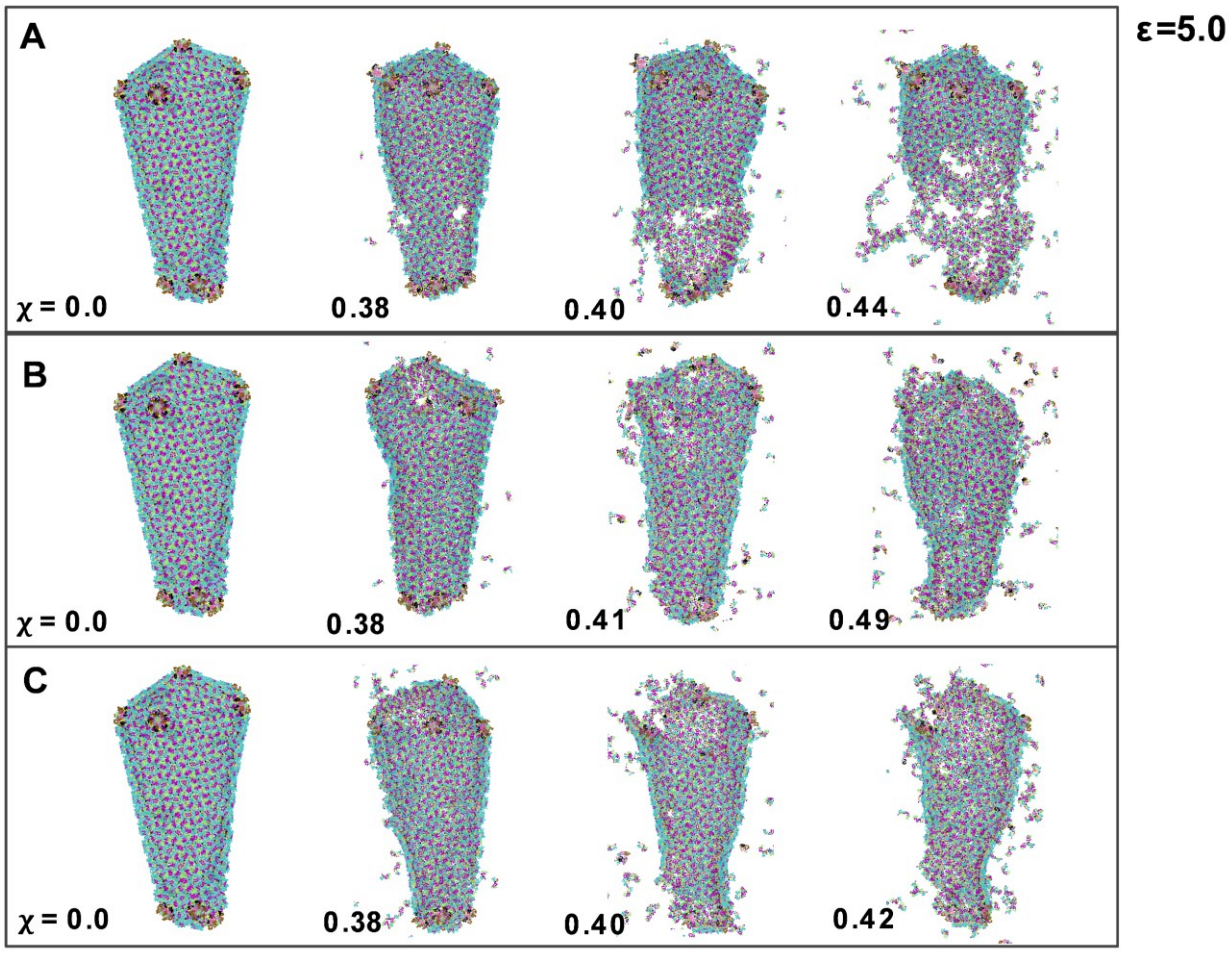

Figure 9: Three representative capsid uncoating pathways for ε = 5.0 kcal/mol: (A) Cracks originating in the midsection of the lattice eventually tear apart the capsid, resulting in a crumbling, heavily ruptured structure. (B) Distortions and cracks form openings interspersed along the capsid walls, with the most prominent ones appearing near the wider end. (C) Cracks emerging near the wider end of the lattice gradually widen and propagate along the vertical axis, ultimately causing extensive rupturing of the entire capsid. Here, χ represents the fraction of the total genome size, and the approximate CG timestamp corresponding to a configuration can be given as $\chi \times 3000 \times 10^4$ CG steps.

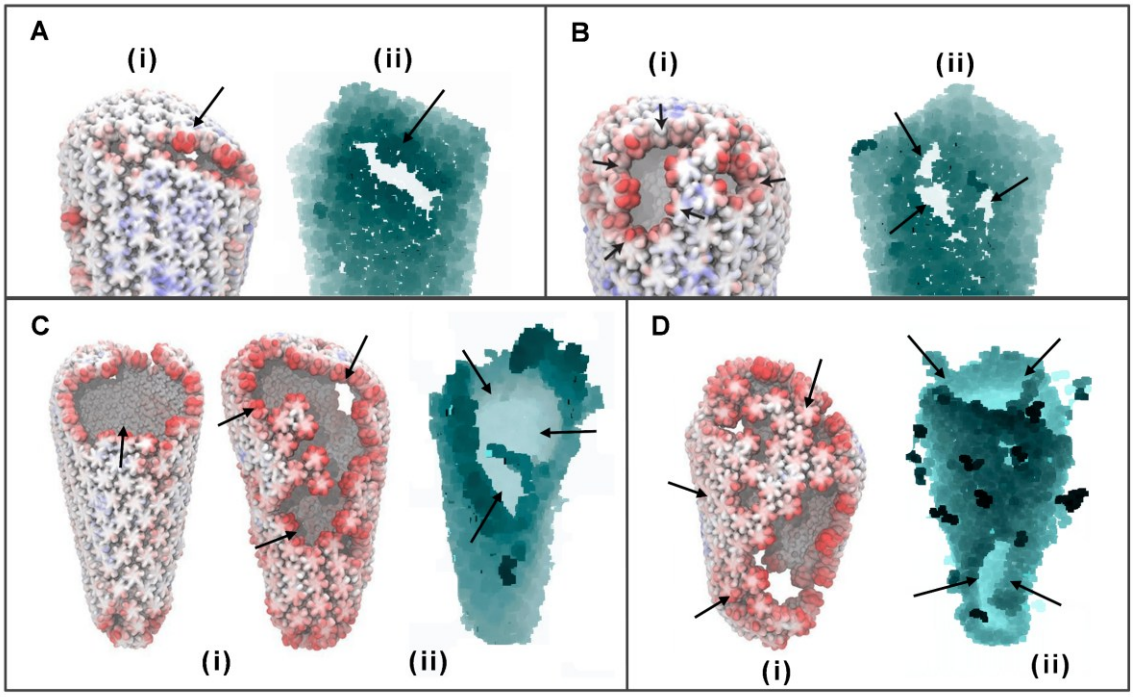

Figure 10: Comparison of CG-KMC predicted capsid rupturing patterns with previously reported cryo-ET images: Subfigures in (i) are cryo-ET images taken from Ref. (*22*) and those in (ii) show CG-KMC rupturing patterns predicted in this work. Points of rupture are indicated by black arrows. ε = 2.0 kcal/mol for A(ii), ε = 3.5 kcal/mol for B(ii) and C(ii) and ε = 5.0 kcal/mol for D(ii). The groups of dark particles in D(ii) are fragments ruptured from the capsid walls..

We also present here in **Figure 10** a comparison between representative CG-KMC predicted capsid rupturing patterns with the cryo-ET images of HIV-1 cores during rupture, as reported in Ref. (*22*). Structures from panels A8, A9, A10 and A12 of Figure 5 and panel 6 of Figure S5 of Ref. (*22*) are used for comparison. As is evident, CG-KMC predicts cracks near the wider end of the cone-shaped lattice in the regions of higher curvature, which aligns well with the cryo-ET experiments (as in **Figures 10A-B**). More severe rupturing patterns, such as wider holes near the broader end of the cone (**Figure 10C**), or cracks spanning the full length of the capsid (**Figure 10D**) are also predicted.

In addition to a visual analysis of the pathways, we generate root-mean-square-displacement (RMSD) profiles of these pathways as a function of the genome size, with the intact capsid as the reference. From **Figure 11** it is clearly seen that there is a sharp increase in the RMSD at the critical genome size that corresponds to capsid rupturing. The relative sharpness or "rate" of disassembly progressively increases as the CA-DNA attractive interactions increase. This is consistent with mechanical and biophysical expectations:

strengthened CA–DNA interactions increase the effective confinement of the dsDNA, amplifying outward-directed forces on the capsid shell. The resulting rise in internal pressure and elastic strain drives more extensive and pronounced rupture. Notably, the average RMSD for ε = 3.5 kcal/mol is elevated even at early stages, seemingly reflecting the inherently stochastic nature of RT. Before reaching the critical size, the DNA also has greater mobility to explore different regions inside the capsid. At and beyond the critical size, the DNA strands push against the walls much more aggressively, leading to lower strand mobility but increased mechanical pressure. This behavior is also in correspondence with the experimentally observed condensing effect of the NC protein. Thus, despite the absence of explicit NC protein in our CG model, our experiment directed switching scheme implicitly mimics the effect of the NC protein on the dsDNA.

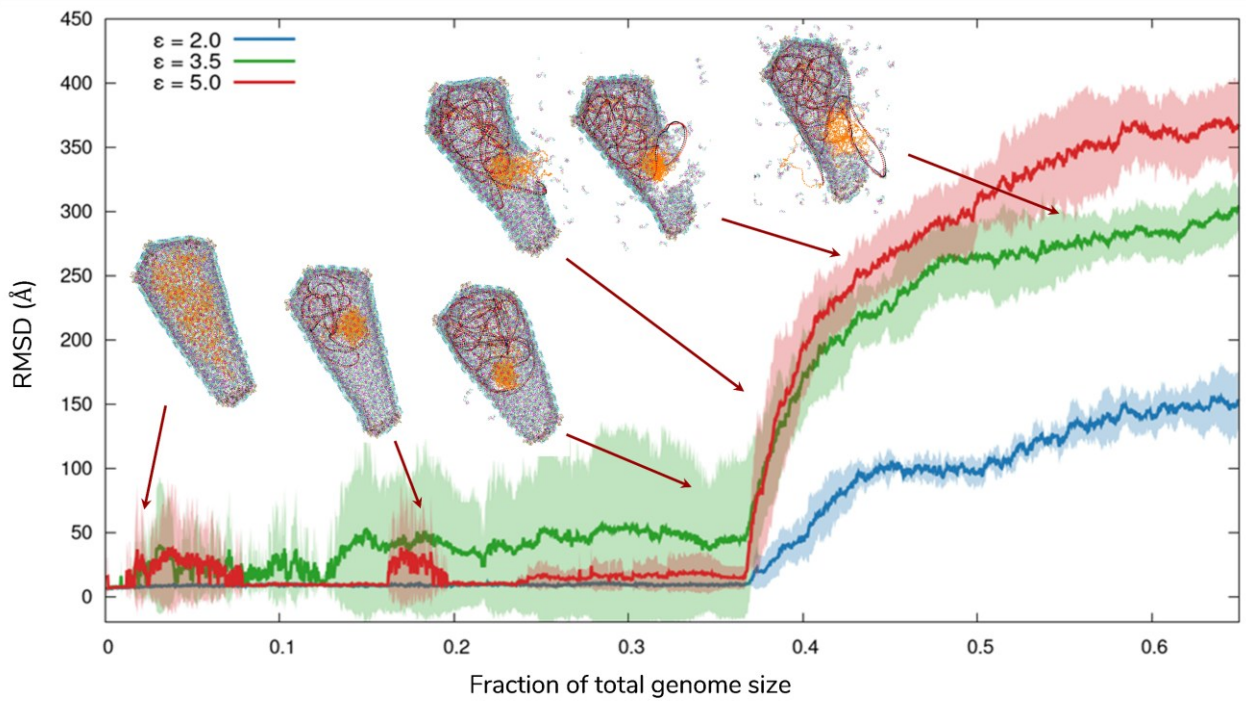

Figure 11: Root-mean-square deviation (RMSD) profiles are shown as a function of the fraction of the total genome size (χ) for ε = 2.0 (blue), 3.5 (green), and 5.0 (green) kcal/mol. An increase in RMSD indicates a departure from the reference configuration—here, the intact capsid—and thus corresponds to the onset and progression of uncoating. The shaded regions represent the standard deviation of the RMSD values, averaged over three replicas for each ε. Regions with elevated RMSD values reflect the inherently stochastic nature of RT. Representative snapshots of capsid rupturing, along with concurrent dsDNA protrusion, are shown at the corresponding χ values at which these events occur. Consistent with experimental observations, the first signs of rupturing appear at roughly one-third of the total genome length. The steepness of the RMSD slope reflects the rate of rupturing progression, which is most pronounced for ε = 5.0. As different values of ε produce distinct uncoating pathways, only one representative pathway (for ε = 3.5) is shown here.

Kinetic analysis of uncoating events relative to genome size

We also investigated the relative "rate" of rupturing events. While directly correlating uncoating events in real time with CG-MD timesteps is challenging, it is possible to relate them to genome size and changes in CA–DNA interactions. An approximate CG "timestamp" for a given configuration can be estimated as: $\chi \times 3000 \times 10^4$ CG steps, where χ is the fraction of the genome size, 3000 is the total number of CG beads in our RNA/DNA model and each KMC move is performed through short CG runs of 10^4 steps (see Methods). For each of the CA-DNA interactions (ε = 2.0, 3.5 and 5.0 kcal/mol), we qualitatively assess the relative kinetics of uncoating. However, the stochastic nature of the process makes it difficult to generalize precise uncoating timelines. For ε = 2.0 kcal/mol, as mentioned in the previous section, the uncoating is comparatively 'slower' than the other cases. Due to the experimentally motivated switch, first stages of rupturing appear at about ~ one-third the total size of the genome (**Figure 11**). Multiple pathways exhibiting cracks at different regions were predicted (**Figure 7**), including the midsection of the capsid lattice (**Figure 7A**) and near the wider and narrower ends (**Figures 7B-C**).

In all these pathways, cracks propagate slowly and lead to significant rupturing at about half of the genome size (χ = 0.45-0.6), characterized by detachment of pentamers near the wider end and/or 'opening up' of the capsid walls at the midsection. It is interesting to note that for certain cases, such as **Figure 7B**, the tip remains intact up to about half of the genome size, and then abruptly uncoats (χ = 0.42-0.53). This can be attributed to the fact that at high curvature regions such as the tip and in local regions of the wide end, the lattice is unable to experience strain, and thus ruptures quickly to release the tension. A similar pattern is observed for the pathway in **Figure 8A** for ε = 3.5 kcal/mol, where the tip ruptures abruptly from a seemingly intact one to a structure with a disassembled lower lattice over a few CG-KMC moves. Even for **Figure 8B**, once a wide patch opens in the lattice, it propagates rapidly toward the wider end of the capsid (χ = 0.38-0.5), further highlighting the role of capsid curvature, as both pentamers

and hexamers detach from the lattice. However, the propagation of cracks after the formation of the first patches is the fastest for ε = 5.0 kcal/mol, consistent with our hypothesis (**Figure 9**). For all the pathways analyzed, the initial local patches tend to form near the midsection of the lattice. From there, uncoating proceeds either through further rupture within the midsection itself or through crack propagation toward the tip, ultimately leading to pentamer and hexamer detachment. On an overall note, it can be stated that with an increase in the strength of CA-DNA interactions, the rate of progression of strain and cracks to partial disassembly of the capsid increases drastically after the first signs of uncoating. This is consistent with our RMSD results (**Figure 11**) and analysis of the interfaces (**Figure 12**).

Analysis of dimeric and trimeric interfaces during capsid uncoating

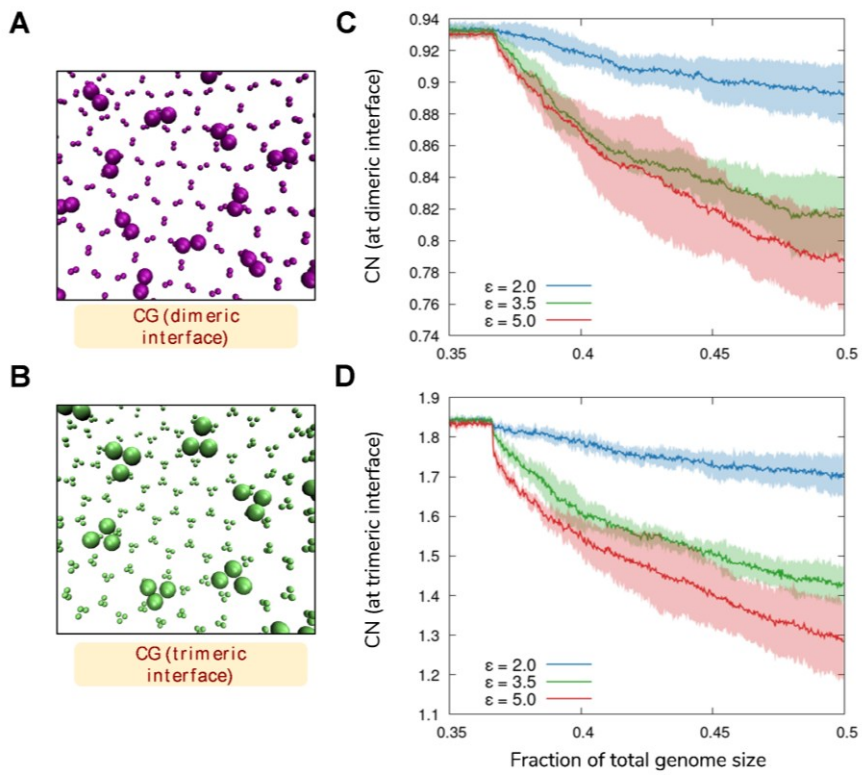

Figure 12: For each value of ε, the evolution of the dimeric and trimeric interfaces of the capsid is shown. The coordination number (CN) of a CG bead is ~1 at the dimeric interface and ~2 at the trimeric interface. An increase in the COM–COM distances at these interfaces, or a decrease in the CN, indicates capsid instability—in this case, capsid uncoating. (A) CG representation of the dimeric interface. (B) CG representation of the trimeric interface. In both cases, as RT progresses, the CN begins to decrease near the critical genome length, with the sharpest decline observed for ε = 5.0 kcal/mol and the most gradual decline for ε = 2.0 kcal/mol.

While RMSD evolution as a function of the genome size serves as an effective metric for evaluating the overall deviation from the reference structure, a more robust analysis of capsid

uncoating is to investigate how the interfaces defining the capsid lattice evolve. A recent paper on the two structural switches in the capsid (*40*) regulating capsid curvature, host factor binding and capsid entry into the NPC highlighted the structural characteristics of hexameric dimeric and trimeric interfaces influenced by host cofactors such as NUP153. At the CG resolution using a previous mapping scheme (*28*), the dimeric interface, characterized by V181, W184, M185 and other residues, corresponds to a pair of CG beads. Similarly, the trimeric interface, characterized by I201, L202, L205 and other residues, corresponds to a triplet (**Figure 12**). This corresponds to bead types 37 and 40 for the dimeric and trimeric interfaces respectively in our CG model. Thus, the central geometric 'cavity' in the dimeric/trimeric interfaces corresponds to the distance between a pair of CG beads or the area of a triangle formed by a triplet of CG beads. One can monitor the evolution of these interfaces more simply by computing the coordination number (CN) of a representative CG bead at each interface. Thus, for a stable capsid, the CN of a bead at the dimeric and trimeric interfaces is ~1 and ~2 respectively, using a 20 Å cutoff. As uncoating destabilizes the capsid, it is expected to drastically reduce the CN at each interface, depending on the degree of rupturing. **Figure 12** clearly demonstrates a reduction in the CN at each interface as RT or uncoating progresses. In agreement with the RMSD analysis, increasing the attractive parameters of the CA-DNA interactions leads to a greater degree of rupturing, as the CN demonstrates a much sharper dip at both interfaces due to uncoating. Structurally, this corresponds to a deformation of the hexameric CA-CA units that constitute the capsid lattice. This destabilizes the dimeric and trimeric interfaces between hexamers, progressively compromising lattice integrity and driving partial capsid disassembly.

Discussion and Concluding Remarks

Investigating the mechanistic and temporal aspects of capsid uncoating driven by RT is highly non-trivial, both experimentally and computationally. From a computational perspective, the primary challenge lies in scale: a mature, solvated HIV-1 capsid undergoing RT would contain > 100 million atoms, making all-atom MD simulations prohibitively expensive, especially given the required long timescales. Experimentally, the intrinsic flexibility and

structural heterogeneity of the viral RNA introduce additional complications. Its ability to adopt multiple conformational states makes high-resolution structure determination by x-ray crystallography or cryo-EM difficult (*41*), thereby limiting the construction of systematic "bottom-up" CG RNA models. Taken together, these challenges necessitate large-scale integrative CG simulations, as implemented in this work, which combine statistical mechanical frameworks with experimentally motivated modeling.

To the best of our knowledge, this work presents the first computational framework that explicitly captures HIV-1 capsid uncoating as a dynamic consequence of RT. By developing and implementing CG-KMC, we successfully simulated the kinetic growth of stiff dsDNA from flexible ssRNA inside the capsid, and, by adding an experimentally motivated switch, used to mimic the release of the condensing effect of NC, we simulated capsid uncoating. Configurations of capsid uncoating predicted by our model are in good agreement with cryo-ET images, thereby validating at least in part the integrative computational approach. It should be noted that our CA–RNA and CA–DNA interactions, as well as the experimentally motivated switch, are necessarily "top-down" or phenomenological due to the lack of high-quality genome structures. In contrast, the capsid itself is modeled bottom-up using a rigorous statistical mechanical framework, which supports our overall approach which is primarily focused on the mechanism of RT-initiated capsid rupture.

We also systematically analyzed capsid rupturing, particularly as a function of different CA-DNA interactions. These CA-DNA interactions might be probed through mutational analysis similar to Ref. (*42*). Our model predicted distinct predominant pathways corresponding to each interaction—stronger interactions led to more pronounced rupturing. It is important to note that uncoating is a stochastic process, making it remarkable that our model can predict diverse uncoating patterns, ranging from local patches at the capsid midsection to substantial disassembly at the narrow end. Our model effectively captures how dsDNA gradually grows, occupies most of the capsid interior, and, upon reaching a critical length, initiates uncoating, progressively leading to severe disassembly.

We also analyzed the relative timescale of rupture events with respect to genome size. As CG simulations eliminate fast degrees of freedom, CG timesteps cannot be directly equated to real timescales. Thus, although our CG-KMC algorithm integrates Monte Carlo steps with MD to

enable sequential conversion of a single-stranded helix to double-stranded DNA, it does not incorporate experimentally determined rate constants, such as dNTP addition to RNA, which are not very well characterized yet.

It is essential to note that uncoating is a complex phenomenon involving the disruption of CA-CA interactions across both hexamers and pentamers. These CA-CA interactions are primarily hydrophobic (*15, 43, 44*), particularly those at the CTD trimer interface, which play a critical role in capsid assembly and stability. A mature capsid also contains hydrophobic pockets with distinct functions, such as the FG-binding pocket between two monomers (described later). Electrostatic and solvation effects may further influence uncoating (*21*). Although solvent-free, our CG model incorporates the hydrophobic and electrostatic physics of a mature capsid through a bottom-up approach, effectively capturing uncoating events despite the inherent difficulty of modeling hydrophobic effects accurately by CG models(*45*).

Another important consideration is the presence of cofactors that facilitate capsid activity, such as nuclear entry and uncoating. One such factor, the IP6 anion, promotes capsid assembly and stability (*13, 46, 47*). As discussed in the Methods section, while IP6 is not explicitly included in the CG-MD simulations, it is implicitly present because the bottom-up capsid model was constructed from all-atom trajectories that included a solvated capsid patch and IP6. Moreover, the capsid contains several hydrophobic binding sites, such as the FG-binding pocket between two CA monomers (*48*). Upon nuclear entry, these FG-binding sites interact with phenylalanine-glycine-rich (FG-rich) domains of intrinsically disordered nucleoporins such NUP98 and NUP153 scattered across the NPC. While our previous study on capsid entry through the NPC included some of these domains (*28*), such components have not yet been incorporated into the RT model presented here. Other relevant cofactors include CPSF6 and Cyclophilin A (*48-50*), both of which play important roles in nuclear entry, though their effects on RT remain to be studied. We also aim to investigate the effects of the capsid inhibitor Lenacapavir (LEN) (*29, 51-53*) on RT driven capsid uncoating. Our overall aim is to construct a model capable of simulating reverse transcription concurrently with nuclear entry through the NPC in the presence of these cofactors. We will also incorporate the effect of explicit reverse transcriptase and capsid binding inhibitors (such as NVP and PF74) into our CG model in the future (*32*).

In conclusion, this work represents a new advance in modeling reverse transcription–driven capsid uncoating and contributes more broadly to our understanding of the virus capsid properties as well as the HIV-1 life cycle in general.

Methods

Integrative CG capsid and RNA modeling

A bottom-up CG capsid model, used for investigating capsid docking into the nuclear pore (*28*), was used in this work. Details of the all-atom MD simulations and the subsequent construction of the CG capsid model can be found in the SI of Ref. (*28*). In brief, a "bottom-up" CG capsid model was constructed from underlying all-atom simulations of three CA hexamers complexed with IP6 and FG peptides. FG peptides play a critical role in regulating CA-NPC interactions, and IP6 is well known for regulating capsid stability (*22, 46*) and influencing capsid uncoating (*5*). Thus, IP6 has been implicitly incorporated into our CG model. We also developed and implemented a more phenomenological (or "top-down") model of the RNA and the CA-RNA/CA-DNA interactions. Two copies of CG single-stranded RNA (ssRNA) are enclosed within the capsid, each minimally representing a ~ 9 kb genome with 3000 beads. A modified Lennard-Jones interaction with a soft core was used for the interactions between the capsid and DNA:

$$E_{\mathrm{sc-LJ}} = 4\varepsilon\lambda^{n}\left(\frac{1}{\left[\alpha_{\mathrm{LJ}}(1-\lambda)^{2}+(r/\sigma)^{6}\right]^{2}}-\frac{1}{\left[\alpha_{\mathrm{LJ}}(1-\lambda)^{2}+(r/\sigma)^{6}\right]}\right) \tag{1}$$

Modeling the CA-RNA and CA-DNA interactions is important as they facilitate capsid uncoating. The CA-DNA interactions were modeled using $E_{\text{sc-LJ}}$, and it was the corresponding ε – representing the well depth or strength of attraction – that was varied to investigate how uncoating patterns depend on the CA-DNA interactions. Before a critical threshold, Eq (1) used ε = 0.5 kcal/mol, σ = 12.5 Å and λ=0.6. After an experimentally directed "switch" is activated (described later), σ = 7.0 Å and λ =0 .6, while ε can be 2.0 or 3.5 or 5.0 kcal/mol (as analyzed extensively in the Results section). In Eq (1) n=2 and $\alpha_{\mathrm{LJ}} = 0.5$ were used throughout. dNTPs or

deoxynucleotide triphosphates were represented by a one-site CG model. $E_{sc\text{-}LJ}$ was also used for modeling CA-dNTP interactions.

Modified Gaussian potentials were used for modeling the electrostatic interactions between the two charged C-terminal CG beads of the capsid with the RNA polymer:

$$E_{\text{gauss}} = \frac{H}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(r - r_0)^2}{2\sigma^2}\right) \tag{2}$$

Where, H = - 4.5 kcal/mol, r_0 = 12.5 Å, σ = 1.2 Å for the CA-RNA interactions, and H = - 2 kcal/mol for the CA-dNTP interactions.

To prevent CG beads from overlapping with each other, soft repulsive cosine interactions were used:

$$E_{\text{rep}} = A\left[1 + \cos\left(\frac{\pi r}{r_c}\right)\right] \tag{3}$$

Where, A = 10 kcal/mol for RNA-DNA and 20 kcal/mol for DNA-DNA interactions (between CG beads of different strands), and r_c = 15 Å. Eq (3) was also used for the CA-dNTP interactions (A = 1 kcal/mol and r_c = 12.5 Å). All other interaction parameters, and the bonding/angular parameters can be found in the LAMMPS input files.

CG-KMC for modeling stages of reverse transcription

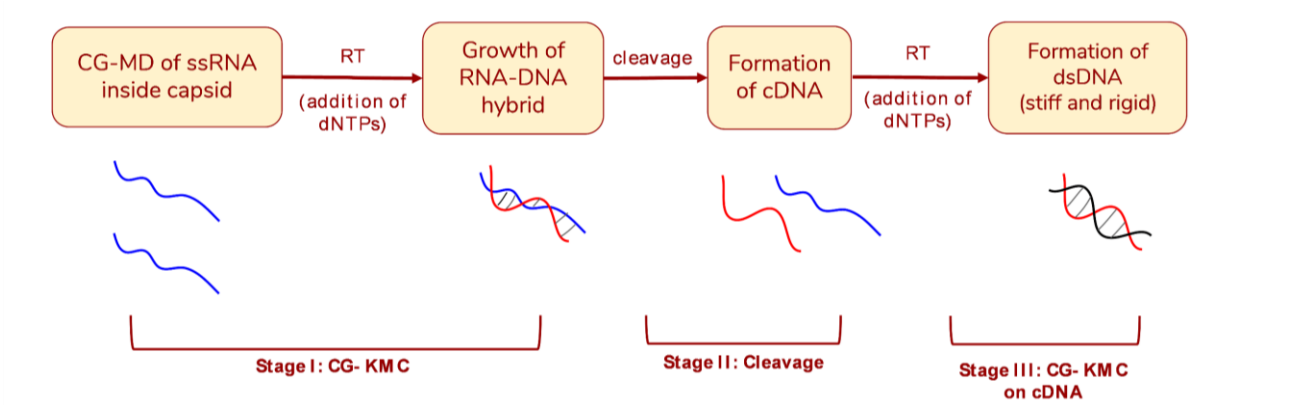

Figure 13: Stages of reverse transcription: CG-MD simulations of dNTP addition to ssRNA (implicitly catalyzed by reverse transcriptase) to form RNA-DNA hybrid. This undergoes cleavage forming cDNA,

which again undergoes RT to finally form dsDNA. This entire protocol can be represented by three stages – Stage I (CG-KMC), Stage II (Cleavage) and Stage III (CG-KMC on cDNA).

Reverse transcription is a complex process comprising multiple stages, including several strand transfer events (*3, 54, 55*). It is computationally intractable to atomistically model such an event as it (i) requires high quality structures of RNA (ii) necessitates the use of very efficient methods capable of modeling bond formation/dissociation. Classical MD with enhanced sampling tools, while being computationally efficient, uses fixed bonding topologies, thus making it impossible to capture reactivity. On another note, *ab initio* or even QM/MM (*56*) simulations are limited to a few hundred picoseconds. While approaches such as multiscale reactive molecular dynamics (MS-RMD) (*57*), and reactive methods combined with CG (*58, 59*) offer a promising approach to enhance chemical reactivity in MD, it is very difficult to efficiently parameterize and devise such methods for large-scale proteins and viruses. To this end, we developed an efficient Coarse-Grained Kinetic Monte Carlo (CG-KMC) approach to model the stages of RT (schematically depicted in **Figure 13**). Additional variants of KMC schemes have been reported in the literature as well (*60*).

From a chemical biology perspective, following the first strand transfer (FST) DNA is synthesized along an ssRNA strand. This is followed by minus strand synthesis and a primer generation, followed by the degradation of RNA and plus strand synthesis. Following second strand transfer (SST), a second DNA strand is generated, finally yielding dsDNA (*5*). As the exact events characterizing RT are very complex, we propose a three-stage approach to model RT: (1) Growth of the RNA-DNA hybrid from ssRNA, (2) Dissociation (or "cleavage") of the RNA-DNA hybrid, (*61*) (loosely corresponding to RNA degradation) to form a single-stranded complementary (cDNA) strand, and (3) Formation of dsDNA from cDNA.

The simulations start with a condensed "ball" of RNA embedded inside the capsid. Reverse transcription progresses by the addition of deoxynucleotide triphosphates (dNTPs) to the RNA. Stage I simulates a stepwise, stochastic, sequential addition of dNTPs, denoted by one CG bead, to the RNA. Thus, each CG-KMC move permits the formation of a single bond between an RNA bead and a dNTP bead, such that a bond can be formed when the RNA bead in concern is within 20 Å of a dNTP bead. The reader is directed to the SI for a mathematical formulation of CG-KMC. Subsequently, corresponding bonds and angles are formed between the dNTP beads adding

onto the RNA. This gradually leads to the formation of an RNA-DNA hybrid. It is valuable to note that this RNA-DNA hybrid naturally acquires a helical nature, which qualitatively mimics the helicity of the RNA-DNA hybrid. Following this, the RNA half of the DNA gets degraded sequentially while a new DNA strand forms complementary to the transiently formed single-stranded DNA.

Stage II of the simulation approach models the formation of a single-stranded DNA from the hybrid, in what we can call 'cleavage' of the hybrid. An RNA-DNA bond is stochastically allowed to break once the bond length is shorter than a threshold (10 Å). This generates a single-stranded DNA (ssDNA), also called the complementary DNA (cDNA). It is important to note that in our model only one RNA chain undergoes this entire process, with the second strand staying inert. Strand transfer events involving the second RNA strand (*3, 54, 55*), which is known for introducing genetic diversity via RT, have been neglected here for simplicity.

Once the cDNA is formed, a second round of CG-KMC is carried out (Stage III) to simulate dNTP addition to cDNA, finally forming the double-stranded DNA. We note that the growth of dsDNA is initiated at the two polypurine tracts (PPT): the 3′-PPT and the central PPT (cPPT) (*36, 62-64*) (**Figure S3**). These are regions which are resistant to the nucleolytic degradation activity of RT-associated RNase H. The protocol for modeling this simultaneous growth of dsDNA from the two primers is discussed in detail in the SI. However, owing to structural and kinetic complexity, the three-stranded central DNA flap (*62*) was not explicitly modeled. Higher force constants were used for representing the newly formed bonds and angles, to accurately incorporate the stiffness and rigidity of dsDNA. This rigidity plays a crucial role in the uncoating of the capsid. In Stage III as well, an experimentally-motivated "switch" is activated at ~3.5 kb of the genome size (*32*) to initiate uncoating. This mimics the ultimate failure of the condensing effect of the NC protein (discussed in the Results section) and implicitly recapitulates the conditions that drive RT. A total of five sets of simulations were carried out for Stages I and II. For each set in Stage III, three runs were performed, each corresponding to a different ε value in Eq (1).

CG-MD simulations, data analysis, and visualization

All CG-MD simulations were carried out using LAMMPS (21 July 2020) (*65*). The equations of motion were integrated with the Velocity Verlet algorithm, and the system was periodic in all three dimensions. CG-KMC was performed through an iterative series of short MD simulations, each consisting of 10^4 CG time steps, and each such iteration was carried out for each RNA/DNA bead undergoing bond formation/breaking. The CG timestep used was 50 fs (but it must be noted the CG "time" cannot be easily associated with a real time – the CG time is much longer than all-atom time). A Langevin thermostat (*66*) was used for propagating the nuclei at 310 K. The simulation box taken was 1500 Å in all three spatial dimensions. Each short CG simulation allows exactly one CG-KMC move, which may be either bond making or breaking, at 50% probability. Frames were saved after every 10^4 CG steps. The resulting trajectories and data were visualized with Visual Molecular Dynamics (VMD) (*67*). The analysis was performed using MDAnalysis (*68*), Gnuplot, and custom Python and Bash scripts. Green-Lagrange strain calculations were carried out with OVITO (*69*).

Acknowledgements

Research reported in this publication was supported by the Behavior of HIV in Viral Environments (B-HIVE) Center of the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) under award number U54AI170855. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors also acknowledge the computational resources provided by the Research Computing Center (RCC) of The University of Chicago, Bridges-2 at the Pittsburgh Supercomputing Center through the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, funded by the National Science Foundation grant numbers 2138259, 2138286, 2138307, 2137603, and 2138296, and Frontera at the Texas Advanced Computing Center (TACC), funded by the National Science Foundation (NSF grant OAC-1818253). K.G. also thanks Prof. Vinay Pathak, Dr. Arpa Hudait, Dr. Ryan Burdick and Dr. Curt Waltmann for valuable discussions.

Data Availability Statement

All relevant codes and input scripts for performing CG-KMC and CG-MD simulations in LAMMPS will be made available on GitHub upon publication. Due to the large size of the files, trajectories will be provided upon request.

References

1. E. O. Freed, HIV-1 assembly, release and maturation. *Nat Rev Microbiol* **13**, 484-496 (2015).
2. W. S. Hu, S. H. Hughes, HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* **2**, (2012).
3. J. A. Peliska, S. J. Benkovic, Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science* **258**, 1112-1118 (1992).
4. N. Arhel, Revisiting HIV-1 uncoating. *Retrovirology* **7**, 96 (2010).
5. D. E. Christensen, B. K. Ganser-Pornillos, J. S. Johnson, O. Pornillos, W. I. Sundquist, Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro. *Science* **370**, (2020).
6. R. C. Burdick, C. Li, M. Munshi, J. M. O. Rawson, K. Nagashima, W. S. Hu, V. K. Pathak, HIV-1 uncoats in the nucleus near sites of integration. *Proc Natl Acad Sci U S A* **117**, 5486-5493 (2020).
7. L. B. Gifford, G. B. Melikyan, HIV-1 Capsid Uncoating Is a Multistep Process That Proceeds through Defect Formation Followed by Disassembly of the Capsid Lattice. *ACS Nano* **18**, 2928-2947 (2024).
8. S. Rankovic, J. Varadarajan, R. Ramalho, C. Aiken, I. Rousso, Reverse Transcription Mechanically Initiates HIV-1 Capsid Disassembly. *J Virol* **91**, (2017).
9. A. E. Hulme, O. Perez, T. J. Hope, Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proc Natl Acad Sci U S A* **108**, 9975-9980 (2011).
10. Z. Ambrose, C. Aiken, HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology* **454-455**, 371-379 (2014).
11. O. Cosnefroy, P. J. Murray, K. N. Bishop, HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology* **13**, 58 (2016).
12. S. Mattei, B. Glass, W. J. Hagen, H.-G. Kräusslich, J. A. Briggs, The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354**, 1434-1437 (2016).
13. M. Gupta, A. J. Pak, G. A. Voth, Critical mechanistic features of HIV-1 viral capsid assembly. *Sci Adv* **9**, eadd7434 (2023).
14. B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, W. I. Sundquist, Assembly and analysis of conical models for the HIV-1 core. *Science* **283**, 80-83 (1999).
15. G. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, P. Zhang, Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643-646 (2013).
16. D. S. Goodsell, A. Jewett, A. J. Olson, S. Forli, Integrative modeling of the HIV-1 ribonucleoprotein complex. *PLoS Comput Biol* **15**, e1007150 (2019).
17. B. K. Ganser-Pornillos, M. Yeager, W. I. Sundquist, The structural biology of HIV assembly. *Current opinion in structural biology* **18**, 203-217 (2008).
18. S. Schifferdecker, V. Zila, T. G. Muller, V. Sakin, M. Anders-Osswein, V. Laketa, H. G. Krausslich, B. Muller, Direct Capsid Labeling of Infectious HIV-1 by Genetic Code Expansion Allows Detection of Largely Complete Nuclear Capsids and Suggests Nuclear Entry of HIV-1 Complexes via Common Routes. *mBio* **13**, e0195922 (2022).
19. J. Jin, A. J. Pak, A. E. P. Durumeric, T. D. Loose, G. A. Voth, Bottom-up Coarse-Graining: Principles and Perspectives. *J Chem Theory Comput* **18**, 5759-5791 (2022).
20. M. G. Saunders, G. A. Voth, Coarse-graining methods for computational biology. *Annu Rev Biophys* **42**, 73-93 (2013).

21. J. R. Perilla, K. Schulten, Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nat Commun* **8**, 15959 (2017).
22. A. Yu, E. M. Y. Lee, J. A. G. Briggs, B. K. Ganser-Pornillos, O. Pornillos, G. A. Voth, Strain and rupture of HIV-1 capsids during uncoating. *Proc Natl Acad Sci U S A* **119**, e2117781119 (2022).
23. J. M. A. Grime, J. F. Dama, B. K. Ganser-Pornillos, C. L. Woodward, G. J. Jensen, M. Yeager, G. A. Voth, Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nat Commun* **7**, 11568 (2016).
24. J. M. Grime, G. A. Voth, Early stages of the HIV-1 capsid protein lattice formation. *Biophysical journal* **103**, 1774-1783 (2012).
25. A. J. Pak, J. M. Grime, A. Yu, G. A. Voth, Off-pathway assembly: a broad-spectrum mechanism of action for drugs that undermine controlled HIV-1 viral capsid formation. *Journal of the American Chemical Society* **141**, 10214-10224 (2019).
26. A. Yu, K. A. Skorupka, A. J. Pak, B. K. Ganser-Pornillos, O. Pornillos, G. A. Voth, TRIM5α self-assembly and compartmentalization of the HIV-1 viral capsid. *Nature communications* **11**, 1307 (2020).
27. A. J. Pak, M. Gupta, M. Yeager, G. A. Voth, Inositol hexakisphosphate (IP6) accelerates immature HIV-1 gag protein assembly toward kinetically trapped morphologies. *Journal of the American Chemical Society* **144**, 10417-10428 (2022).
28. A. Hudait, G. A. Voth, HIV-1 capsid shape, orientation, and entropic elasticity regulate translocation into the nuclear pore complex. *Proc Natl Acad Sci U S A* **121**, e2313737121 (2024).
29. A. Hudait, R. C. Burdick, E. K. Bare, V. K. Pathak, G. A. Voth, Lenacapavir-induced Lattice Hyperstabilization is Central to HIV-1 Capsid Failure at the Nuclear Pore Complex and in the Cytoplasm. *eLife*, (2026).
30. W. M. Young, E. W. Elcock, Monte Carlo studies of vacancy migration in binary ordered alloys: I. *Proceedings of the Physical Society* **89**, 735 (1966).
31. D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* **22**, 403-434 (1976).
32. R. C. Burdick, M. Morse, I. Rouzina, M. C. Williams, W. S. Hu, V. K. Pathak, HIV-1 uncoating requires long double-stranded reverse transcription products. *Sci Adv* **10**, eadn7033 (2024).
33. S. G. Sarafianos, B. Marchand, K. Das, D. M. Himmel, M. A. Parniak, S. H. Hughes, E. Arnold, Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol* **385**, 693-713 (2009).
34. D. Warrilow, K. Warren, D. Harrich, Strand transfer and elongation of HIV-1 reverse transcription is facilitated by cell factors in vitro. *PLoS One* **5**, e13229 (2010).
35. F. Xavier Ruiz, E. Arnold, Evolving understanding of HIV-1 reverse transcriptase structure, function, inhibition, and resistance. *Curr Opin Struct Biol* **61**, 113-123 (2020).
36. S. G. Sarafianos, K. Das, C. Tantillo, A. D. Clark, Jr., J. Ding, J. M. Whitcomb, P. L. Boyer, S. H. Hughes, E. Arnold, Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J* **20**, 1449-1461 (2001).
37. R. C. Burdick, V. K. Pathak, Efficient HIV-1 in vitro reverse transcription: optimal capsid stability is required. *Signal Transduct Target Ther* **6**, 13 (2021).
38. H. Gien, M. Morse, M. J. McCauley, J. P. Kitzrow, K. Musier-Forsyth, R. J. Gorelick, I. Rouzina, M. C. Williams, HIV-1 Nucleocapsid Protein Binds Double-Stranded DNA in Multiple Modes to Regulate Compaction and Capsid Uncoating. *Viruses* **14**, (2022).
39. K. D. Hjelmstad, *Fundamentals of structural mechanics* (Prentice-Hall international series in civil engineering and engineering mechanics, Prentice Hall, Upper Saddle River, N.J., 1997), pp. xiv, 420 p.
40. J. C. V. Stacey, A. Tan, J. M. Lu, L. C. James, R. A. Dick, J. A. G. Briggs, Two structural switches in HIV-1 capsid regulate capsid curvature and host factor binding. *Proc Natl Acad Sci U S A* **120**, e2220557120 (2023).
41. L. Levintov, H. Vashisth, Structural and computational studies of HIV-1 RNA. *RNA Biol* **21**, 1-32 (2024).
42. A. E. Hulme, Z. Kelley, E. A. Okocha, T. J. Hope, Identification of capsid mutations that alter the rate of HIV-1 uncoating in infected cells. *J Virol* **89**, 643-651 (2015).

43. A. T. Gres, K. A. Kirby, V. N. KewalRamani, J. J. Tanner, O. Pornillos, S. G. Sarafianos, X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **349**, 99-103 (2015).
44. D. K. Worthylake, H. Wang, S. Yoo, W. I. Sundquist, C. P. Hill, Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution. *Acta Crystallogr D Biol Crystallogr* **55**, 85-92 (1999).
45. K. Ghosh, T. D. Loose, G. A. Voth, Can a coarse-grained water model capture the key physical features of the hydrophobic effect? *J Chem Phys* **159**, (2023).
46. R. A. Dick, K. K. Zadrozny, C. Xu, F. K. M. Schur, T. D. Lyddon, C. L. Ricana, J. M. Wagner, J. R. Perilla, B. K. Ganser-Pornillos, M. C. Johnson, O. Pornillos, V. M. Vogt, Inositol phosphates are assembly co-factors for HIV-1. *Nature* **560**, 509-512 (2018).
47. A. Yu, E. M. Y. Lee, J. Jin, G. A. Voth, Atomic-scale characterization of mature HIV-1 capsid stabilization by inositol hexakisphosphate (IP(6)). *Sci Adv* **6**, (2020).
48. J. Temple, T. N. Tripler, Q. Shen, Y. Xiong, A snapshot of HIV-1 capsid-host interactions. *Curr Res Struct Biol* **2**, 222-228 (2020).
49. Z. Ingram, C. Kline, A. K. Hughson, P. K. Singh, H. L. Fischer, R. Radhakrishnan, G. A. Sowd, N. F. B. Dos Santos, B. K. Ganser-Pornillos, S. C. Watkins, M. Kane, A. N. Engelman, Z. Ambrose, Spatiotemporal binding of cyclophilin A and CPSF6 to capsid regulates HIV-1 nuclear entry and integration. *mBio* **16**, e0016925 (2025).
50. S. Jang, G. J. Bedwell, S. P. Singh, H. J. Yu, B. Arnarson, P. K. Singh, R. Radhakrishnan, A. W. Douglas, Z. M. Ingram, C. Freniere, O. Akkermans, S. G. Sarafianos, Z. Ambrose, Y. Xiong, P. V. Anekal, P. Montero Llopis, V. N. KewalRamani, A. C. Francis, A. N. Engelman, HIV-1 usurps mixed-charge domain-dependent CPSF6 phase separation for higher-order capsid binding, nuclear entry and viral DNA integration. *Nucleic Acids Res* **52**, 11060-11082 (2024).
51. C. Li, R. C. Burdick, R. Siddiqui, S. K. Janaka, R. C. Hsia, W. S. Hu, V. K. Pathak, Lenacapavir disrupts HIV-1 core integrity while stabilizing the capsid lattice. *Proc Natl Acad Sci U S A* **122**, e2420497122 (2025).
52. S. M. Bester, D. Adu-Ampratwum, A. S. Annamalai, G. Wei, L. Briganti, B. C. Murphy, R. Haney, J. R. Fuchs, M. Kvaratskhelia, Structural and Mechanistic Bases of Viral Resistance to HIV-1 Capsid Inhibitor Lenacapavir. *mBio* **13**, e0180422 (2022).
53. M. Gupta, C. Waltmann, N. Renner, Y. Wang, L. C. James, D. A. Jacques, T. Bocking, G. A. Voth, Mechanistic insights into lenacapavir-induced off-pathway HIV-1 capsid assembly. *Proc Natl Acad Sci U S A* **123**, e2524995123 (2026).
54. V. P. Basu, M. Song, L. Gao, S. T. Rigby, M. N. Hanson, R. A. Bambara, Strand transfer events during HIV-1 reverse transcription. *Virus Res* **134**, 19-38 (2008).
55. J. Hardy, V. Mortier, Y. De Meersman, D. Staelens, E. Demecheleer, M. Schauvliege, C. Vermeersch, E. Padalko, C. Verhofstede, Aberrant strand transfer events across multiple stages of reverse transcription shape the heterogeneous landscape of the HIV-1 reservoir. *Commun Biol* **8**, 1703 (2025).
56. M. W. van der Kamp, A. J. Mulholland, Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **52**, 2708-2728 (2013).
57. C. Knight, G. E. Lindberg, G. A. Voth, Multiscale reactive molecular dynamics. *J Chem Phys* **137**, 22A525 (2012).
58. D. Teng, A. V. Mironenko, G. A. Voth, QM/CG-MM: Systematic Embedding of Quantum Mechanical Systems in a Coarse-Grained Environment with Accurate Electrostatics. *J Phys Chem A* **128**, 6061-6071 (2024).
59. K. Ghosh, D. Teng, G. A. Voth, Systematic Embedding of All-Atom Reactive Molecular Dynamics into a Coarse-Grained Environment. *J Chem Theory Comput* **21**, 8456-8467 (2025).
60. M. Faber, S. Klumpp, Kinetic Monte Carlo approach to RNA folding dynamics using structure-based models. *Phys Rev E Stat Nonlin Soft Matter Phys* **88**, 052701 (2013).
61. L. Tian, M. S. Kim, H. Li, J. Wang, W. Yang, Structure of HIV-1 reverse transcriptase cleaving RNA in an RNA/DNA hybrid. *Proc Natl Acad Sci U S A* **115**, 507-512 (2018).
62. V. Zennou, C. Serguera, C. Sarkis, P. Colin, E. Perret, J. Mallet, P. Charneau, The HIV-1 DNA flap stimulates HIV vector-mediated cell transduction in the brain. *Nat Biotechnol* **19**, 446-450 (2001).
63. L. Hameau, J. Jeusset, S. Lafosse, D. Coulaud, E. Delain, T. Unge, T. Restle, E. Le Cam, G. Mirambeau, Human immunodeficiency virus type 1 central DNA flap: dynamic terminal product of plus-

strand displacement dna synthesis catalyzed by reverse transcriptase assisted by nucleocapsid protein. *J Virol* **75**, 3301-3313 (2001).
64. C. Hu, D. T. Saenz, H. J. Fadel, W. Walker, M. Peretz, E. M. Poeschla, The HIV-1 central polypurine tract functions as a second line of defense against APOBEC3G/F. *J Virol* **84**, 11981-11993 (2010).
65. A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. I. Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, S. J. Plimpton, LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput Phys Commun* **271**, 108171 (2022).
66. A. Brünger, C. L. Brooks, M. Karplus, Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters* **105**, 495-500 (1984).
67. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J Mol Graph Model* **14**, 33-38 (1996).
68. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319-2327 (2011).
69. A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering* **18**, 015012 (2010).