The Impact of Gag Non-Cleavage Site Mutations on HIV-1 Viral Fitness from Integrative Modelling and Simulations

Firdaus Samsudin[a], Samuel Ken-En Gan[a,c,d,*], and Peter J. Bond[a,b,*]

[a]Bioinformatics Institute (A*STAR), 30 Biopolis Street, #07-01 Matrix, Singapore 138671, Singapore.

[b]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore.

[c]Antibody & Product Development Lab – Large Molecule Innovation, Experimental Drug Development Centre (A*STAR), Singapore 138670.

[d]p53 Laboratory (A*STAR), Singapore 138648

*To whom correspondence should be addressed:
Email: peterjb@bii.a-star.edu.sg
Email: samuel_gan@eddc.a-star.edu.sg

Summary

The high mutation rate in retroviruses is one of the leading causes of drug resistance. In human immunodeficiency virus type-1 (HIV-1), synergistic mutations in its protease and the protease substrate – the Group-specific antigen (Gag) polyprotein – work together to confer drug resistance against protease inhibitors and compensate the mutations affecting viral fitness. Some Gag mutations can restore Gag-protease binding, yet most Gag-protease correlated mutations occur outside of the Gag cleavage site. To investigate the molecular basis for this, we now report multiscale modelling approaches to investigate various sequentially cleaved Gag products in the context of clinically relevant mutations that occur outside of the cleavage sites, including simulations of the largest Gag proteolytic product in its viral membrane-bound state. We found that some mutations, such as G123E and H219Q, involve direct interaction with cleavage site residues to influence their local environment, while certain mutations in the matrix domain lead to the enrichment of lipids important for Gag targeting and assembly. Collectively, our results reveal why non-cleavage site mutations have far-reaching implications outside of Gag proteolysis, with important consequences for drugging Gag maturation intermediates and tackling protease inhibitor resistance.

Introduction

RNA viruses are among the most adaptable pathogens threatening human health. With their high mutation rates[1], they are orders of magnitude more adaptable than most DNA-based organisms[2], resulting in their ability not only to escape the immune system and become drug resistant, but jump across host species boundaries causing emerging infectious diseases that could turn into global pandemics[3–6]. HIV-1 represents one of the most extensively characterized RNA viruses, with mathematical modelling showing that many clinically established drug resistance mutations could occur within the first replication cycle[7], and every single possible point mutation could occur thousands of times within a single day, leading to rapid drug resistance[8], hence leading to the need to treat HIV patients with a cocktail of multiple antiviral drugs with different mechanisms[9]. Understanding the role of these mutations is thus an imperative aspect in novel antiviral drug development.

Most clinically approved HIV treatments target various key enzymes that are critical to the life cycle of the virus, one of which is the HIV-1 protease. HIV-1 protease cleaves the structural polyproteins Gag and Gag-Pol into their mature components to create the infectious virion. To inhibit this, protease inhibitors (PIs) bind competitively to the viral protease to prevent Gag substrate proteolysis and virion maturation. PIs form a vital second line component of the antiretroviral therapy (ART) in the management of adults with HIV. As failure of first line regimens develop, a potent second line drug is crucial.[10] This is often associated with multi nucleoside reverse transcriptase inhibitors (NRTIs) resistance. Ritonavir-boosted PIs represents the only available second-line treatment option for the majority of HIV patients worldwide; however, emerging mutations in HIV-1 protease often render PIs ineffective, with some mutations even leading to cross-drug resistance[11–13]. This is further aggravated by synergistic mutations on the Gag polyprotein itself[14–17], which is the rationale for elucidating the drug resistance encoded in Gag. Gag mutations around the protease cleavage site have been shown to restore Gag-protease binding by introducing new chemical interactions and inducing subtle conformational changes that compensate for the loss of affinity due to the mutations on HIV-1 protease[18]. Apart from mutations around the protease cleavage sites, Gag also harbours mutations away from these regions, many of which have been shown to directly contribute towards PI resistance[19–21]. To date, the role of these non-cleavage site mutations is largely unknown, with some analysis pointing to allosteric communication[22]. A study using

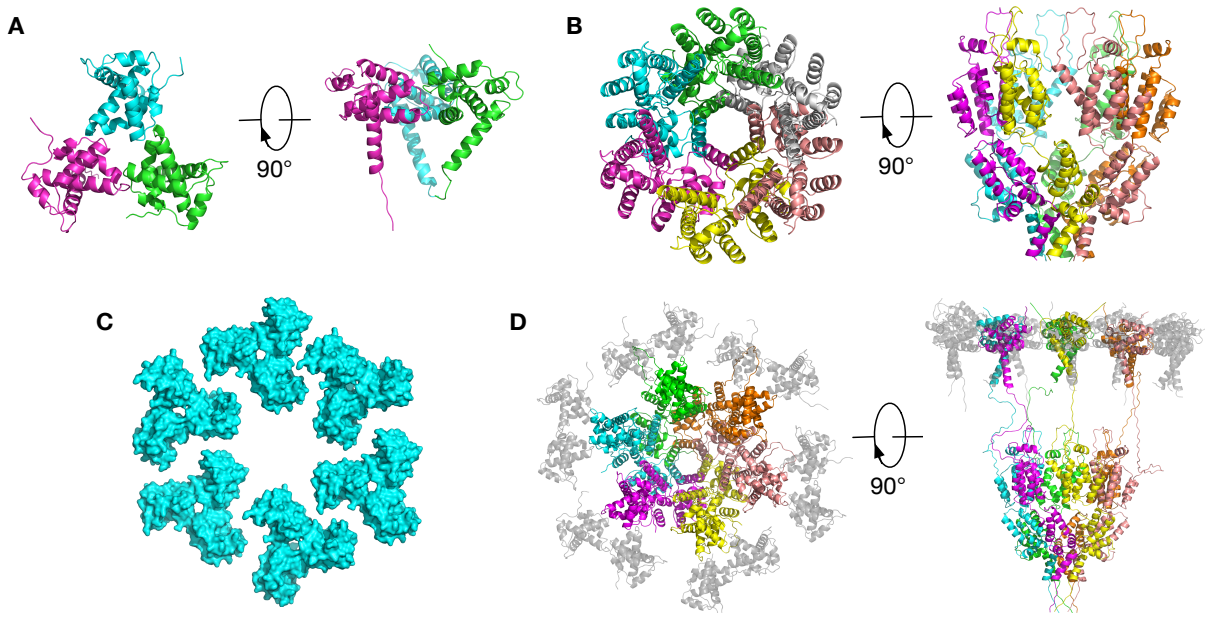
77 paramagnetic NMR spectroscopy showed that Gag can form transient complexes with
78 protease and the binding interface involves non-cleavage site regions that are prone
79 to mutation causing drug resistance[23]. In this study, we aim to use structural modelling
80 and simulations to decipher how Gag non-cleavage site mutations contribute towards
81 the overall fitness of the HIV-1 virus.

82 The HIV-1 Gag polyprotein is a 500 amino acid precursor protein containing the
83 matrix (MA), capsid (CA), nucleocapsid (NC) and P6 domains, as well as two spacer
84 peptides, SP1 and SP2. During the late phase of the HIV-1 replication cycle, Gag is
85 sequentially cleaved by the protease enzyme into these domains (Figure S1), which
86 subsequently form an infectious mature virion[24]. Each domain plays a specific role
87 during maturation; for example, the MA domain drives full-length Gag to assemble at
88 the plasma membrane of the host cell, while the CA and NC domains encapsulate the
89 viral RNA genome. The CA domain also interacts with CypA, which is a host cell
90 protein that is incorporated into the virion and is essential for capsid uncoating[25,26].
91 Due to its functional significance in the HIV-1 life cycle, Gag is an attractive target for
92 therapeutic agents and to date, several drugs that target the CA domain have been
93 identified[27]. Nevertheless, a Gag inhibitor is yet to be clinically approved. For example,
94 Bevirimat is a drug candidate that stabilises the immature Gag lattice by preventing
95 proteolysis between the CA and SP1 domains[28]. A clinical trial, however, showed a
96 reduced response amongst certain patients due to the highly polymorphic nature of
97 Gag across different circulating HIV-1 subtypes and various resistant mutations[29,30],
98 which highlights the importance of understanding the role of Gag non-cleavage site
99 mutations.

100 Over the past few years, integrative modelling and molecular dynamics (MD)
101 simulations, in tandem with advances in structural biology, have provided valuable
102 insights into the molecular mechanism underlying viral function and dynamics[31]. In the
103 study of HIV-1, this ranges from multiscale simulations of the entire viral capsid shell
104 and its assembly pathways[32–37], to simulations of individual Gag proteins with host cell
105 components, such as the CA domain with CypA[38], kinesin-1 adaptor protein FEZ1[39]
106 and inositol hexakisphosphate (IP6)[40], as well as the MA domain with a model plasma
107 membrane[41]. A previous structural model of full-length monomeric HIV-1 Gag revealed
108 allosteric communications between non-cleavage site mutations and the first Gag
109 cleavage site[22], providing a glimpse into how residues far away from the protease
110 cleavage site could affect proteolysis. However, this model did not take into account

111 *in vivo* Gag oligomerisation, which is instrumental during virus particle maturation.
112 Crystal structures show that the MA domain exists as a trimer[42], whilst the CA domain
113 forms a hexamer[43]. An electron microscopy (EM) study of the MA protein in a PIP2
114 containing membrane showed that under higher order conditions, it organizes into
115 hexamers-of-trimers[44].

116 Structural models of sequentially cleaved Gag are imperative for understanding
117 the conformational changes involved upon proteolysis, which may improve our overall
118 knowledge of Gag mutations and potential drugging of these intermediates. Using
119 available structures of HIV-1 Gag domains, we built an integrative model of the
120 complete oligomeric Gag polyprotein cleavage product (Figure 1) bound to a viral
121 membrane model (Figure 2). Based on coarse grained (CG) MD simulations using the
122 Martini forcefield of the wild-type (WT) and mutant Gag variants, supported by careful
123 calibration against atomic-resolution sampling, we found that non-cleavage site
124 mutations can interact with cleavage site residues and potentially alter their local
125 environment, while mutations on the MA domain confer stronger binding to the plasma
126 membrane. Subsequent all-atom MD simulations of the CA domain revealed how
127 mutations in this region modulate CypA interaction as well as stabilise oligomer
128 formation. Overall, our study uncovers how these distant mutations can affect various
129 processes during HIV-1 viral maturation and contribute towards its overall fitness.
130



131
132
133 **Figure 1: Modelling HIV-1 Gag MA-CA-SP1.** (A) The crystal structure of the MA trimer (PDB: 1HIW).
134 (B) The crystal structure of the immature CA hexamer (PDB: 5L93). (C) The hexamer of trimers model

135 for the MA domains as supported by EM. (D) A model of the MA-CA-SP1 hexamer. The central subunits
136 of the MA trimers are connected to the CA domains and coloured as in (B), whereas the peripheral MA
137 subunits are coloured grey.
138
139 **Results**
140 ***Membrane dynamics of the HIV-1 Gag MA-CA-SP1 polyprotein***
141 The Gag polyprotein is sequentially cleaved by HIV-1 protease during viral maturation
142 (Figure S1). It is unknown which state of viral maturation is affected by non-cleavage
143 site mutations. As such, we aim to investigate the effects of these mutations on both
144 immature and mature states of Gag, where structures are available. The first cleavage
145 occurs between the SP1 and NC domains to produce the MA-CA-SP1 polyprotein,
146 which represents the largest Gag cleavage product. A hexameric model of the HIV-1
147 Gag MA-CA-SP1 polyprotein was built using the crystal structures of the MA trimer[42]
148 and CA hexamer[43] (see Methods section for further details) (Figure 1). The crystal
149 structure of the immature CA hexamer was aligned with six copies of MA trimers
150 (hexamers-of-trimers) following the arrangement observed in membrane bilayers[44]. A
151 solution NMR structure of an N-terminal fragment of immature HIV-1 Gag that includes
152 MA and the C-terminal portion of CA shows that the linker between MA and CA
153 domains is inherently flexible and adopts multiple conformation[45]. This is supported by
154 another NMR study of an HIV-1 Gag fragment consisting of the MA, CA, SP1 and NC
155 domains that showed individual domains orient in a manner semi-independent of each
156 other, with adjoining linkers that are intrinsically disordered[46]. As such, we constructed
157 a flexible intervening linker between the MA and CA domains *ab initio*. Gatanaga et
158 al. discovered PI-resistant Gag variants with non-cleavage site mutations mapped to
159 the MA-CA-SP1 complex (Table S1, Figure S1)[20]. These include seven unique
160 mutations, with five located in the MA domain and two in the CA domain. We
161 generated models for three variants (specifically variant numbers 3, 6, and 7, as
162 outlined in Table S1) that cover all of these seven unique mutations (Figure 2A). These
163 models were built by substituting the wild-type residues with the corresponding mutant
164 version, assuming that these point mutations do not have any large structural impact
165 upon Gag. To understand the conformational dynamics of the polyprotein and the
166 potential roles of these non-cleavage site mutations, we performed four independent
167 15 μs CG MD simulations of the MA-CA-SP1 model bound to a realistic HIV-1
168 membrane model (Simulations 1-4 in Table S2). The membrane model contained

169 phosphatidylcholine (PC), phosphatidyl ethanolamine (PE) and phosphatidylserine
170 (PS) lipids, as well as PIP2, sphingomyelin and cholesterol, as determined by a
171 previous lipidomics study[47] (see Methods section for further details), while the MA
172 domain N-terminus was myristoylated (Figure 2B).

173 As previously predicted[41], the myristate group on the N-terminal glycine residue
174 of MA domain effectively anchored the protein to the membrane. Within the first few
175 nanoseconds, the linker connecting the MA and CA domains contracted, effectively
176 pulling the CA-SP1 domain towards the membrane (Figure 2B). To verify the MA-CA
177 linker contraction observed during CG sampling, we also performed four independent
178 500 ns atomistic simulations of the MA-CA-SP1 monomer with positional restraints
179 applied to the backbone atoms of lipid binding residues to mimic membrane
180 association (Simulation 5 in Table S2). Additionally, we conducted CG simulations of
181 the same monomeric system (Simulation 6 in Table S2). Both of these additional sets
182 of simulations showed a contraction of the MA-CA linker, resulting in a similar
183 distribution of distances between the MA and CA domains compared to the CG
184 hexamer simulations (Figure S2). While the atomistic simulations initially sampled a
185 wider spread of distances, after around 300 ns this converged to 4-6 nm, reproducing
186 the CG distributions. This validates our CG hexamer model and shows that the linker
187 region does not maintain an extended conformation, but rather contracts to allow direct
188 contact between the MA and CA domains.

189 Curvature has been shown to be important in modulating the interface of the
190 CA lattice[43,48]. Structural flexibility of the CA N-terminal and C-terminal domains give
191 rise to a curved CA-CA interface, and oligomerisation of this curved interface at the
192 plasma membrane of infected cells results in membrane bending and subsequent
193 budding of the virus particles. The structure of the CA hexamer used in our modelling
194 was taken from the intrinsically curved CA lattice within immature virus-like particles[43].
195 However, the MA domains were arranged to be flat to match the flat model membrane
196 at the beginning of the simulations (Figure 1D). Interestingly, over the course of the
197 simulations, as the MA-CA linker contracted allowing direct contacts between these
198 two domains, the MA domain spontaneously adapted to the curved CA hexamer
199 (Figure 2B and S3). Consequently, the curved MA domain induced some degree of
200 curvature in the membrane. This shows how a curved CA-CA interface could affect
201 surrounding Gag domains and the membrane, thereby affecting viral release.

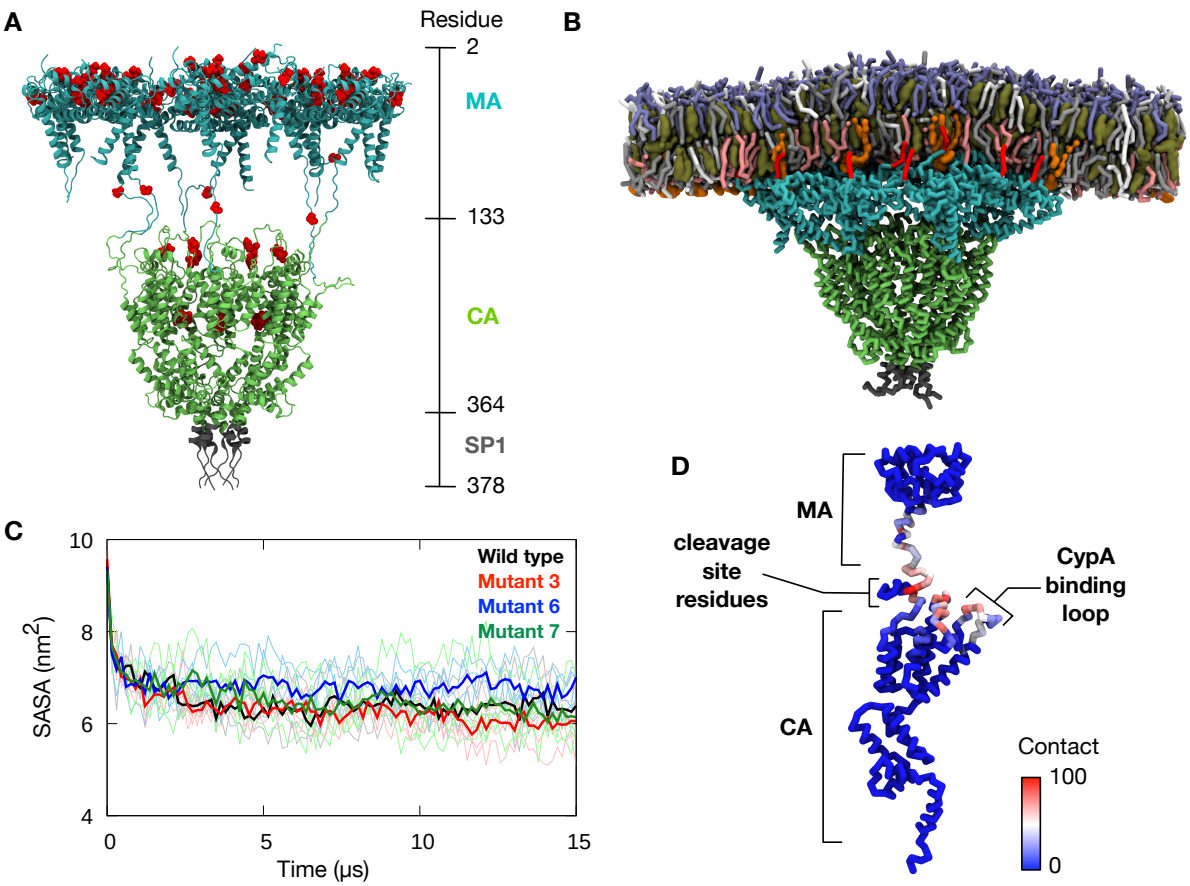


Figure 2: HIV-1 Gag MA-CA-SP1 model. (A) An atomic model of MA-CA-SP1 hexamer built using the crystal structures of MA trimer[42] and immature CA hexamer[43], shown in cartoon representation. The MA domain (cyan), CA domain (green), and SP1 domain (dark grey) are shown in different colours. The positions of seven unique non-cleavage site mutations[20] are shown in red. Residue numbering for each domain and their approximate positions are shown on the right. (B) The final snapshot from one of the 15 μs CG simulations of a membrane-bound WT MA-CA-SP1 hexamer showing contraction of the linker connecting MA and CA domains. Protein backbone and lipids are shown in licorice representation. N-terminal myristate (red), PC (white), PE (light grey), PS (pink), PIP2 (orange), sphingomyelin (purple) and cholesterol (dark green) are shown in different colours. (C) Solvent accessible surface area (SASA) of the MA-CA cleavage site region (residue 128-137) over the course of 15 μs CG simulations. Thin lines represent the six individual subunits and thick lines show the running average. Data were averaged over four independent simulations. (D) Average percentage of contact made by the cleavage site residues with the rest of the protein, mapped onto the backbone of a single subunit of MA-CA-SP1 from the WT simulations. Data were averaged over six subunits and four independent simulations. The cut-off distance used for contact analysis was 0.6 nm.

Interaction between cleavage site residues and non-cleavage site mutations

The linker contraction described above resulted in a decrease in solvent accessibility of the cleavage site residues between the MA and CA domains from ~10 nm^2 to ~6

222 nm^2 (Figure 2C and S4). In the first round of proteolysis, Gag is cleaved at the SP1/NC
223 cleavage site to produce MA-CA-SP1 and NC-SP2-P6 intermediates (Figure S1). In
224 the second round of proteolysis, P6 is cleaved off the latter at a rate measured in vitro
225 approximately 9-fold slower than the initial cleavage, while MA is cleaved off the former
226 at an estimated 14-fold slower rate than the first round of proteolysis[49]. Cleavage at
227 the MA/CA site therefore happens around 1.6 times more slowly than at the SP2/P6
228 site. Our simulations suggest that this may be caused by the reduction in exposure of
229 the MA/CA cleavage site, whereby at the end of the simulations, the solvent accessible
230 surface area (SASA) of the cleavage site is also approximately 1.6 times smaller than
231 at the beginning of the simulations, due to the contraction of the linker region between
232 MA and CA domains. This suggests that the MA-CA linker dynamics could account for
233 the difference in the rates observed in Gag cleavage, and that the SP2-P6 linker
234 adopts a more extended and less buried conformation.

235 This linker contraction also led to the cleavage site residues interacting with
236 neighbouring subdomains, predominantly with the C-terminal helix of the MA domain,
237 N-terminal loop of the CA domain, and several inter-helical loops on the CA domain,
238 including the CypA binding loop (Figure 2D). Two PI-resistant mutations were
239 identified at the interaction sites – G123E and H219Q. To understand how these
240 mutations may affect the cleavage site residues, contact analysis with a cut-off
241 distance of 0.6 nm was performed between the cleavage site residues and these two
242 specific positions in the WT and mutant proteins, particularly focusing on potential
243 inter- and intra-subunit interactions. Due to their close proximity, we found that the
244 residue at position 123 made contact primarily with the N-terminal portion of the
245 cleavage site from the same Gag subunit (Figure 3A and 3B). Both WT (glycine) and
246 mutant (glutamate) residues showed a similar percentage of contact over the course
247 of the simulations. “Back-mapping” of CG simulation snapshots to transform them into
248 atomic resolution, performed using a flexible geometric approach[50] (details in Methods
249 section), suggested that a glutamate residue at this position may interact with polar
250 residues on the cleavage site, such as Q130 and Y132 (Figure 3C). We performed
251 three independent 200 ns atomistic simulations of a single MA-CA-SP1 subunit to
252 refine the back-mapped structure, and found that the glutamate residue indeed
253 interacted primarily with Y132 (Figure S5). Additionally, the residue also contacted
254 residues N131 and Q130. In contrast, a glycine residue at this position interacted
255 primarily with V128 and S129. Given the change in the overall size and charge of the

256 residue in the WT and mutants (from small and neutral to large and acidic), the G123E
257 mutation alters the accessibility and electrostatic properties in the vicinity of the
258 cleavage site and would therefore be expected to directly interfere with proteolysis.
259 While the interactions of G123E with the cleavage site residues occur within
260 the same subunit of the MA-CA-SP1 polyprotein, inter-subunit interactions were
261 observed for H219, whereby this residue contacted the cleavage site of an adjacent
262 subunit (Figure 3D). The frequency of such interactions, however, was notably lower
263 than that displayed by G123E, and interestingly, the mutation to glutamine resulted in
264 a reduced contact with the cleavage site (Figure 3E). Back-mapping to atomistic
265 representation suggested potential for interactions with polar residues such as N131
266 and N137 (Figure 3F). Atomistic simulations of two adjacent MA-CA-SP1 subunits
267 using the back-mapped structure showed that histidine at this position made
268 intermittent contacts with I134, V135, Q136 and N137, whilst glutamine interacted
269 primarily with Y132 (Figure S5). Given that the WT histidine may potentially become
270 protonated, it may affect local electrostatic surface properties of the Gag cleavage site
271 to modulate the catalytic efficiency of the protease enzyme.
272 Overall, our data suggest that even though these mutations occur far from the
273 protease binding site, they can physically interact with the cleavage site residues,
274 either within the same Gag subunit or with a neighbouring subunit, as the linker region
275 between MA and CA domain contracts following the first proteolytic cleavage.
276

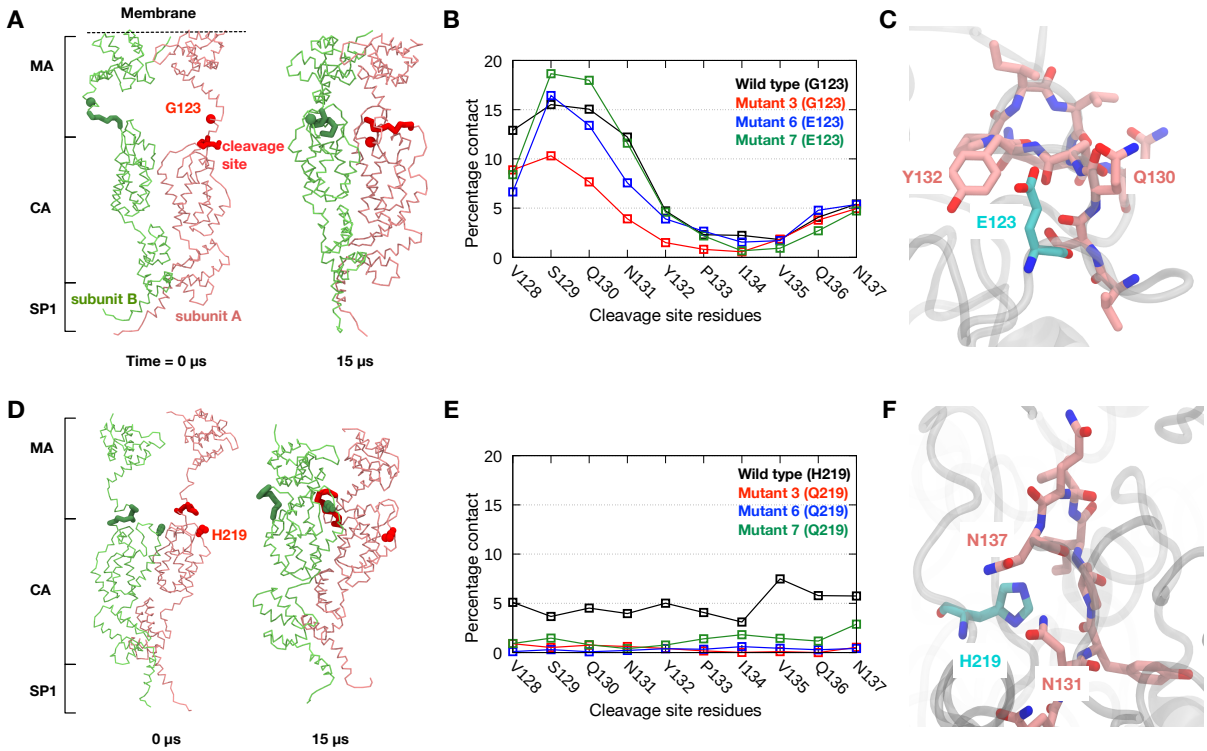


Figure 3: Interactions between non-cleavage site mutations and cleavage site residues. (A) Snapshot at the beginning and end of the CG simulations of WT MA-CA-SP1 showing the interaction of residue G123 (sphere representation) and cleavage site residues (thick licorice representation). Two subunits are shown in red and green. (B) Percentage of contacts across the entire CG simulation sampling made between G123/E123 with each of the residues of the MA-CA cleavage site. This is averaged over the six subunits of MA-CA-SP1 and the four independent trajectories. The cut-off distance used for contact analysis was 0.6 nm. (C) Atomistic model, derived from the final snapshot of one of the CG simulations, highlighting E123 in mutant Gag (cyan) and nearby cleavage site residues (pink). (D) Snapshot at the beginning and end of the CG simulations showing the interaction between H219 from one subunit to cleavage site residues on an adjacent subunit. (E) Contact analysis similar to (B) for H219/Q219. (F) Atomistic model derived from the CG simulations of WT Gag highlighting H219 in WT Gag (cyan) and surrounding cleavage site residues (pink).

Mutations in the MA domain modulate membrane binding

Four of the mutations in the MA domain map onto the plasma membrane associating surface of our model, and three mutations – E12K, E40K and L75R – involve a switch to basic amino acids. We hypothesized that these mutations may therefore alter key electrostatic interactions with lipids. Indeed, in the three mutant variants, there is a notable increase in positive electrostatic charge on the membrane binding region compared to WT (Figure S6). In the WT Gag MA domain, the positive charge is concentrated on the peripheral region of the membrane binding surface of the MA trimer. In the mutant variants, in particular mutants 6 and 7, the positive charge covers

300 almost the entire membrane binding surface including the central region at the
301 interface of the MA trimer subunits.
302 Previous biochemical and structural studies showed that PIP2 is instrumental
303 in targeting Gag to the plasma membrane of the host cell, and that it binds directly to
304 the MA domain[51,52]. To understand how the changes in electrostatic surface properties
305 caused by the PI-resistant mutations affect membrane binding, we determined the
306 percentage of PIP2 interacting with MA throughout our CG simulations. This is based
307 on the number of PIP2 molecules found within 0.6 nm of MA normalized by the total
308 number of PIP2 molecules within the membrane. Over the course of the 15 μs
309 simulations, there was a significant enrichment of PIP2 lipids around the MA domain,
310 consistently observed across all trajectories, from around 30% at the beginning of the
311 simulations to 60-80% by the end (Figure 4A). A number-density map averaged over
312 the last 5 μs of the simulations showed a significantly higher density of PIP2 around
313 the edges of MA hexamer-of-trimers as well as at the interface between MA trimers
314 compared to areas further away from the protein (Figure 4B and S7A). This was due
315 to electrostatic attraction between the anionic PIP2 headgroups and the positively
316 charged membrane-peripheral surface of the MA domain. Interestingly, we did not
317 observe any increment in the percentage of anionic PS lipids despite the higher
318 concentration of PS compared to PIP2 in the membrane (Figure S7B and S7C). This
319 may be due to the higher negative charge on the headgroups of PIP2 lipids. PIP2
320 enrichment around the MA domain is in agreement with previous experimental studies
321 showing that PIP2 is responsible for anchoring HIV-1 Gag to the plasma membrane[53].
322 While both WT and mutant variants displayed PIP2 enrichment, all three mutants
323 showed a larger degree of accumulation of PIP2 around MA. More importantly, the
324 E40K and L75R mutations, which map around the interface of the MA trimer subunits,
325 formed novel binding sites for PIP2 lipids (Figure 4C, 4D). Our simulations therefore
326 suggest that these non-cleavage site mutations enhance interactions between the MA
327 domain and PIP2 lipids, and may consequently improve membrane targeting of HIV-
328 1 Gag during viral assembly.
329

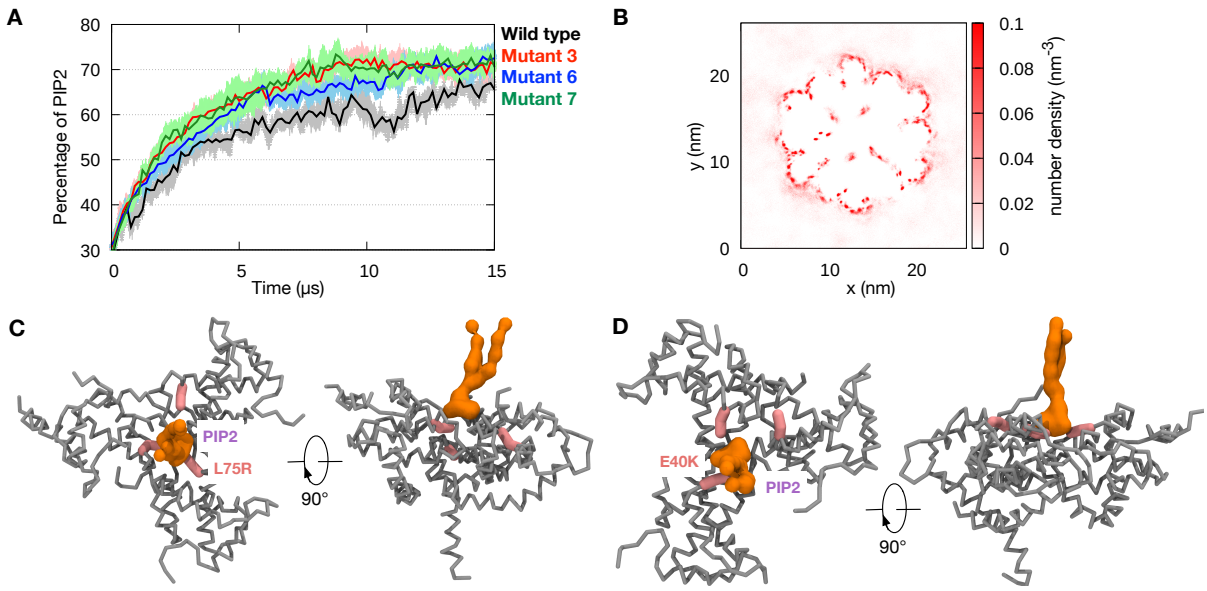


Figure 4: PIP2 is enriched around MA domain. (A) The percentage of PIP2 found within 0.6 nm of the MA domain during CG simulations of WT and mutant MA-CA-SP1 models. Error bars indicate standard deviation from four independent simulations. (B) Two-dimensional number density map of PIP2 for CG simulations of mutant 3 (Simulation 2 in Table S2) averaged over the last 5 μs. (C) Snapshots from one of the simulations of mutant 3 showing PIP2 binding at the interface of the MA trimer. The position of mutant residue, L75R, is shown in pink stick representation, whereas PIP2 is shown in orange surface representation. The rest of the MA domain is shown in grey. (D) Snapshot from simulations of mutant 7 (Simulation 4 in Table S2) showing PIP2 binding facilitated by E40K mutation, displayed in similar colours and representation as in (C).

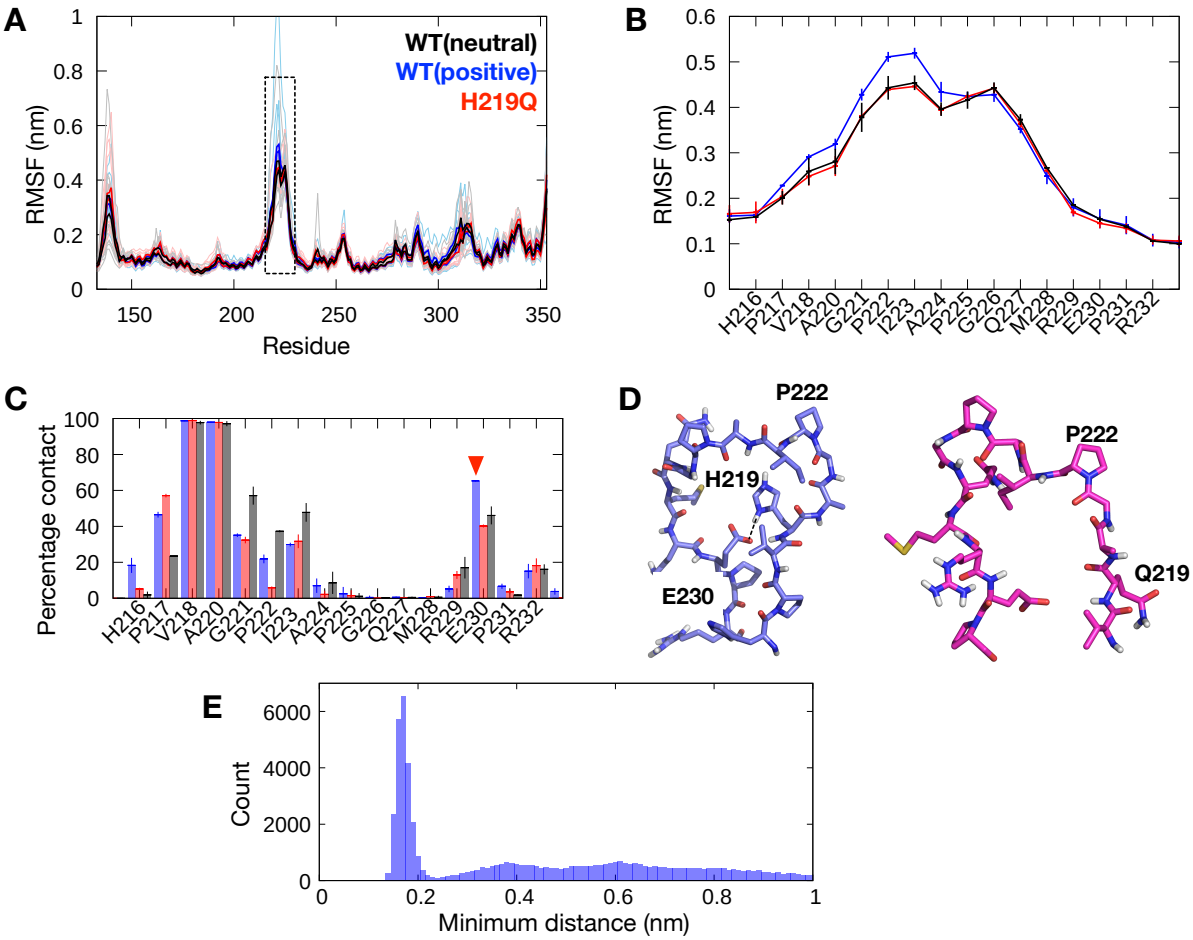
Mutations in CypA-binding loop optimize CypA interaction

During viral maturation, CA molecules undergo a large conformational change to form a conical capsid encapsulating the viral genome. We next sought to elucidate the impact of non-cleavage site mutations on the mature CA using the crystal structure of the mature CA hexamer[40]. We first investigated the interactions between mature CA and Cyclophilin A (CypA). CypA is a host protein that is recruited by HIV-1 Gag into mature virion particles. While the exact role of CypA in HIV-1 replication remains unclear, it is thought that the prolyl isomerase activity destabilises interactions between the subunits of the CA hexamer and therefore promotes viral core disassembly during the early stage of the replication cycle[26,54,55]. An optimal population of CypA in the infected target cell has been shown to be critical in modulating HIV-1 infectivity[56]. At a very low CypA concentration, the HIV-1 replication rate is severely reduced as the CA subunits are tightly bound to one another, which hinders virion uncoating. On the other hand, at a very high CypA concentration, interactions between

355 CA subunits are greatly destabilised, resulting in an unstable virion core and thereby
356 delayed virion maturation[57]. CypA binds to CA via an exposed proline-rich
357 unstructured loop region in the N-terminal domain of the latter[55]. Interestingly, this loop
358 houses H219Q, a mutation found in all seven variants from Gatanaga et al. (Table
359 S1). To investigate the potential role of this mutation *vis-à-vis* CA-CypA interactions,
360 we performed two independent 500 ns atomistic MD simulations of the WT and H219Q
361 mutant of the mature CA hexamer either in its CypA-free (apo) or CypA-bound state
362 (Simulations 7-12 in Table S2). The predicted pKa values of H219 in all six chains of
363 the hexamer were determined using the PROPKA3 programme[58] and were found to
364 range between 6.3-6.5 (Table S3). As these pKa values are close to physiological pH,
365 it is possible that H219 may exist in either protonated or unprotonated form. As such,
366 the simulations of WT CA were performed with H219 in both protonation states. IP6
367 has been shown to be an important co-factor for HIV-1 capsid maturation and
368 assembly[40,59–61]. In the immature Gag hexamer, IP6 interacts with K290 and K359,
369 thus facilitating the formation of a six-helix bundle found between the CA and SP1
370 domains. Proteolytic cleavage between these two domains releases IP6 to
371 subsequently bind to residue R18 in the mature CA, promoting assembly of mature
372 CA hexamers and the lattice. As such, we included the IP6 molecule in all of our
373 atomistic simulations of the mature CA hexamer (see Methods section for further
374 details).

375 Our apo simulations revealed that the CypA binding loop was the most dynamic
376 region of the entire CA molecules (Figure 5A). This is consistent with a recent structure
377 of a CA tubular assembly determined by magic-angle spinning NMR by Manman et al.
378 showing that the loop adopts at least four distinct conformations[62]. No noticeable
379 difference was found between the flexibility of this loop in simulations with the neutral
380 H219 *versus* the H219Q mutant (Figure 5B). Interestingly, when this histidine residue
381 was protonated, we saw a slight increase in the root mean square fluctuation (RMSF)
382 values of the loop, especially around G221-P222, which is the putative substrate for
383 the CypA rotamase activity[63,64]. Cluster analysis performed on structures extracted
384 from the simulations revealed that the CypA binding loop containing the protonated
385 H219 was the most heterogenous with the largest distribution of root mean square
386 deviation (RMSD) values, the highest number of distinct clusters, and the fewest
387 structures grouped into the top three clusters (Figure S8). An increased association
388 was observed between protonated H219 and E230, found on the opposite side of the

389 loop (Figure 5C). Similarly, the central structure of the top cluster from protonated
390 H219 simulations showed formation of a salt bridge between the protonated histidine
391 and E230 (Figure 5D), which was absent from the top three clusters extracted from
392 simulations with neutral H219 and H219Q mutant, as well as the ensemble of wild type
393 CA structures solved by Manman et al.[62] (Figure S8). However, this salt bridge was
394 intermittent (Figure 5E), such that the CypA-binding loop may adopt alternative
395 conformational states when H219 is protonated. This means that under low pH
396 conditions, the transition between these two states increases the overall flexibility of
397 the loop, which consequently may make it more difficult for CypA to bind. Mutation of
398 H219 to glutamine would therefore abolish the ability of the loop to adopt these two
399 conformational states, reducing the loop flexibility, and facilitate initial CypA binding.
400 This would be advantageous inside the endosome which is associated with a lower
401 local pH, to ensure sufficient CypA binding for core disassembly[65].
402



403
404 **Figure 5: The dynamics of CypA-binding loop.** (A) Per-residue root mean square fluctuations
405 (RMSF) of the CA from two independent 500 ns apo simulations. Thin lines show the RMSF of each

individual CA subunit, whilst the thick lines show the average from each simulation. Simulation with neutral H219 is shown in black, positively charged H219 in blue, and mutant H219Q in red. Dotted box indicates the loop region where CypA binds. (B) Enlarged RMSF plot for CypA-binding loop. Average from the two simulations is shown and error bars indicate standard deviations between repeat simulations. (C) Atomic contact analysis performed between residues at position 219 and the rest of the CypA-binding loop, including E230 (indicated by red arrowhead). Average values from two simulations are shown and error bars indicate standard deviations between repeat simulations. Cut-off distance for contact analysis is 0.4 nm. (D) Representative structures of the CypA-binding loop from the simulations with positively charged H219 (left) and mutant H219Q (right) calculated using cluster analysis. The central structures (the structure with the lowest RMSD from all other structures) of the top clusters are shown (more details in Figure S8). In the former, H219 can form a salt bridge interaction with residue E230 found on the opposite side of the loop, represented as dotted line. (E) The distribution of minimum distance between the hydrogen atoms bonded to the two nitrogen atoms on the side chain of the protonated H219 and the two oxygen atoms on the side chain of E230. Data taken from all six CA subunits and both repeat simulations.

Since the interactions between H219 and CypA form a part of the binding interface[55], this raises the question of whether a mutation to glutamine would undermine the strength of binding. To investigate the potential effects of the H219Q mutation upon CypA binding affinity, we performed two independent 500 ns atomistic MD simulations of a mature CA hexamer with each of the subunits bound to a CypA molecule (Figure 6A). While the average buried surface area between CA and CypA in the WT and mutant CA systems were comparable throughout the simulations, we observed transient detachment of CypA molecules in two of the CA subunits from the H219Q variant (Figure 6B). Based on the crystal structure[55], H219 forms hydrogen bonds with residues in the CypA active site, including N71. Our simulations revealed that this histidine residue also interacts with other polar residues such as T73, N102, and Q111 on CypA. In the H219Q mutant, the contacts made by the glutamine residue followed a largely similar pattern, although the frequency of interactions was noticeably lower compared to both neutral and protonated histidine (Figure 6C). Both of these analyses of CypA-bound CA simulations suggest that the H219Q mutation may potentially weaken CypA binding to the CA domain. This mutation is therefore likely beneficial in a CypA-rich environment whereby a high population of CypA in target cells may destabilise the viral core. Taken together, the H219Q mutation on the CypA-binding loop may play a crucial role in fine-tuning CypA interactions, especially in low endosomal pH and CypA-enriched infected target cells.

442 It is worth noting that cryo-EM studies of the large HIV-1 capsid lattice showed
443 that a single CypA molecule could simultaneously bind two CA hexamers (Figure
444 S9)[38,66]. To one hexamer, CypA binds to one CA subunit via canonical interactions
445 with the CypA-binding loop discussed in this study. To the other neighbouring
446 hexamer, CypA binds to two adjacent CA subunits via non-canonical interactions,
447 facilitated by residues 254-257 as well as 227-229. We compared the RMSF values
448 of these residues between our three sets of apo simulations involving neutral H219,
449 protonated H219 and mutated H219Q variants (Simulations 7-9 in Table S2). There
450 was no significant difference between the RMSF values for these residues, suggesting
451 that the protonation and mutation of H219 would not affect the dynamics of residues
452 involved in non-canonical binding of CypA. Interestingly, the cryo-EM structures show
453 that these two non-canonical binding interfaces have a much smaller interaction area
454 compared to the canonical interface, suggesting a weaker interaction. This is
455 supported by the lower pairwise interaction energies of the non-canonical interface
456 from previous MD simulations[38]. Collectively, these non-canonical CypA-CA interfaces
457 act as a secondary binding site, once CypA binds to the canonical CypA-binding loop.
458 While we focused on the primary CypA binding site, since mutation and protonation of
459 residue H219 does not affect the dynamics of the residues at the non-canonical
460 interfaces, it is conceivable that CypA binding at these secondary sites may not be
461 regulated by the H219Q mutation.
462

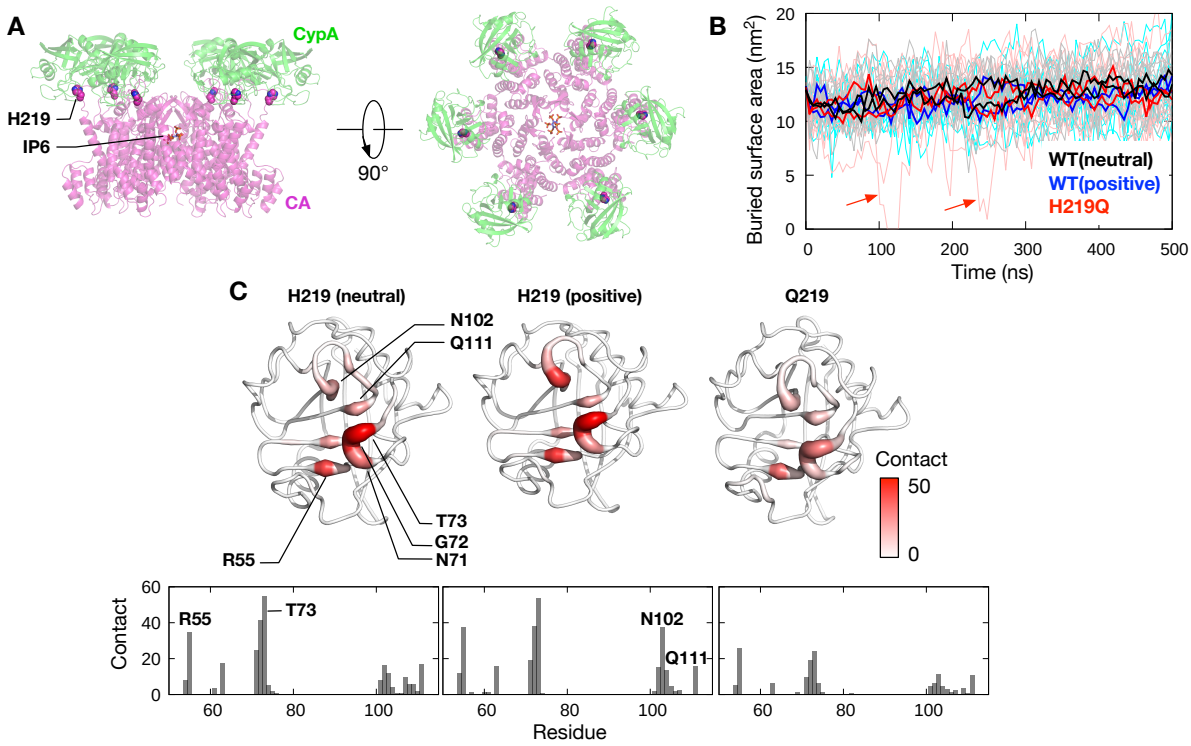


Figure 6: Atomistic simulations of mature CA with CypA. (A) Side and top views of mature CA hexamer (PDB: 6BHT) with each subunit bound to CypA (aligned using PDB: 1AK4). Residue H219 is shown in van der Waals representation, while IP6 bound in the center of the hexamer is shown in stick representation. (B) Buried surface area between CA and CypA throughout 500 ns simulations. Data taken from two independent simulations of CA with neutral H219 (black), protonated H219 (blue) and H219Q mutant (red). Thin lines indicate values from each of the six subunits, whilst the thick lines show the running averages. Red arrows show transient dissociation of CypA from CA in H219Q mutant simulations. (C) Residues on CypA that made contacts specifically with the side chains of neutral and protonated H219 (left and centre, respectively) from the WT CA simulations and Q219 (right) from the mutant CA simulations. CypA is shown in ribbon representation, the colour and thickness of which represent the percentage of contact made with CA during the simulations. A corresponding plot of per-residue contact percentages for each system is shown underneath, with key residues labelled. The cut-off distance used for contact analysis was 0.4 nm.

Mutation in the CA hexameric interface stabilizes oligomerisation

Finally, another non-cleavage site mutation in the CA domain occurs at position 199, in which a glutamine residue is mutated to a histidine[20]. In immature Gag, this residue is located inside the CA domain pointing towards the internal cavity. Our CG simulations show that this residue did not form any major interactions with other CA residues as it was fully exposed to the solvent within the cavity of the CA. However, as CA undergoes large conformational changes during maturation, the Q199 residue forms a part of the hexameric interface of a mature CA domain (Figure 7A). This

residue bridges the N-terminal domain of one CA subunit and the C-terminal domain of an adjacent subunit via polar interactions with residues such as Y301, K314 and Q311. We hypothesized that a mutation to histidine may alter the interactions between the two CA subdomains and therefore affect oligomerisation.

To test this hypothesis, we performed two repeat 500 ns atomistic simulations of mature CA for the WT and Q199H variants, the latter in either protonated or unprotonated states (Simulations 7, 13 and 14 in Table S2) (Figure 7B). We found that both neutral and charged histidine retained all major interactions as observed for glutamine in the WT protein; for example, the two most prevalent interacting partners from the neighbouring subunit were Y301 and L343, via polar and van der Waals interactions, respectively. Similar to the WT simulations, these residues were found to interact with H199 throughout most of the trajectories. However, the mutant simulations also showed additional contacts with some residues compared to WT. Thus, neutral H199 made significantly more frequent interactions with K314 and V313, while protonated H199 could form a salt bridge with E319. Protonation of H199 also modulates some interactions at the hexameric interface; for example, contact with R305 is noticeably reduced due to electrostatic repulsion, whilst intermittent contacts were observed with Q311 and E312, which explains the large standard deviations between the two repeat simulations. Pair interaction energies of both neutral and protonated H199 with surrounding residues from the neighbouring subunit are significantly more negative than that of Q199 (Figure 7C), indicating stronger interactions made by this histidine residue at the CA hexameric interface compared to the wild type glutamine. Our simulations therefore suggest that a histidine residue at position 199 potentially strengthens the binding interface between CA subunits to enhance the stability of the hexameric complex.

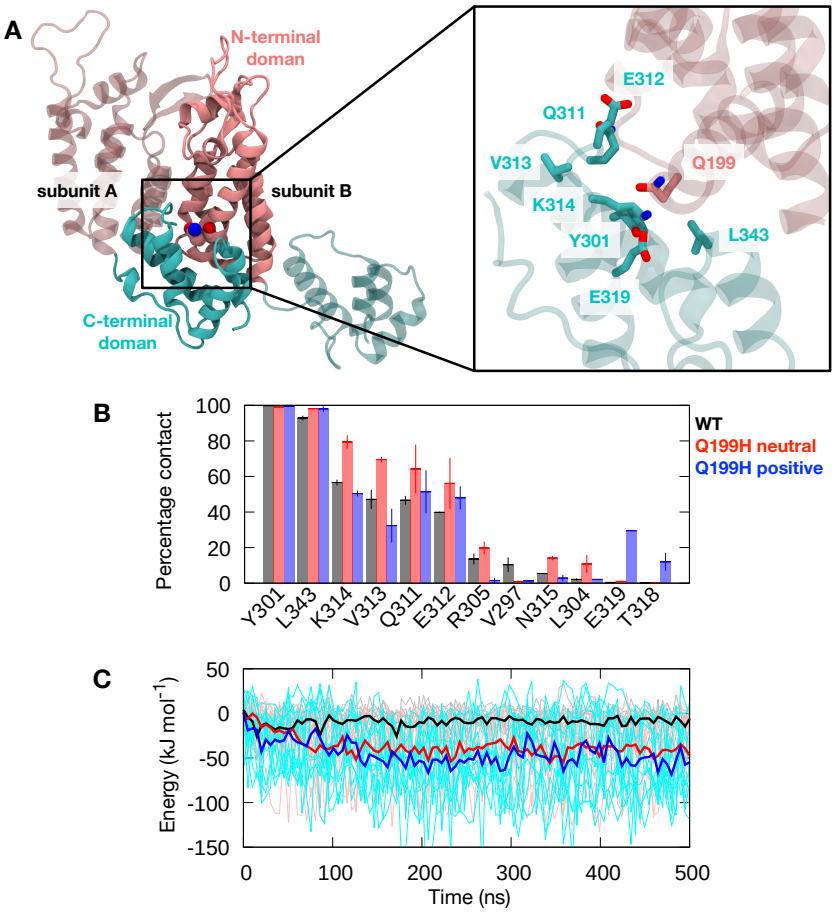


Figure 7: The role of the Q199H mutation in CA oligomerisation. (A) (Left) A snapshot of two subunits of the mature WT CA hexamer in cartoon representation. The Q199 residue bridging the N-terminal domain (pink) of one subunit and the C-terminal domain (cyan) of an adjacent subunit is shown in van der Waals representation. (Right) Enlarged image of the hexameric interface showing residues found within 0.4 nm of the Q199 residue. (B) Percentage contact made by residue Q199 (WT), neutral or positively charged H199 (mutants) with residues from adjacent CA subunits. Contacts are averaged over all six subunits and two independent simulations. Error bars indicate standard deviations between repeat simulations. The cut-off distance used for contact analysis was 0.4 nm. (C) Short-range Coulombic pairwise interaction energies between residue at position 199 on one subunit and surrounding residues on the adjacent subunit. Thin lines indicate values from each of the six subunits and two independent simulations, while thick lines show the running averages.

Discussion

Given the possible benefits of drugging Gag together with the protease, we set out to study cleaved Gag structures using multiscale modelling and simulations of the HIV-1 Gag protein to understand the effect of clinically relevant mutations that occur outside of the cleavage sites. We built a model of the MA-CA-SP1 polyprotein complex in its multimeric state, which constitutes the largest cleavage product of the immature Gag protein. Our simulations of the membrane-bound complex suggest that the long linker

532 connecting the MA and CA domains is flexible and may contract, enabling interactions
533 between protease cleavage site residues and the C-terminal region of the MA domain,
534 as well as several regions on the CA domain including the CypA binding loop. Two
535 clinical mutations, G123E and H219Q, are found in these regions and alter the
536 accessibility and electrostatic properties of the protease binding site. Mutations in the
537 MA domain are concentrated on the predicted membrane binding surface and at least
538 three of them – E12K, E40K and L75R – render the surface more positively charged
539 compared to wild type Gag. Our CG simulations of these mutant variants show more
540 pronounced interactions with PIP2 lipids, leading to the hypothesis that these
541 mutations improve Gag targeting and anchoring to the plasma membrane during viral
542 assembly. Atomic-resolution simulations of the mature Gag CA domain showed that
543 the H219Q mutation may play a role in optimizing CypA interactions by controlling the
544 conformational flexibility of the CypA binding loop and fine-tuning the CypA binding
545 affinity. We also found that a mutation in the CA domain, Q199H, likely enhances
546 oligomerisation by providing more stabilising interactions between adjacent subunits.
547 In the field of HIV and other related retroviruses, multiscale modelling and
548 simulations have been utilised to study CA in different maturation states, providing
549 important insights into their structural dynamics and interactions with drugs and host
550 cell proteins[38,67–69]. For example, landmark microsecond-timescale atomic simulations
551 of the entire HIV-1 capsid revealed important physicochemical properties, such as its
552 electrostatics, dynamics, and water/ion permeability, as well as pointing towards long-
553 range allosteric motions in the shell[35]. Simulations of the monomeric MA domain shed
554 light on the molecular mechanism underlying membrane anchoring by the myristoyl
555 group and PIP2 lipids[41]. Our study reports the first complete oligomeric model of the
556 proteolytic product of HIV-1 Gag polyprotein, allowing the investigation of interactions
557 between residues on the protease binding site and distant residues on the MA and CA
558 domains, especially those from different subunits. Such inter-subunit “crosstalk” would
559 be difficult to identify using individual crystal structures of mature Gag domains or a
560 monomeric Gag model. Understanding structural changes that occur upon proteolysis
561 could provide important clues on HIV-1 protease activity. For example, the MA-CA
562 linker contraction observed in our simulations resulted in a significant decrease in
563 solvent accessibility of the protease binding site, which helps to explain the lower
564 cleavage rate of the MA-CA domain compared to the SP2-P6 domain[49]. From the
565 perspective of anti-HIV drug development, targeting Gag proteolytic products in

566 addition to the mature form, by virtue of identifying intermediate structural models, may
567 allow for the discovery of Gag inhibitors that can inhibit multiple stages of maturation[14].
568 Since accumulation of deleterious mutations could negatively impact virion
569 survival, Gag polyprotein and protease co-compensate to outcompete protease
570 inhibitors and restore viral fitness[14,15,70]. While HIV-1 protease acquires drug resistant
571 mutations to evade PI activity, its ability to bind and subsequently cleave the Gag
572 protein is impaired. Thus, Gag gains additional mutations to compensate for fitness
573 defects to recover protease function. Structural studies show that associated
574 mutations on cleavage sites do not restore interactions lost due to mutations on the
575 protease, but rather establish novel interactions surrounding the site of protease
576 mutations as well as induce conformational changes to drive better binding[18]. Apart
577 from cleavage site mutations, mutations in more distant regions have been shown to
578 contribute towards recovery from the reduced viral fitness caused by protease
579 mutations[19,20]. Our study shows that these non-cleavage site mutations have far
580 reaching implications outside of Gag proteolysis, from improving membrane binding
581 to fine-tuning CypA interactions and stabilising the CA hexamer. A recent single
582 genome sequencing study revealed that, indeed, most Gag-protease correlated
583 mutations occur outside of the Gag cleavage site with the viral MA and CA domains
584 representing the largest subsets of the mutations[70]. Interestingly, a high concentration
585 of compensatory mutations was uncovered within the globular domain of MA (which
586 binds to the plasma membrane) and the CypA binding loop of CA, further corroborating
587 our simulation results. Mutations and deletions of residues on the C-terminal region of
588 the MA domain, which have direct contact with cleavage site residues based on our
589 simulations, have also been recently shown to confer resistance to PI lopinavir[71]. Our
590 results also highlight that the residues on the CypA binding loop can interact directly
591 with MA-CA cleavage site residues, thus potentially influencing their local
592 electrostatics and surface accessibility, although determining the exact interactions
593 made between these residues and subsequent molecular effects would require further
594 studies. Additionally, PS lipids are known to be exposed on the outer leaflet of HIV-1
595 membrane and may facilitate HIV-1 entry into the host cells[72,73], and it is possible that
596 the highly basic surface generated by the non-cleavage site mutations could attract
597 anionic PS lipids, similarly to the PIP2 enrichment around the MA domain observed
598 here. While our study assumes that these non-cleavage site mutations do not have
599 any large impact on the overall structure of Gag, it is possible that these point

mutations could result in local structural changes. Further structural and computational studies would thus be of interest to directly elucidate the impact of these mutations on Gag structure. It is worth noting that compensatory non-cleavage site mutations also occur in the NC and P6 domains. Gatanaga et al. discovered five such mutations with clinical importance[20], and many clinical drug resistant mutations can occur within the first replication cycle[7]. While modelling these domains is beyond the scope of the current study, it is certainly worth exploring in the future, specifically regarding how these mutations affect interactions with the viral RNA genome. Overall, our data have revealed new insights into the role of Gag non-cleavage site mutations on HIV-1 viral fitness. Given that HIV-1 is one of the fastest mutating RNA viruses, an in-depth comprehension of emerging mutations will aid in the general understanding of viral drug resistance, as well as the emergence of novel viruses that are able to cross species during zoonotic infections.

615 **Materials and Methods**

616 ***Model building and system setup***

617 The model of HIV-1 Gag MA-CA-SP1 was built using two structural templates: MA
618 trimers (PDB: 1HIW)[42] and the immature CA hexamer (PDB: 5L93)[43]. To the best of
619 our knowledge, the former represents the only available crystal structure of the MA
620 protein in its trimeric form, whilst the latter represents the highest resolution structure
621 of a complete immature CA hexamer. Six copies of MA trimers were arranged based
622 on EM images showing that they organize as hexamer-of-trimers on a model
623 membrane bilayer[44]. The CA hexamer was placed such that the N-termini of its
624 subunits align with the C-termini of the central subunits of the MA trimers. The MA and
625 CA domains are connected by a flexible, disordered linker as shown by solution NMR
626 studies[45,46]. Modeller version 9.21[74] was used to build the loops connecting the MA
627 trimers and CA hexamer (Figure S1) using the discreet optimized protein energy
628 (DOPE)-based loop modelling protocol, and the model with the lowest DOPE score
629 was chosen[75]. Stereochemical assessment using Ramachandran analysis[76] showed
630 only one outlier residue, confirming that the model was structurally reasonable. The
631 atomistic MA-CA-SP1 structure was subsequently converted to CG representation
632 using the MARTINI 2.2 force field[77] and the elastic network model, ElNeDyn[78], was
633 imposed to maintain the integrity of the secondary and tertiary structure. This CG
634 model was validated by comparing the MA-CA domain dynamics to all-atom
635 simulations of a monomeric MA-CA-SP1 protein (more details below). The N-terminal
636 glycine residue on each MA subunit was myristoylated based on parameters from
637 Charlier et al.[41] Three mutant variants were generated using PyMOL[79] based on non-
638 cleavage site mutations published by Gatanaga et al.[20] The wild-type residues were
639 substituted with the mutant versions using the PyMOL mutagenesis tool, assuming
640 these point mutations would not result in any large structural deviations from wild-type
641 Gag.

642 A 30 x 30 nm^2 patch representing the HIV-1 membrane model was built using
643 the CHARMM-GUI Martini Maker Bilayer Builder[80,81] and the lipid composition was
644 based on a previous HIV-1 lipidomic study[47]. This membrane was asymmetric: the
645 upper leaflet was made of 10% phosphatidyl choline (PC), 10% phosphatidyl ethanol
646 amine (PE), 30% sphingomyelin, and 50% cholesterol, whereas the lower leaflet was
647 made of 10% PC, 20% PE, 15% phosphatidyl serine (PS), 5% phosphatidylinositol-
648 4,5-bisphosphate (PIP2), and 50% cholesterol. The HIV-1 membrane is enriched in

649 saturated lipid species especially for PC; we therefore modelled the lipid tails of PC as
650 1,2-dipalmitoyl (DP), whereas the rest of the lipids were modelled with 1-palmitoyl-2-
651 oleoyl (PO) lipid tails. Following minimization and equilibration procedures according
652 to the CHARMM-GUI, the model membrane was further simulated for 1 μs to allow
653 better mixing of the lipid components prior to simulations in the presence of protein.
654
655 ***Molecular dynamics simulations***
656 The HIV-1 Gag MA-CA-SP1 CG model was placed underneath the membrane
657 such that the myristoylated N-termini were inserted into lower leaflet of the membrane.
658 Energy minimization was then performed using the steepest descent method to
659 remove any overlapping beads. The system was solvated with standard MARTINI
660 water molecules and neutralized with 0.15 M NaCl ions. Further energy minimization
661 was performed. The system was then equilibrated for 10 ns whereby positional
662 restraints with a force constant of 1000 kJ mol^{-1} nm^{-2} were imposed on all of the
663 backbone atoms of the protein. Temperature coupling using the V-rescale thermostat
664 with a time constant of 1 ps[82] was applied to maintain the temperature at 310 K. Semi-
665 isotropic pressure coupling using the Berendsen barostat with a time constant of 5 ps
666 was utilized to maintain the pressure of the system at 1 atm. Electrostatics were
667 calculated using the reaction field method. The der Waals interactions were computed
668 using a potential shift Verlet scheme. The short-range cut-offs for both of these were
669 set to 1.1 nm. After the equilibration simulations, four independent production runs of
670 15 μs were performed using different distributions of initial velocities. The same
671 protocols used for the equilibration simulation were used for the production runs in the
672 absence of position restraints, except for the pressure coupling in which the Parrinello-
673 Rahman barostat was used with a coupling constant of 12.0 ps[83].
674 Atomistic MD simulations were performed to understand the role of the H219Q
675 mutations on CypA binding and Q199H on CA oligomerisation. The structure of the
676 mature HIV-1 CA hexamer in complex with inositol IP6 (PDB: 6BHT)[40] was aligned to
677 that of the CA N-terminal domain bound to CypA (PDB: 1AK4)[55]. This alignment
678 generated a structure of the CA hexamer bound to six copies of CypA (Figure 6A).
679 Mutations in the CA were performed using PyMOL. The PROPKA3 programme[58] was
680 used to calculate the pKa of H219 (Table S3). The proteins were parameterized using
681 the CHARMM36 force field[84], whereas the CHARMM-GUI Ligand Reader & Modeller[85]
682 was used to generate the parameters for the IP6 molecule. The complex was solvated

683 with TIP3P water molecules and 0.15 M NaCl was added to neutralize the system.
684 The steepest descent minimization protocol was used to remove overlapping atoms.
685 A short 100 ps equilibration simulation with position restraints (force constant of 1000
686 kJ mol^{-1} nm^{-2}) applied to all heavy atoms of the protein was conducted. The
687 temperature of the system was maintained at 310 K using the Nosé-Hoover thermostat
688 with a time constant of 1.0 ps[86,87]. The pressure was kept at 1 atm using an isotropic
689 coupling to the Parrinello-Rahman barostat with a time constant of 5.0 ps[83].
690 Electrostatic interactions were measured using the smooth particle mesh Ewald (PME)
691 method with a real-space cut off of 1.2 nm[88]. The van der Waals interactions were
692 calculated using the force switch smoothing function applied between 1.0 and 1.2 nm,
693 and truncated at 1.2 nm. An integration time step of 2 fs was used with all covalent
694 bonds involving hydrogens constrained using the LINCS algorithm[89]. After
695 equilibration, the positional restraints were removed and two independent 500 ns
696 production runs were conducted using the same setup for each of the WT and mutant
697 variants starting with different initial velocities. To assess whether the systems were
698 properly equilibrated, we calculated the RMSD of the Cα atoms of CA (Simulation 7-
699 14 in Table S2) and CypA (Simulation 10-12 in Table S2) for all six chains of the
700 hexamer (Figure S10). We found that in most cases the RMSD reached a plateau after
701 around 100 ns.

702 Additional atomistic and CG simulations were performed to verify the
703 conformation of the MA-CA linker. A single subunit of the HIV-1 Gag MA-CA-SP1
704 model was placed in solution with 0.15 M NaCl ions. Positional restraints with a force
705 constant of 1000 kJ mol^{-1} nm^{-2} were applied to the backbone atoms of membrane
706 binding residues (residue 2-53 and 72-90) to mimic membrane binding. Four
707 independent 500 ns simulations were performed for both atomistic and CG systems
708 using the same protocols described above.

709 To refine interactions between the linker region and mutant residues observed
710 in the CG simulations of the MA-CA-SP1 hexamer, we converted the structure of the
711 hexamer from the final frame of one of the simulations to atomistic representation
712 using the CHARMM-GUI All-Atom Converter. The back-mapping protocol was
713 performed using a flexible geometric approach via the *backward.py* program[50]. In this
714 approach, the atomic particles are mapped to the weighted average of the CG beads,
715 followed by geometrical corrections whereby the atomic particles are repositioned
716 (when necessary) to reconstruct chiral centres and double bond configurations. Force

field-based relaxation consisting of a series of energy minimizations and position restrained MD simulations was then conducted. A single subunit of the MA-CA-SP1 was subsequently used to refine interactions with the G123E mutant, whereas two adjacent subunits were used for interactions with the H219Q mutant. The protein was inserted into a box of water with 0.15 M NaCl ions. Positional restraints with a force constant of 1000 kJ mol^{-1} nm^{-2} were imposed on the backbone atoms of residues outside of the MA-CA linker and CypA binding loop (residue 2-117, 147-215 and 233-376). For each system, three independent 200 ns simulations were performed using the parameters described above.

All simulations were performed using GROMACS 2018[90]. The list of simulations performed is provided in Table S2. Analysis of the electrostatic surface charge of the MA domain was performed in PyMOL using the APBS plugin[91]. Cluster analysis of the CypA binding loop (residue 216-232) was performed on structures generated from simulation trajectories using the GROMACS *gmx cluster* tool. The clustering algorithm described by Daura et al.[92] was applied using a cutoff of 0.3 nm.

Acknowledgments

This work was supported by BII and EDDC of A*STAR. Simulations were performed on the petascale computer cluster ASPIRE-1 at the National Supercomputing Centre of Singapore (NSCC) and the A*STAR Computational Resource Centre (A*CRC).

Competing interests

The authors declare no competing interests.

743 **References**

744 1. Yeo, J. Y., Goh, G. R., Tran-To Su, C. & Ken-En Gan, S. The determination of
745 HIV-1 RT mutation rate, its possible allosteric effects, and its implications on
746 drug resistance. *Viruses* **12**, 1–22 (2020).
747 2. Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in
748 viruses: Patterns and determinants. *Nat. Rev. Genet.* **9**, 267–276 (2008).
749 3. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science*
750 *(80-.).* **310**, 676–679 (2005).
751 4. Kuiken, T. *et al.* Host species barriers to influenza virus infections. *Science*
752 *(80-.).* **312**, 394–397 (2006).
753 5. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in
754 China. *Nature* (2020). doi:10.1038/s41586-020-2008-3
755 6. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of
756 probable bat origin. *Nature* **2019**, (2020).
757 7. Yi Yeo, J., Yap, P., Goh, G., Koh, D. W. & Gan, S. K. HIV-1 mutations in HIV-1
758 Gag , protease , RT p66 and when they appear : Insights from an in vitro BSL2
759 assay on mutation rates and types . *bioRxiv* 1–26 (2020).
760 8. Perelson, A. S. Modelling viral and immune system dynamics. *Nat. Rev.*
761 *Immunol.* **2**, 28–36 (2002).
762 9. Collier, D. A., Monit, C. & Gupta, R. K. The Impact of HIV-1 Drug Escape on
763 the Global Treatment Landscape. *Cell Host Microbe* **26**, 48–60 (2019).
764 10. Gregson, J. *et al.* Global epidemiology of drug resistance after failure of WHO
765 recommended first-line regimens for adult HIV-1 infection: A multicentre
766 retrospective cohort study. *Lancet Infect. Dis.* **16**, 565–575 (2016).
767 11. Su, C. T. T., Ling, W. L., Lua, W. H., Haw, Y. X. & Gan, S. K. E. Structural
768 analyses of 2015-updated drug-resistant mutations in hiv-1 protease: An
769 implication of protease inhibitor cross-resistance. *BMC Bioinformatics* **17**,
770 (2016).
771 12. Ajose, O., Mookerjee, S., Mills, E. J., Boulle, A. & Ford, N. Treatment
772 outcomes of patients on second-line antiretroviral therapy in resource-limited
773 settings: A systematic review and meta-analysis. *Aids* **26**, 929–938 (2012).
774 13. Collier, D. *et al.* Virological outcomes of second-line protease inhibitor-based
775 treatment for human immunodeficiency virus type 1 in a high-prevalence Rural
776 South African setting: A competing-risks prospective cohort analysis. *Clin.*

777 *Infect. Dis.* **64**, 1006–1016 (2017).
778 14. Su, C. T.-T., Koh, D. W.-S. & Gan, S. K.-E. Reviewing HIV-1 Gag Mutations in
779 Protease Inhibitors Resistance: Insights for Possible Novel Gag Inhibitor
780 Designs. *Molecules* **24**, 3243 (2019).
781 15. Clavel, F. & Mammano, F. Role of gag in HIV resistance to protease inhibitors.
782 *Viruses* **2**, 1411–1426 (2010).
783 16. Sutherland, K. A. *et al.* Gag-Protease Sequence Evolution Following Protease
784 Inhibitor Monotherapy Treatment Failure in HIV-1 Viruses Circulating in East
785 Africa. *AIDS Res. Hum. Retroviruses* **31**, 1032–1037 (2015).
786 17. Gupta, R. K., Kohli, A., Mccormick, A. L. & Towers, G. J. Full length HIV-1 Gag
787 determines protease inhibitor susceptibility within in vitro assays. *Aids* **24**,
788 1651–1655 (2010).
789 18. Özen, A., Lin, K. H., Yilmaz, N. K. & Schiffer, C. A. Structural basis and distal
790 effects of Gag substrate coevolution in drug resistance to HIV-1 protease.
791 *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15993–15998 (2014).
792 19. Myint, L. *et al.* Gag Non-Cleavage Site Mutations Contribute to Full Recovery
793 of Viral Fitness in Protease Inhibitor-Resistant Human Immunodeficiency Virus
794 Type 1. *Antimicrob. Agents Chemother.* **48**, 444–452 (2004).
795 20. Gatanaga, H. *et al.* Amino acid substitutions in Gag protein at non-cleavage
796 sites are indispensable for the development of a high multitude of HIV-1
797 resistance against protease inhibitors. *J. Biol. Chem.* **277**, 5952–5961 (2002).
798 21. Sutherland, K. A. *et al.* Evidence for reduced drug susceptibility without
799 emergence of major protease mutations following protease inhibitor
800 monotherapy failure in the SARA trial. *PLoS One* **10**, 1–14 (2015).
801 22. Su, C. T. T., Kwoh, C. K., Verma, C. S. & Gan, S. K. E. Modeling the full length
802 HIV-1 Gag polyprotein reveals the role of its p6 subunit in viral maturation and
803 the effect of non-cleavage site mutations in protease drug resistance. *J.*
804 *Biomol. Struct. Dyn.* **36**, 4366–4377 (2018).
805 23. Deshmukh, L., Louis, J. M., Ghirlando, R. & Clore, G. M. Transient HIV-1 Gag-
806 protease interactions revealed by paramagnetic NMR suggest origins of
807 compensatory drug resistance mutations. *Proc. Natl. Acad. Sci. U. S. A.* **113**,
808 12456–12461 (2016).
809 24. Bell, N. M. & Lever, A. M. L. HIV Gag polyprotein: Processing and early viral
810 particle assembly. *Trends Microbiol.* **21**, 136–144 (2013).

811 25. Franke, E. K., Yuan, H. E. H. & Luban, J. Specific incorporation of cyclophilin a
812 into HIV-1 virions. *Nature* **372**, 359–362 (1994).
813 26. Braaten, D., Franke, E. K. & Luban, J. Cyclophilin A is required for an early
814 step in the life cycle of human immunodeficiency virus type 1 before the
815 initiation of reverse transcription. *J. Virol.* **70**, 3551–60 (1996).
816 27. Carnes, S. K., Sheehan, J. H. & Aiken, C. Inhibitors of the HIV-1 capsid, a
817 target of opportunity. *Curr. Opin. HIV AIDS* **13**, 359–365 (2018).
818 28. Purdy, M. D. *et al.* MicroED structures of HIV-1 Gag CTD-SP1 reveal binding
819 interactions with the maturation inhibitor bevirimat. *Proc. Natl. Acad. Sci. U. S.*
820 *A.* **115**, 13258–13263 (2018).
821 29. Seclén, E. *et al.* High prevalence of natural polymorphisms in Gag (CA-SP1)
822 associated with reduced response to Bevirimat, an HIV-1 maturation inhibitor.
823 *AIDS* **24**, 467–469 (2010).
824 30. Verheyen, J. *et al.* High prevalence of bevirimat resistance mutations in
825 protease inhibitor-resistant HIV isolates. *Aids* **24**, 669–673 (2010).
826 31. Marzinek, J. K., Huber, R. G. & Bond, P. J. Multiscale modelling and simulation
827 of viruses. *Current Opinion in Structural Biology* (2020).
828 doi:10.1016/j.sbi.2019.12.019
829 32. Ayton, G. S., Noid, W. G. & Voth, G. Multiscale modeling of biomolecular
830 systems: in serial and in parallel. *Curr. Opin. Struct. Biol.* **17**, 192–8 (2007).
831 33. Grime, J. M. A. & Voth, G. A. Early stages of the HIV-1 capsid protein lattice
832 formation. *Biophys. J.* **103**, 1774–1783 (2012).
833 34. Grime, J. M. A. *et al.* Coarse-grained simulation reveals key features of HIV-1
834 capsid self-assembly. *Nat. Commun.* **7**, 11568 (2016).
835 35. Perilla, J. R. & Schulten, K. Physical properties of the HIV-1 capsid from all-
836 atom molecular dynamics simulations. *Nat. Commun.* **8**, (2017).
837 36. Pak, A. J. *et al.* Immature HIV-1 lattice assembly dynamics are regulated by
838 scaffolding from nucleic acid and the plasma membrane. *Proc. Natl. Acad. Sci.*
839 *U. S. A.* **114**, E10056–E10065 (2017).
840 37. Pak, A. J., Grime, J. M. A., Yu, A. & Voth, G. A. Off-Pathway Assembly: A
841 Broad-Spectrum Mechanism of Action for Drugs That Undermine Controlled
842 HIV-1 Viral Capsid Formation. *J. Am. Chem. Soc.* **141**, 10214–10224 (2019).
843 38. Liu, C. *et al.* Cyclophilin A stabilizes the HIV-1 capsid through a novel non-
844 canonical binding site. *Nat. Commun.* **7**, (2016).

845 39. Huang, P.-T. *et al.* FEZ1 Is Recruited to a Conserved Cofactor Site on Capsid
846 to Promote HIV-1 Trafficking. *Cell Rep.* **28**, 2373-2385.e7 (2019).
847 40. Dick, robert A. *et al.* Inositol phosphates are assembly co-factors for HIV-1.
848 *Nature* **560**, 509–512 (2018).
849 41. Charlier, L. *et al.* Coarse-grained simulations of the HIV-1 matrix protein
850 anchoring: Revisiting its assembly on membrane domains. *Biophys. J.* **106**,
851 577–585 (2014).
852 42. Hill, C. P., Worthylake, D., Bancroft, D. P., Christensen, A. M. & Sundquist, W.
853 I. Crystal structures of the trimeric human immunodeficiency virus type 1
854 matrix protein: implications for membrane association and assembly. *Proc.*
855 *Natl. Acad. Sci. U. S. A.* **93**, 3099–104 (1996).
856 43. Schur, F. K. M. *et al.* An atomic model of HIV-1 capsid-SP1 reveals structures
857 regulating assembly and maturation. *Science (80-.).* **353**, 506–508 (2016).
858 44. Alfadhli, A., Barklis, R. L. & Barklis, E. HIV-1 matrix organizes as a hexamer of
859 trimers on membranes containing phosphatidylinositol-(4,5)-bisphosphate.
860 *Virology* **387**, 466–472 (2009).
861 45. Tang, C., Ndassa, Y. & Summers, M. F. Structure of the N-terminal 283-
862 residue fragment of the immature HIV-1 Gag polyprotein. *Nat. Struct. Biol.* **9**,
863 537–543 (2002).
864 46. Deshmukh, L., Ghirlando, R. & Clorea, G. M. Conformation and dynamics of
865 the Gag polyprotein of the human immunodeficiency virus 1 studied by NMR
866 spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3374–3379 (2015).
867 47. Brugger, B. *et al.* The HIV lipidome: A raft with an unusual composition. *Proc.*
868 *Natl. Acad. Sci.* **103**, 2641–2646 (2006).
869 48. Mattei, S., Glass, B., Hagen, W. J. H., Kräusslich, H. G. & Briggs, J. A. G. The
870 structure and flexibility of conical HIV-1 capsids determined within intact
871 virions. *Science (80-.).* **354**, 1434–1437 (2016).
872 49. Pettit, S. C. *et al.* The p2 domain of human immunodeficiency virus type 1 Gag
873 regulates sequential proteolytic processing and is required to produce fully
874 infectious virions. *J. Virol.* **68**, 8017–27 (1994).
875 50. Wassenaar, T. A., Pluhackova, K., Böckmann, R. A., Marrink, S. J. &
876 Tieleman, D. P. Going backward: A flexible geometric approach to reverse
877 transformation from coarse grained to atomistic models. *J. Chem. Theory*
878 *Comput.* **10**, 676–690 (2014).

879 51. Ono, A., Ablan, S. D., Lockett, S. J., Nagashima, K. & Freed, E. O.
880 Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the
881 plasma membrane. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 14889–14894 (2004).
882 52. Saad, J. S. *et al.* Structural basis for targeting HIV-1 Gag proteins to the
883 plasma membrane for virus assembly. *Proc. Natl. Acad. Sci.* **101**, 14889–
884 14894 (2006).
885 53. Ono, A. & Freed, E. O. Plasma membrane rafts play a critical role in HIV-1
886 assembly and release. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 13925–13930
887 (2001).
888 54. Thali, M. *et al.* Functional association of cyclophilin a with HIV-1 virions. *Nature*
889 **372**, 363–365 (1994).
890 55. Gamble, T. R. *et al.* Crystal structure of human cyclophilin A bound to the
891 amino-terminal domain of HIV-1 capsid. *Cell* **87**, 1285–1294 (1996).
892 56. Sokolskaja, E., Sayah, D. M. & Luban, J. Target Cell Cyclophilin A Modulates
893 Human Immunodeficiency Virus Type 1 Infectivity. *J. Virol.* **78**, 12800–12808
894 (2004).
895 57. Yin, L., Braaten, D. & Luban, J. Human immunodeficiency virus type 1
896 replication is modulated by host cyclophilin A expression levels. *J. Virol.* **72**,
897 6430–6 (1998).
898 58. Olsson, M. H. M., SØndergaard, C. R., Rostkowski, M. & Jensen, J. H.
899 PROPKA3: Consistent treatment of internal and surface residues in empirical
900 pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
901 59. Campbell, S. *et al.* Modulation of HIV-like particle assembly in vitro by inositol
902 phosphates. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10875–10879 (2001).
903 60. Munro, J. B. *et al.* A Conformational Transition Observed in Single HIV-1 Gag
904 Molecules during In Vitro Assembly of Virus-Like Particles. *J. Virol.* **88**, 3577–
905 3585 (2014).
906 61. Datta, S. A. K. *et al.* Interactions between HIV-1 Gag Molecules in Solution: An
907 Inositol Phosphate-mediated Switch. *J. Mol. Biol.* **365**, 799–811 (2007).
908 62. Lu, M. *et al.* Atomic-resolution structure of HIV-1 capsid tubes by magic-angle
909 spinning NMR. *Nat. Struct. Mol. Biol.* **27**, 863–869 (2020).
910 63. Harrison, R. K. & Stein, R. L. Substrate Specificities of the Peptidyl Prolyl Cis-
911 Trans Isomerase Activities of Cyclophilin and FK-506 Binding Protein:
912 Evidence for the Existence of a Family of Distinct Enzymes. *Biochemistry* **29**,

913 3813–3816 (1990).
914 64. Ke, H., Mayrose, D. & Cao, W. Crystal structure of cyclophilin A complexed
915 with substrate Ala-Pro suggests a solvent-assisted mechanism of cis-trans
916 isomerization. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3324–3328 (1993).
917 65. Yezid, H., Konate, K., Debaisieux, S., Bonhoure, A. & Beaumelle, B.
918 Mechanism for HIV-1 tat insertion into the endosome membrane. *J. Biol.*
919 *Chem.* **284**, 22736–22746 (2009).
920 66. Ni, T. *et al.* Intrinsic curvature of the HIV-1 CA hexamer underlies capsid
921 topology and interaction with cyclophilin A. *Nat. Struct. Mol. Biol.* **27**, 855–862
922 (2020).
923 67. Zhao, G. *et al.* Mature HIV-1 capsid structure by cryo-electron microscopy and
924 all-atom molecular dynamics. *Nature* **497**, 643–646 (2013).
925 68. Ayton, G. S. & Voth, G. A. Multiscale computer simulation of the immature
926 HIV-1 virion. *Biophys. J.* **99**, 2757–2765 (2010).
927 69. Goh, B. C. *et al.* Atomic Modeling of an Immature Retroviral Lattice Using
928 Molecular Dynamics and Mutagenesis. *Structure* **23**, 1414–1425 (2015).
929 70. Codoñer, F. M. *et al.* Gag-protease coevolution analyses define novel
930 structural surfaces in the HIV-1 matrix and capsid involved in resistance to
931 Protease Inhibitors. *Sci. Rep.* **7**, (2017).
932 71. Datir, R. *et al.* In Vivo Emergence of a Novel Protease Inhibitor Resistance
933 Signature in HIV-1 Matrix. *bioRxiv* 865840 (2019). doi:10.1101/865840
934 72. Callahan, M. K. *et al.* Phosphatidylserine on HIV Envelope Is a Cofactor for
935 Infection of Monocytic Cells. *J. Immunol.* **170**, 4840–4845 (2003).
936 73. Zaitseva, E. *et al.* Fusion Stage of HIV-1 Entry Depends on Virus-Induced Cell
937 Surface Exposure of Phosphatidylserine. *Cell Host Microbe* **22**, 99-110.e7
938 (2017).
939 74. Sali, A. & Blundell, T. Comparative protein modelling by satisfaction of spatial
940 restraints. *J. Mol. Biol.* **234**, 779–815 (1994).
941 75. Shen, M., Devos, D., Melo, F. & Sali, A. A composite score for predicting
942 errors in protein structure models. *Protein Sci.* **15**, 1653–1666 (2006).
943 76. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry
944 of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
945 77. Monticelli, L. *et al.* The MARTINI coarse-grained force field: Extension to
946 proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008).

78. Periole, X., Cavalli, M., Marrink, S. & Ceruso, M. A. Combining an Elastic Network With a Coarse-Grained Molecular Force Field : Structure , Dynamics , and Intermolecular Recognition. *J. Chem. Theory Comput.* **5**, 2531–2543 (2009).
79. DeLano, W. The PyMOL molecular graphics system. (2002).
80. Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **97**, 50–58 (2009).
81. Qi, Y. *et al.* CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *J. Chem. Theory Comput.* **11**, 4486–4494 (2015).
82. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
83. Parrinello, M. & Rahman, A. Polymorphic Transitions in Single Crystals: a New Molecular Dynamics Method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
84. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
85. Kim, S. *et al.* CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J. Comput. Chem.* **38**, 1879–1886 (2017).
86. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
87. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
88. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577 (1995).
89. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* **18**, 1463–1472 (1997).
90. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
91. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–41 (2001).

981 92. Daura, X. *et al.* Peptide Folding: When Simulation Meets Experiment. *Angew.*
982 *Chemie* **38**, 236–240 (1999).
983