The Structural Basis for Gag Non-Cleavage Site Mutations in Determining HIV-1 Viral Fitness

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

Abstract

The high mutation rate in retroviruses is one of the leading causes of rampant drug resistance and the emergence of novel infectious diseases. 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 effects of the mutations on viral fitness. Some Gag mutations have been reported to restore Gag-protease binding, yet most Gag-protease correlated mutations occur outside of the Gag cleavage site. To rationalize this, we 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. Simulations of the complete oligomeric structure of the largest Gag proteolytic product in its viral membrane-bound state revealed how non-cleavage site mutations can directly interact with cleavage site residues to affect their local environment, facilitated by conformational changes upon lipid interaction. Mutations in the matrix domain led to the enrichment of phosphatidylinositol bisphosphate (PIP2) lipids – whose association is essential for Gag targeting and assembly on the plasma membrane – facilitated by creation of novel PIP2 binding sites. Additional studies of the mature CA hexamer suggest that some mutations can modulate recruitment of cyclophilin A (CypA) into the mature virion, as well as stabilise the oligomerisation of the viral core. Collectively, our results reveal that 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 mechanism. 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, most prominently 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 major component of the highly active antiretroviral therapy (HAART) in the management of HIV; however, emerging mutations in HIV-1 protease often render PIs ineffective, with some mutations even leading to cross-drug resistance[9]. This is further aggravated by synergistic mutations on the Gag polyprotein itself[10,11], making Gag an attractive drug target in addition to the protease[10]. 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[12]. 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[13,14]. To date, the role of these non-cleavage site mutations is largely unknown, with some analysis pointing to allosteric communication[15]. In this study, we aim to use structural modelling and simulations to decipher how Gag non-cleavage site mutations contribute towards the overall fitness of the HIV-1 virus.

The HIV-1 Gag polyprotein is a 500 amino acid precursor protein containing the matrix (MA), capsid (CA), nucleocapsid (NC) and P6 domains, as well as two spacer peptides, SP1 and SP2. During the late phase of the HIV-1 replication cycle, Gag is sequentially cleaved by the protease enzyme into these domains (Figure S1), which subsequently form an infectious mature virion[16]. Each domain plays a specific role during maturation; for example, the MA domain drives full-length Gag to assemble at the plasma membrane of the host cell, while the CA and NC domains encapsulate the viral RNA genome. The CA domain also interacts with CypA, which is a host cell protein that is incorporated into the virion and is essential for capsid uncoating[17,18]. Due to its functional significance in the HIV-1 life cycle, Gag is an attractive target for therapeutic agents and to date, several drugs that target the CA domain have been identified[19]. Nevertheless, a Gag inhibitor is yet to be clinically approved. For example, Bevirimat is a drug candidate that stabilises the immature Gag lattice by preventing proteolysis between the CA and SP1 domains[20]. A clinical trial, however, showed a reduced response amongst certain patients due to a high prevalence of Gag polymorphisms and resistant mutations[21,22], which highlights the importance of understanding the role of Gag non-cleavage site mutations.

Over the past few years, integrative modelling and molecular dynamics (MD) simulations, in tandem with advances in structural biology, have provided valuable insights into the molecular mechanism underlying viral function and dynamics[23]. In the study of HIV-1, this ranges from multiscale simulations of the entire viral capsid shell and its assembly pathways[24–29], to simulations of individual Gag proteins with host cell components, such as the CA domain with CypA[30], kinesin-1 adaptor protein FEZ1[31] and inositol hexakisphosphate (IP6)[32], as well as the MA domain with a model plasma membrane[33]. A previous structural model of full-length monomeric HIV-1 Gag revealed allosteric communications between non-cleavage site mutations and the first Gag cleavage site[15], providing a glimpse into how residues far away from the protease cleavage site could affect proteolysis. However, this model did not take into account *in vivo* Gag oligomerisation, which is instrumental during virus particle maturation. Crystal structures show that the MA domain exists as a trimer[34], whilst the CA domain forms a hexamer[35]. An electron microscopy (EM) study of the MA protein in a PIP2 containing membrane showed that under higher order conditions, it organizes into hexamers-of-trimers[36].

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

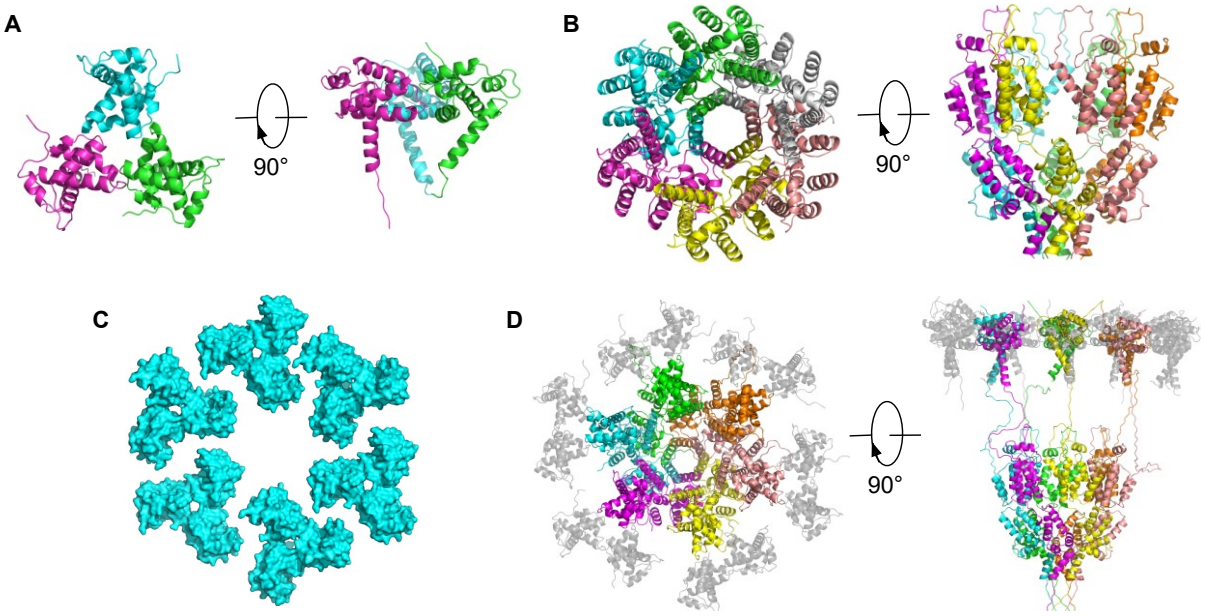

Figure 1: Modelling HIV-1 Gag MA-CA-SP1. (A) The crystal structure of the MA trimer (PDB: 1HIW). (B) The crystal structure of the immature CA hexamer (PDB: 5L93). (C) The hexamer of trimers model for the MA domains as supported by EM. (D) A model of the MA-CA-SP1 hexamer. The central subunits of the MA trimers are connected to the CA domains and coloured as in (B), whereas the peripheral MA subunits are coloured grey.

Results

Membrane dynamics of the HIV-1 Gag MA-CA-SP1 polyprotein

The Gag polyprotein is sequentially cleaved by HIV-1 protease during viral maturation (Figure S1). The first cleavage occurs between the SP1 and NC domains to produce the MA-CA-SP1 polyprotein, which represents the largest Gag cleavage product. A hexameric model of the HIV-1 Gag MA-CA-SP1 polyprotein was built using the crystal structures of the MA trimer[34] and CA hexamer[35], with the intervening linker between the MA and CA domains constructed *ab initio* (see Methods section for further details) (Figure 1). The crystal structure of the immature CA hexamer was aligned with six copies of MA trimers (hexamers-of-trimers) following the arrangement observed in membrane bilayers[36]. Gatanaga et al. discovered PI-resistant Gag variants with non-cleavage site mutations mapped to the MA-CA-SP1 complex (Table S1, Figure S1)[14]. These include seven unique mutations, with five located in the MA domain and two in the CA domain. We generated models for three variants (specifically variant numbers 3, 6, and 7, as outlined in Tables S1-S2) that cover all of these seven unique mutations (Figure 2A). To understand the conformational dynamics of the polyprotein and the potential roles of these non-cleavage site mutations, we performed four independent 15 μs CG MD simulations of the MA-CA-SP1 model bound to a realistic HIV-1 membrane model (Table S2). The membrane model contained phosphatidylcholine (PC), phosphatidyl ethanolamine (PE) and phosphatidylserine (PS) lipids, as well as PIP2, sphingomyelin and cholesterol, as determined by a previous lipidomics study[37] (see Methods section for further details), whilst the MA domain N-terminus was myristoylated (Figure 2B).

As previously predicted[33], the myristate group on the N-terminal glycine residue of MA domain effectively anchored the protein to the membrane. Within the first few nanoseconds, the linker connecting the MA and CA domains contracted, effectively pulling the CA-SP1 domain towards the membrane (Figure 2B). To verify these domain-domain dynamics observed during CG sampling, we also performed four independent 500 ns atomistic simulations of the MA-CA-SP1 monomer with positional restraints applied to the backbone atoms of lipid binding residues to mimic membrane association. Additionally, we also conducted CG simulations of the same monomeric system. Both of these additional sets of simulations showed a contraction of the MA-CA linker, resulting in a similar distribution of distances between the MA and CA domains compared to the CG hexamer simulations (Figure S2). While the atomistic

simulations initially sampled a wider spread of distances, after around 300 ns this converged to 4-6 nm, reproducing the CG distributions. This validates our CG hexamer model and shows that the linker region does not maintain an extended conformation, but rather contracts to allow direct contact between the MA and CA domains.

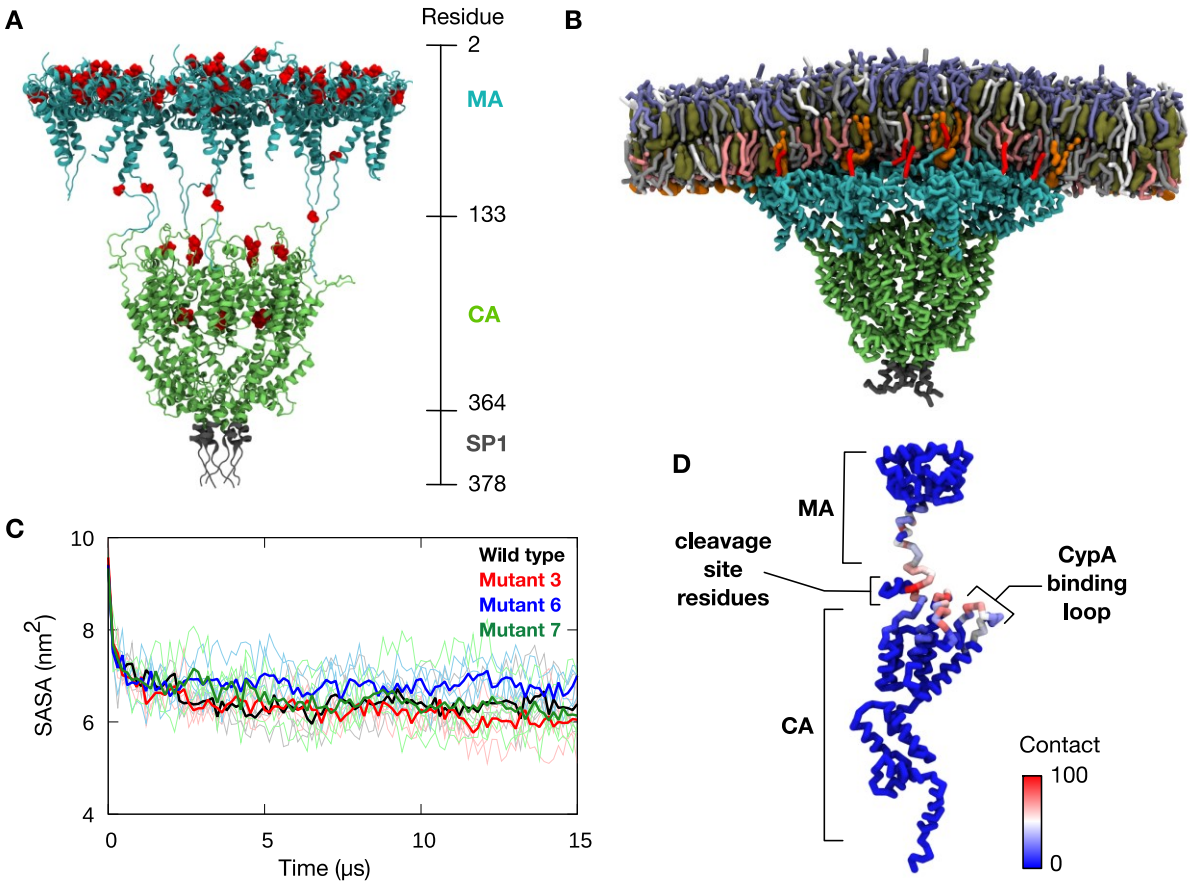

Figure 2: HIV-1 Gag MA-CA-SP1 model. (A) An atomic model of MA-CA-SP1 hexamer built using the crystal structures of MA trimer[34] and immature CA hexamer[35], shown in cartoon representation. The MA domain (cyan), CA domain (green), and SP1 domain (dark grey) are shown in different colours. The positions of seven unique non-cleavage site mutations[14] 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.

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 nm^2 (Figure 2C). After the first round of Gag proteolysis, the MA domain is cleaved off from the MA-CA-SP1 polyprotein, and the P6 domain is cleaved from the NC-SP2-P6 polyprotein (Figure S1). However, the rate at which the former occurs is noticeably slower than that of the latter[38]. Our simulations suggest that this may be caused by the reduction in exposure of the cleavage site, due to the contraction of the linker region between MA and CA domains. This linker contraction also led to the cleavage site residues interacting with neighbouring subdomains, predominantly with the C-terminal helix of the MA domain and several inter-helical loops on the CA domain, including the CypA binding loop (Figure 2D).

Two PI-resistant mutations were identified at the interaction sites – G123E and H219Q. To understand how these mutations may affect the cleavage site residues, contact analysis was performed between the cleavage site residues and these two specific positions in the WT and mutant proteins, particularly focusing on potential inter- and intra-subunit interactions. Due to their close proximity, we found that the residue at position 123 made contact primarily with the N-terminal portion of the cleavage site from the same Gag subunit (Figure 3A and 3B). Both WT (glycine) and mutant (glutamate) residues showed a similar percentage of contact over the course of the simulations. “Back-mapping” of CG simulation snapshots to transform them into atomic resolution suggested that a glutamate residue at this position may interact with polar residues on the cleavage site, such as Q130 and Y132 (Figure 3C). We performed three independent 200 ns atomistic simulations of a single MA-CA-SP1 subunit to refine the back-mapped structure, and indeed found that the glutamate residue interacted primarily with Y132 (Figure S3). Additionally, the residue also contacted residues N131 and Q130. In contrast, a glycine residue at this position interacted primarily with V128 and S129. Given the change in the overall size and charge of the residue in the WT and mutants (from small and neutral to large and acidic), the G123E mutation alters the accessibility and electrostatic properties in the

vicinity of the cleavage site and would therefore be expected to directly interfere with proteolysis.

While the interactions of G123E with the cleavage site residues occur within the same subunit of the MA-CA-SP1 polyprotein, inter-subunit interactions were observed for H219, whereby this residue contacted the cleavage site of an adjacent subunit (Figure 3D). The frequency of such interactions, however, was notably lower than that displayed by G123E, and interestingly, mutation to glutamine resulted in a reduced contact with the cleavage site (Figure 3E). Back-mapping to atomistic representation suggested potential for interactions with polar residues such as N131 and N137 (Figure 3F). Atomistic simulations of two adjacent MA-CA-SP1 subunits using the back-mapped structure showed that histidine at this position made intermittent contacts with I134, V135, Q136 and N137, whilst glutamine interacted primarily with Y132 (Figure S3). Given that the WT histidine may potentially become protonated, it may affect local electrostatic surface properties of the Gag cleavage site to modulate the catalytic efficiency of the protease enzyme.

Overall, our data suggest that even though these mutations occur far from the protease binding site, they can physically interact with the cleavage site residues, either within the same Gag subunit or with a neighbouring subunit, as the linker region between MA and CA domain contracts following the first proteolytic cleavage.

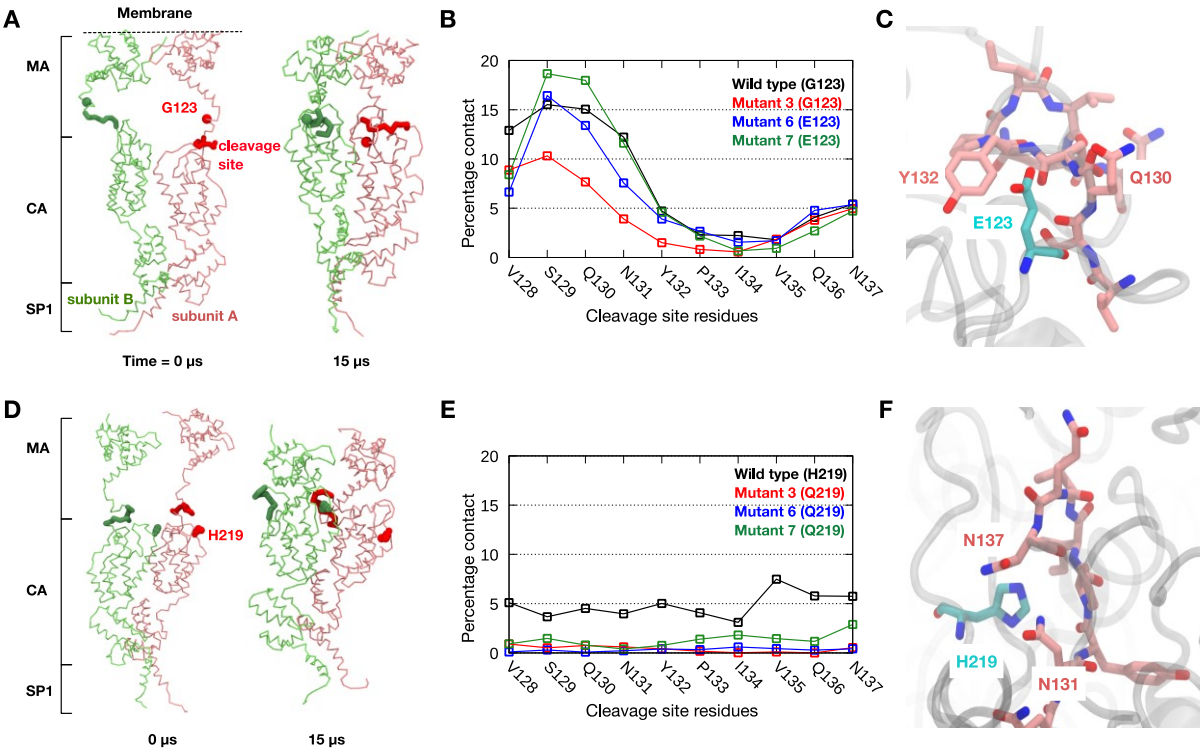

Figure 3: Interactions between non-cleavage site mutations and cleavage site residues. (A) Snapshot at the beginning and end of the CG simulations of WT MA-CA-SP1 showing the interaction of residue G123 (sphere representation) and cleavage site residues (thick licorice representation). Two subunits are 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. (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 S4). 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

almost the entire membrane binding surface including the central region at the interface of the MA trimer subunits.

Previous biochemical and structural studies showed that PIP2 is instrumental in targeting Gag to the plasma membrane of the host cell, and that it binds directly to the MA domain[39,40]. To understand how the changes in electrostatic surface properties caused by the PI-resistant mutations affect membrane binding, we determined the number of PIP2 lipids interacting with the MA domains throughout our CG simulations. Over the course of the 15 μs simulations, there was a significant enrichment of PIP2 lipids around the MA domain, consistently observed across all trajectories, from around 30% at the beginning of the simulations to 60-80% by the end (Figure 4A). This was due to electrostatic attraction between the anionic PIP2 headgroups and the positively charged membrane-peripheral surface of the MA domain. Interestingly, we did not observe any increment in the amount of anionic PS lipids despite its higher concentration compared to PIP2 in the membrane (Figure 4B). This may be due to the higher negative charge on the headgroups of PIP2 compared to PS lipids. PIP2 enrichment around the MA domain is in agreement with previous experimental studies showing that PIP2 is responsible for anchoring HIV-1 Gag to the plasma membrane[41]. While both WT and mutant variants displayed PIP2 enrichment, all three mutants showed a larger degree of accumulation of PIP2 around MA. More importantly, the E40K and L75R mutations, which map around the interface of the MA trimer subunits, formed novel binding sites for PIP2 lipids (Figure 4C, 4D). Our simulations therefore suggest that these non-cleavage site mutations enhance interactions between the MA domain and PIP2 lipids, and may consequently improve membrane targeting of HIV-1 Gag during viral assembly.

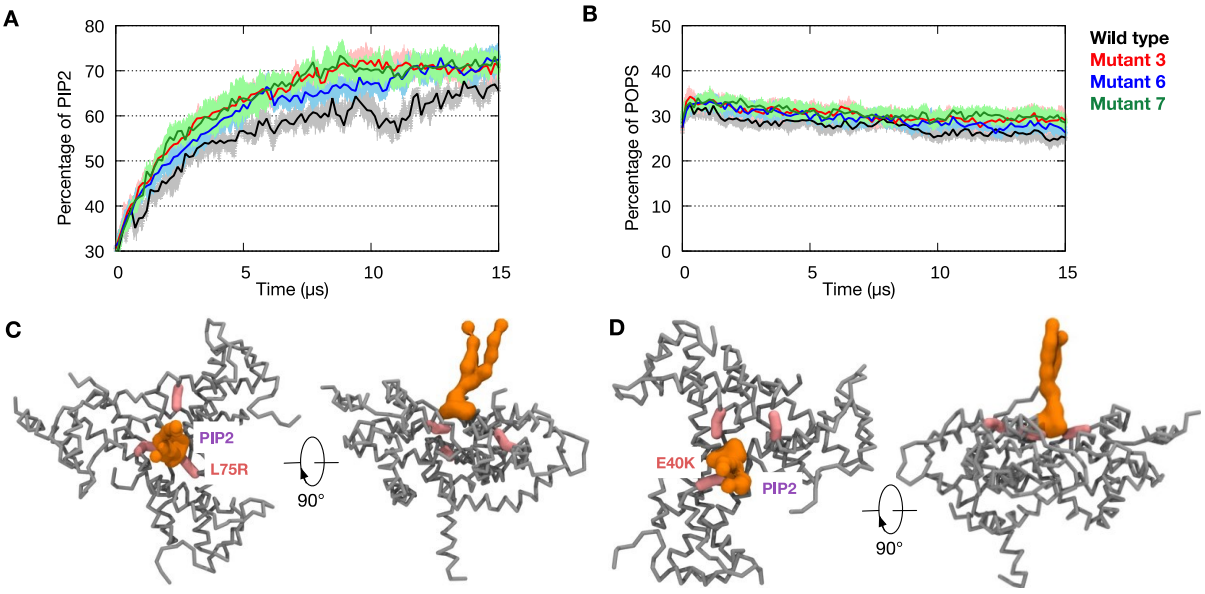

Figure 4: PIP2 is enriched around MA domain. (A) The percentage of PIP2 found within 0.6 nm of the MA domain during CG simulations of WT and mutant MA-CA-SP1 models. Error bars indicate standard deviation from four independent simulations. (B) The same analysis as in (A) performed for PS lipids. (C) Representative 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) Representative snapshot from simulations of mutant 7 showing PIP2 binding facilitated by E40K mutation, displayed in similar colours and representation as in (C).

Mutations in CypA-binding loop optimize CypA recruitment

Cyclophilin A (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[18,42,43]. Recruiting an optimal concentration of CypA is critical to viral fitness. 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 CA subunits are greatly destabilised, resulting in an unstable virion core and thereby delayed virion maturation[44]. CypA binds to CA via an exposed proline-rich unstructured loop region in the N-terminal domain of the latter[43]. Interestingly, this loop houses one of the most prevalent PI-resistant non-cleavage site mutations, H219Q. This mutation is found in all seven variants from Gatanaga et al (Table S1). To investigate the potential role of this mutation *vis-à-vis* CA-CypA interactions, we

performed two independent 500 ns atomistic MD simulations of the WT and H219Q mutant of the mature CA hexamer either in its CypA-free (apo) or CypA-bound state. As the imidazole ring of a histidine side chain has a pKa value that is close to physiological pH, it is possible that H219 may exist in either protonated or unprotonated form. As such, the simulations of WT CA were performed with H219 in both protonation states. IP6 has been shown to be an important assembly co-factor for HIV-1[32]. Crystal structures show that IP6 binds to the centre of the Gag CA hexamer and promotes the formation of the mature capsid lattice. As such, we included the IP6 molecule in all of our atomistic simulations of mature CA hexamer (see Methods section for further details).

Our apo simulations revealed that the CypA binding loop was the most dynamic region of the entire CA molecules (Figure 5A). No noticeable difference was found between the flexibility of this loop in simulations with the neutral H219 *versus* the H219Q mutant. Interestingly, when this histidine residue was protonated, we saw a slight increase in the RMSF values of the loop, especially around G221-P222, which is the putative substrate for the CypA rotamase activity. An increased association was observed between protonated H219 and E230, found on the opposite side of the loop (Figure 5B), as a result of salt bridge formation (Figure 5C). However, this salt bridge was intermittent (Figure 5D), such that the CypA-binding loop may adopt alternative conformational states when H219 is protonated. This means that under low pH conditions, the transition between these two states increases the overall flexibility of the loop, which consequently may make it more difficult for CypA to bind. Mutation of H219 to glutamine would therefore abolish the ability of the loop to adopt these two conformational states, reducing the loop flexibility, and facilitate initial CypA binding. This would be advantageous inside the endosome which is associated with a lower local pH, to ensure sufficient CypA binding and incorporation into the mature virion for subsequent core disassembly[45].

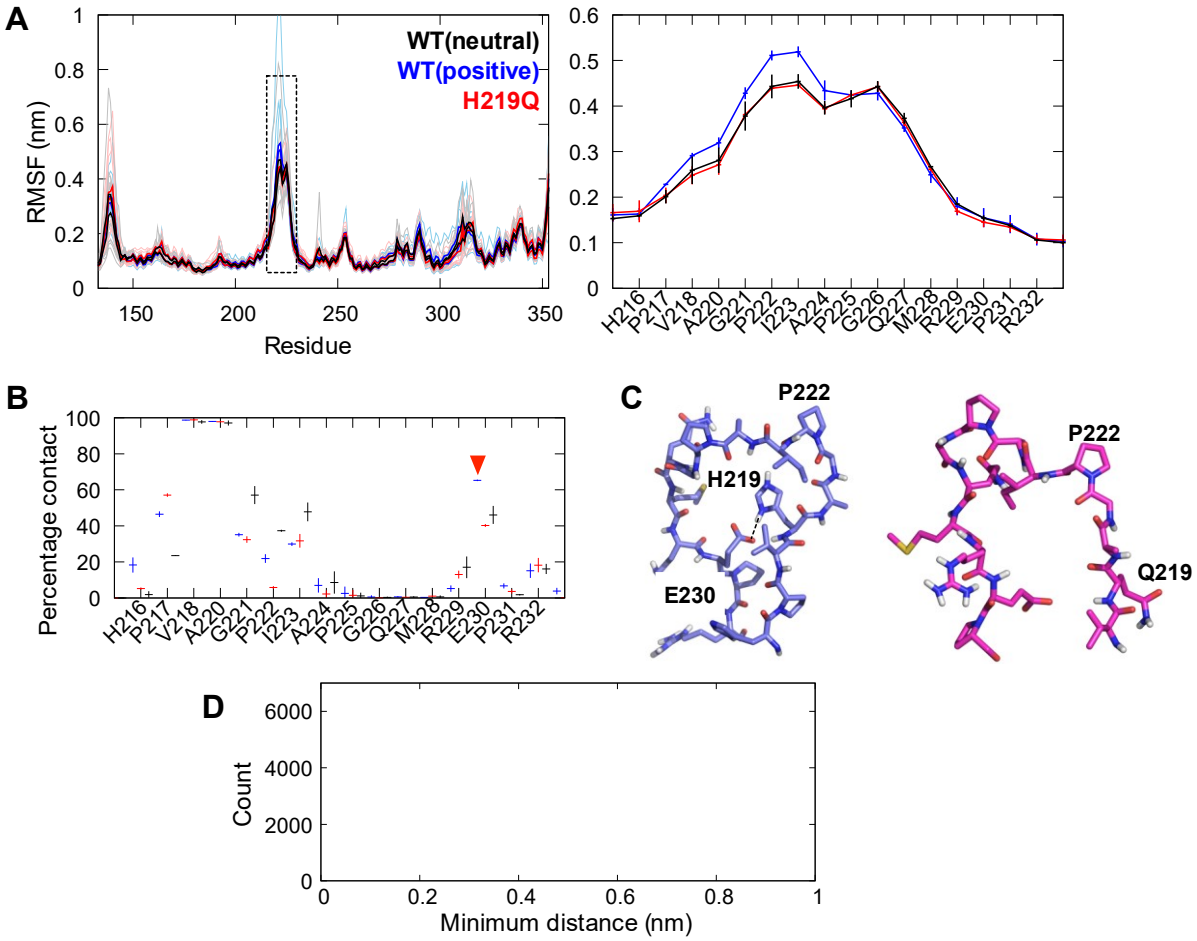

Figure 5: The dynamics of CypA-binding loop. (A) (Left) Per-residue root mean square fluctuations (RMSF) of the CA from two independent 500 ns apo simulations. Thin lines show the RMSF of each individual CA subunit, whilst the thick lines show the average from each simulation. Simulation with neutral H219 is shown in black, positively charged H219 in blue, and mutant H219Q in red. Dotted box indicates the loop region where CypA binds. (Right) Enlarged RMSF plot for CypA-binding loop. Average from the two simulations is shown and error bars indicate standard deviations between repeat simulations. (B) 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. (C) Representative structures of the CypA-binding loop from the simulations with positively charged H219 (left) and mutant H219Q (right) calculated using cluster analysis (representative structures of the top clusters are shown). 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. (D) 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[43], 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[43], 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 over-recruitment of CypA may destabilise the viral core. Taken together, the H219Q mutation on the CypA-binding loop may play a crucial role in fine-tuning the packaging of CypA into virions, especially in low endosomal pH and CypA-enriched host environments. It is worth noting that a cryo-EM study showed CypA could simultaneously bind two CA subunits from adjacent pairs of hexamers through a second, non-canonical binding site[30]. However, previous MD simulations have revealed that the interaction at this second site is weaker than the canonical binding site; we therefore did not focus on this second binding site. Nevertheless, as the binding is facilitated primarily by residues A220 and P222, both of which became more mobile upon H219 protonation in on our apo simulations (Figure 5A), it is conceivable that CypA binding at this second site may also be regulated by the H219Q mutation.

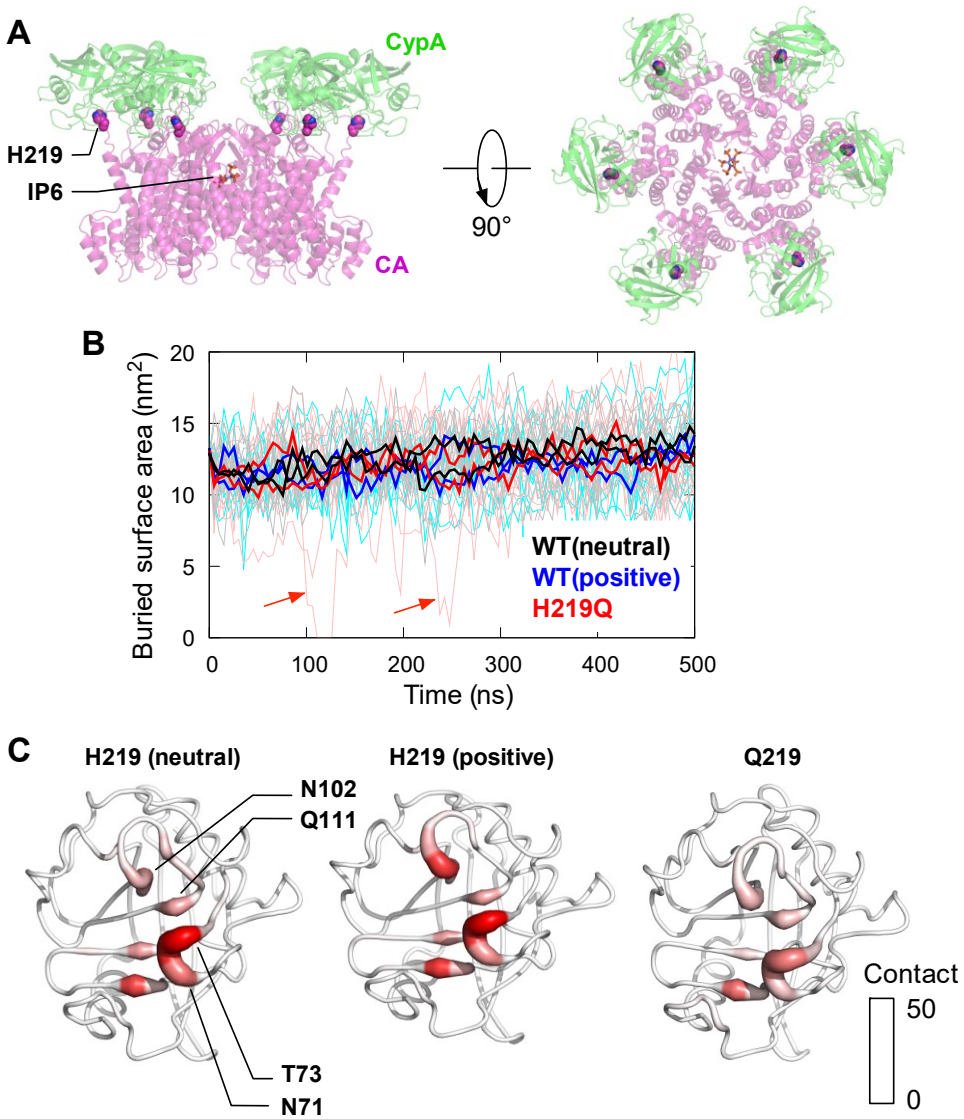

Figure 6: Atomistic simulations of mature CA with CypA. (A) Side and top views of mature CA hexamer (PDB: 6BHT) with each subunit bound to CypA (aligned using PDB: 1AK4). Residue H219 is shown in van der Waals representation, while IP6 bound in the center of the hexamer is shown in stick representation. (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.

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[14]. 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 (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, whilst protonated H199 could form a salt bridge with E319. Our simulations therefore suggest that a histidine residue at position 199 has the capability to interact with more residues in neighbouring subunits compared to WT, and thus potentially strengthen the binding interface between CA subunits to enhance the stability of the hexameric complex.

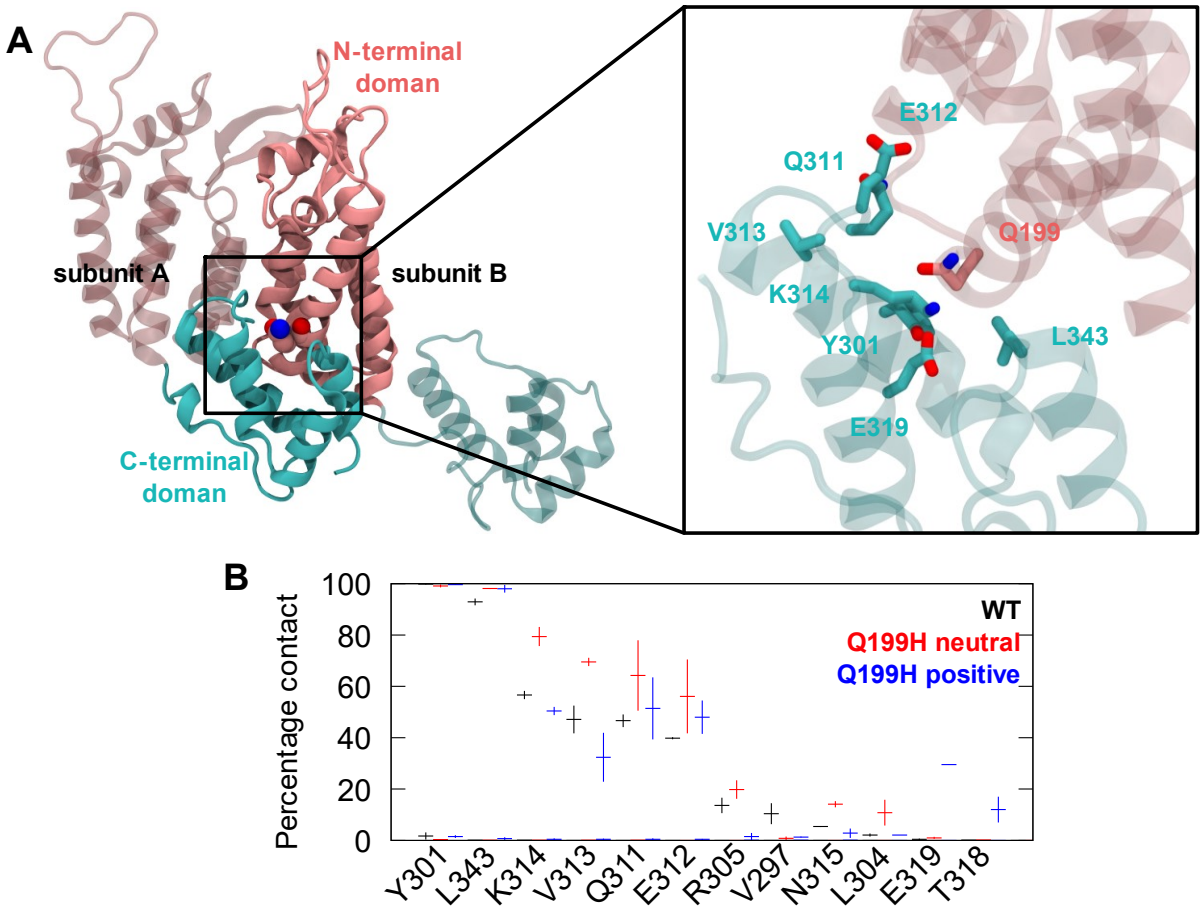

Figure 7: The role of the Q199H mutation in CA oligomerisation. (A) (Left) A representative snapshot of two subunits of the mature WT CA hexamer in cartoon representation. The Q199 residue bridging the N-terminal domain (pink) of one subunit and the C-terminal domain (cyan) of an adjacent subunit is 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.

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 connecting the MA and CA domains is flexible and may contract, enabling interactions between protease cleavage site residues and the C-terminal region of the MA domain,

as well as several regions on the CA domain including the CypA binding loop. Two clinical mutations, G123E and H219Q, are found in these regions and alter the accessibility and electrostatic properties of the protease binding site. Mutations in the MA domain are concentrated on the predicted membrane binding surface and at least three of them – E12K, E40K and L75R – render the surface more positively charged compared to wild type Gag. Our CG simulations of these mutant variants show more pronounced interactions with PIP2 lipids, leading to the hypothesis that these mutations improve Gag targeting and anchoring to the plasma membrane during viral assembly. Atomic-resolution simulations of the mature Gag CA domain showed that the H219Q mutation may play a role in optimizing CypA recruitment by controlling the conformational flexibility of the CypA binding loop and fine-tuning the CypA binding affinity. We also found that a mutation in the CA domain, Q199H, likely enhances oligomerisation by providing more stabilising interactions between adjacent subunits.

In the field of HIV and other related retroviruses, multiscale modelling and simulations have been utilised to study CA in different maturation states, providing important insights into their structural dynamics and interactions with drugs and host cell proteins[30,46–48]. For example, landmark microsecond-timescale atomic simulations of the entire HIV-1 capsid revealed important physicochemical properties, such as its electrostatics, dynamics, and water/ion permeability, as well as pointing towards long-range allosteric motions in the shell[27]. Simulations of the monomeric MA domain shed light on the molecular mechanism underlying membrane anchoring by the myristoyl group and PIP2 lipids[33]. Our study reports the first complete oligomeric model of the proteolytic product of HIV-1 Gag polyprotein, allowing the investigation of interactions between residues on the protease binding site and distant residues on the MA and CA domains, especially those from different subunits. Such inter-subunit “crosstalk” would be difficult to identify using individual crystal structures of mature Gag domains or a monomeric Gag model. Understanding structural changes that occur upon proteolysis could provide important clues on HIV-1 protease activity. For example, the MA-CA linker contraction observed in our simulations resulted in a significant decrease in solvent accessibility of the protease binding site, which helps to explain the lower cleavage rate of the MA-CA domain compared to the SP2-P6 domain[38]. From the perspective of anti-HIV drug development, targeting Gag proteolytic products in addition to the mature form, by virtue of identifying intermediate structural models, may allow for the discovery of Gag inhibitors that can inhibit multiple stages of maturation[10].

Since accumulation of deleterious mutations could negatively impact virion survival, Gag polyprotein and protease co-compensate to outcompete protease inhibitors and restore viral fitness[10,11,49]. While HIV-1 protease acquires drug resistant mutations to evade PI activity, its ability to bind and subsequently cleave the Gag protein is impaired. Thus, Gag gains additional mutations to compensate for fitness defects to recover protease function. Structural studies show that associated mutations on cleavage sites do not restore interactions lost due to mutations on the protease, but rather establish novel interactions surrounding the site of protease mutations as well as induce conformational changes to drive better binding[12]. Apart from cleavage site mutations, mutations in more distant regions have been shown to contribute towards recovery from the reduced viral fitness caused by protease mutations[13,14]. Our study shows that these non-cleavage site mutations have far reaching implications outside of Gag proteolysis, from improving membrane binding to fine-tuning CypA recruitment and stabilising the CA hexamer. A recent single genome sequencing study revealed that, indeed, most Gag-protease correlated mutations occur outside of the Gag cleavage site with the viral MA and CA domains representing the largest subsets of the mutations[49]. Interestingly, a high concentration of compensatory mutations was uncovered within the globular domain of the MA, which binds to the plasma membrane, and the CypA binding loop of the CA, further corroborating our simulation results. Our results also highlight that the residues on the CypA binding loop can interact directly with MA-CA cleavage site residues, thus potentially influencing their local electrostatics and surface accessibility, although determining the exact interactions made between these residues and subsequent molecular effects would require further studies. For example, PS lipids are known to be exposed on the outer leaflet of HIV-1 membrane and may facilitate HIV-1 entry into the host cells[50,51], and it is possible that the highly basic surface generated by the non-cleavage site mutations could attract anionic PS lipids, similarly to the PIP2 enrichment around the MA domain observed here. 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[14], 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-

519 cleavage site mutations on HIV-1 viral fitness. Given that HIV-1 is one of the fastest
520 mutating RNA viruses, an in-depth comprehension of emerging mutations will aid in
521 the general understanding of viral drug resistance, as well as the emergence of novel
522 viruses that are able to cross species during zoonotic infections.
523
524

Methods

Model building and system setup

The model of HIV-1 Gag MA-CA-SP1 was built using two structural templates: MA trimers (PDB: 1HIW)[34] and the immature CA hexamer (PDB: 5L93)[35]. Six copies of MA trimers were arranged based on EM images showing that they organize as hexamer-of-trimers on a model membrane bilayer[36]. The CA hexamer was placed such that the N-termini of its subunits align with the C-termini of the central subunits of the MA trimers. Modeller version 9.21[52] was used to build the loops connecting the MA trimers and CA hexamer (Figure S1) using the discreet optimized protein energy (DOPE)-based loop modelling protocol, and the model with the lowest DOPE score was chosen[53]. Stereochemical assessment using Ramachandran analysis[54] showed only one outlier residue, confirming that the model was structurally reasonable. The atomistic MA-CA-SP1 structure was subsequently converted to CG representation using the MARTINI 2.2 force field[55] and the elastic network model, ElNeDyn[56], was imposed to maintain the integrity of the secondary and tertiary structure. The N-terminal glycine residue on each MA subunit was myristoylated based on parameters from Charlier et al[33]. Three mutant variants were generated using PyMOL[57] based on non-cleavage site mutations published by Gatanaga et al[14].

A 30 x 30 nm^2 patch representing the HIV-1 membrane model was built using the CHARMM-GUI Martini Maker Bilayer Builder[58,59] and the lipid composition was based on a previous HIV-1 lipidomic study[37]. This membrane was asymmetric: the upper leaflet was made of 10% phosphatidyl choline (PC), 10% phosphatidyl ethanol amine (PE), 30% sphingomyelin, and 50% cholesterol, whereas the lower leaflet was made of 10% PC, 20% PE, 15% phosphatidyl serine (PS), 5% phosphatidylinositol-4,5-bisphosphate (PIP2), and 50% cholesterol. The HIV-1 membrane is enriched in saturated lipid species especially for PC; we therefore modelled the lipid tails of PC as 1,2-dipalmitoyl (DP), whereas the rest of the lipids were modelled with 1-palmitoyl-2-oleoyl (PO) lipid tails. Following minimization and equilibration procedures according to the CHARMM-GUI, the model membrane was further simulated for 1 μs to allow better mixing of the lipid components prior to simulations in the presence of protein.

Molecular dynamics simulations

The HIV-1 Gag MA-CA-SP1 CG model was placed underneath the membrane such that the myristoylated N-termini were inserted into lower leaflet of the membrane.

Energy minimization was then performed using the steepest descent method to remove any overlapping beads. The system was solvated with standard MARTINI water molecules and neutralized with 0.15 M NaCl ions. Further energy minimization was performed. The system was then equilibrated for 10 ns whereby positional restraints with a force constant of 1000 kJ mol^{-1} nm^{-2} were imposed on all of the backbone atoms of the protein. Temperature coupling using the V-rescale thermostat with a time constant of 1 ps[60] was applied to maintain the temperature at 310 K. Semi-isotropic pressure coupling using the Berendsen barostat with a time constant of 5 ps was utilized to maintain the pressure of the system at 1 atm. Electrostatics were calculated using the reaction field method. The der Waals interactions were computed using a potential shift Verlet scheme. The short-range cut-offs for both of these were set to 1.1 nm. After the equilibration simulations, four independent production runs of 15 μs were performed using different distributions of initial velocities. The same protocols used for the equilibration simulation were used for the production runs in the absence of position restraints, except for the pressure coupling in which the Parrinello-Rahman barostat was used with a coupling constant of 12.0 ps[61].

Atomistic MD simulations were performed to understand the role of the H219Q mutations on CypA binding and Q199H on CA oligomerisation. The structure of the mature HIV-1 CA hexamer in complex with inositol IP6 (PDB: 6BHT)[32] was aligned to that of the CA N-terminal domain bound to CypA (PDB: 1AK4)[43]. This alignment generated a structure of the CA hexamer bound to six copies of CypA (Figure 6A). Mutations in the CA were performed using PyMOL. The proteins were parameterized using the CHARMM36 force field[62], whereas the CHARMM-GUI Ligand Reader & Modeller[63] was used to generate the parameters for the IP6 molecule. The complex was solvated with TIP3P water molecules and 0.15 M NaCl was added to neutralize the system. The steepest descent minimization protocol was used to remove overlapping atoms. A short 100 ps equilibration simulation with position restraints (force constant of 1000 kJ mol^{-1} nm^{-2}) applied to all heavy atoms of the protein was conducted. The temperature of the system was maintained at 310 K using the Nosé-Hoover thermostat with a time constant of 1.0 ps[64,65]. The pressure was kept at 1 atm using an isotropic coupling to the Parrinello-Rahman barostat with a time constant of 5.0 ps[61]. Electrostatic interactions were measured using the smooth particle mesh Ewald (PME) method with a real-space cut off of 1.2 nm[66]. The van der Waals interactions were calculated using the force switch smoothing function applied

between 1.0 and 1.2 nm, and truncated at 1.2 nm. An integration time step of 2 fs was used with all covalent bonds involving hydrogens constrained using the LINCS algorithm[67]. After equilibration, the positional restraints were removed and two independent 500 ns production runs were conducted using the same setup for each of the WT and mutant variants starting with different initial velocities.

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

To refine interactions between the linker region and mutant residues observed in the CG simulations of the MA-CA-SP1 hexamer, we converted the structure of the hexamer from the final frame of one of the simulations to atomistic representation using the CHARMM-GUI All-Atom Converter[68]. 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[69]. 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[70].

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).

627 **References**

628 1. Yeo, J. Y., Goh, G. R., Tran-To Su, C. & Ken-En Gan, S. The determination of
629 HIV-1 RT mutation rate, its possible allosteric effects, and its implications on
630 drug resistance. *Viruses* **12**, 1–22 (2020).
631 2. Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in
632 viruses: Patterns and determinants. *Nat. Rev. Genet.* **9**, 267–276 (2008).
633 3. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science*
634 *(80-.).* **310**, 676–679 (2005).
635 4. Kuiken, T. *et al.* Host species barriers to influenza virus infections. *Science*
636 *(80-.).* **312**, 394–397 (2006).
637 5. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in
638 China. *Nature* (2020). doi:10.1038/s41586-020-2008-3
639 6. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of
640 probable bat origin. *Nature* **2019**, (2020).
641 7. Yi Yeo, J., Yap, P., Goh, G., Koh, D. W. & Gan, S. K. HIV-1 mutations in HIV-1
642 Gag , protease , RT p66 and when they appear : Insights from an in vitro BSL2
643 assay on mutation rates and types . *bioRxiv* 1–26 (2020).
644 8. Perelson, A. S. Modelling viral and immune system dynamics. *Nat. Rev.*
645 *Immunol.* **2**, 28–36 (2002).
646 9. Su, C. T. T., Ling, W. L., Lua, W. H., Haw, Y. X. & Gan, S. K. E. Structural
647 analyses of 2015-updated drug-resistant mutations in hiv-1 protease: An
648 implication of protease inhibitor cross-resistance. *BMC Bioinformatics* **17**,
649 (2016).
650 10. Su, C. T.-T., Koh, D. W.-S. & Gan, S. K.-E. Reviewing HIV-1 Gag Mutations in
651 Protease Inhibitors Resistance: Insights for Possible Novel Gag Inhibitor
652 Designs. *Molecules* **24**, 3243 (2019).
653 11. Clavel, F. & Mammano, F. Role of gag in HIV resistance to protease inhibitors.
654 *Viruses* **2**, 1411–1426 (2010).
655 12. Özen, A., Lin, K. H., Yilmaz, N. K. & Schiffer, C. A. Structural basis and distal
656 effects of Gag substrate coevolution in drug resistance to HIV-1 protease.
657 *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15993–15998 (2014).
658 13. Myint, L. *et al.* Gag Non-Cleavage Site Mutations Contribute to Full Recovery
659 of Viral Fitness in Protease Inhibitor-Resistant Human Immunodeficiency Virus
660 Type 1. *Antimicrob. Agents Chemother.* **48**, 444–452 (2004).

14. Gatanaga, H. *et al.* Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J. Biol. Chem.* **277**, 5952–5961 (2002).
15. Su, C. T. T., Kwoh, C. K., Verma, C. S. & Gan, S. K. E. Modeling the full length HIV-1 Gag polyprotein reveals the role of its p6 subunit in viral maturation and the effect of non-cleavage site mutations in protease drug resistance. *J. Biomol. Struct. Dyn.* **36**, 4366–4377 (2018).
16. Bell, N. M. & Lever, A. M. L. HIV Gag polyprotein: Processing and early viral particle assembly. *Trends Microbiol.* **21**, 136–144 (2013).
17. Franke, E. K., Yuan, H. E. H. & Luban, J. Specific incorporation of cyclophilin a into HIV-1 virions. *Nature* **372**, 359–362 (1994).
18. Braaten, D., Franke, E. K. & Luban, J. Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. *J. Virol.* **70**, 3551–60 (1996).
19. Carnes, S. K., Sheehan, J. H. & Aiken, C. Inhibitors of the HIV-1 capsid, a target of opportunity. *Curr. Opin. HIV AIDS* **13**, 359–365 (2018).
20. Purdy, M. D. *et al.* MicroED structures of HIV-1 Gag CTD-SP1 reveal binding interactions with the maturation inhibitor bevirimat. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 13258–13263 (2018).
21. Seclén, E. *et al.* High prevalence of natural polymorphisms in Gag (CA-SP1) associated with reduced response to Bevirimat, an HIV-1 maturation inhibitor. *AIDS* **24**, 467–469 (2010).
22. Verheyen, J. *et al.* High prevalence of bevirimat resistance mutations in protease inhibitor-resistant HIV isolates. *Aids* **24**, 669–673 (2010).
23. Marzinek, J. K., Huber, R. G. & Bond, P. J. Multiscale modelling and simulation of viruses. *Current Opinion in Structural Biology* (2020). doi:10.1016/j.sbi.2019.12.019
24. Ayton, G. S., Noid, W. G. & Voth, G. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol.* **17**, 192–8 (2007).
25. Grime, J. M. A. & Voth, G. A. Early stages of the HIV-1 capsid protein lattice formation. *Biophys. J.* **103**, 1774–1783 (2012).
26. Grime, J. M. A. *et al.* Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nat. Commun.* **7**, 11568 (2016).
27. Perilla, J. R. & Schulten, K. Physical properties of the HIV-1 capsid from all-

atom molecular dynamics simulations. *Nat. Commun.* **8**, (2017).

28. Pak, A. J. *et al.* Immature HIV-1 lattice assembly dynamics are regulated by scaffolding from nucleic acid and the plasma membrane. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10056–E10065 (2017).
29. Pak, A. J., Grime, J. M. A., Yu, A. & Voth, G. A. Off-Pathway Assembly: A Broad-Spectrum Mechanism of Action for Drugs That Undermine Controlled HIV-1 Viral Capsid Formation. *J. Am. Chem. Soc.* **141**, 10214–10224 (2019).
30. Liu, C. *et al.* Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat. Commun.* **7**, (2016).
31. Huang, P.-T. *et al.* FEZ1 Is Recruited to a Conserved Cofactor Site on Capsid to Promote HIV-1 Trafficking. *Cell Rep.* **28**, 2373-2385.e7 (2019).
32. Dick, robert A. *et al.* Inositol phosphates are assembly co-factors for HIV-1. *Nature* **560**, 509–512 (2018).
33. Charlier, L. *et al.* Coarse-grained simulations of the HIV-1 matrix protein anchoring: Revisiting its assembly on membrane domains. *Biophys. J.* **106**, 577–585 (2014).
34. Hill, C. P., Worthylake, D., Bancroft, D. P., Christensen, A. M. & Sundquist, W. I. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 3099–104 (1996).
35. Schur, F. K. M. *et al.* An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science (80-.).* **353**, 506–508 (2016).
36. Alfadhli, A., Barklis, R. L. & Barklis, E. HIV-1 matrix organizes as a hexamer of trimers on membranes containing phosphatidylinositol-(4,5)-bisphosphate. *Virology* **387**, 466–472 (2009).
37. Brugger, B. *et al.* The HIV lipidome: A raft with an unusual composition. *Proc. Natl. Acad. Sci.* **103**, 2641–2646 (2006).
38. Pettit, S. C. *et al.* The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.* **68**, 8017–27 (1994).
39. Ono, A., Ablan, S. D., Lockett, S. J., Nagashima, K. & Freed, E. O. Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 14889–14894 (2004).
40. Saad, J. S. *et al.* Structural basis for targeting HIV-1 Gag proteins to the

plasma membrane for virus assembly. *Proc. Natl. Acad. Sci.* **101**, 14889–14894 (2006).

41. Ono, A. & Freed, E. O. Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 13925–13930 (2001).
42. Thali, M. *et al.* Functional association of cyclophilin a with HIV-1 virions. *Nature* **372**, 363–365 (1994).
43. Gamble, T. R. *et al.* Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87**, 1285–1294 (1996).
44. Yin, L., Braaten, D. & Luban, J. Human immunodeficiency virus type 1 replication is modulated by host cyclophilin A expression levels. *J. Virol.* **72**, 6430–6 (1998).
45. Yezid, H., Konate, K., Debaisieux, S., Bonhoure, A. & Beaumelle, B. Mechanism for HIV-1 tat insertion into the endosome membrane. *J. Biol. Chem.* **284**, 22736–22746 (2009).
46. Zhao, G. *et al.* Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643–646 (2013).
47. Ayton, G. S. & Voth, G. A. Multiscale computer simulation of the immature HIV-1 virion. *Biophys. J.* **99**, 2757–2765 (2010).
48. Goh, B. C. *et al.* Atomic Modeling of an Immature Retroviral Lattice Using Molecular Dynamics and Mutagenesis. *Structure* **23**, 1414–1425 (2015).
49. Codoñer, F. M. *et al.* Gag-protease coevolution analyses define novel structural surfaces in the HIV-1 matrix and capsid involved in resistance to Protease Inhibitors. *Sci. Rep.* **7**, (2017).
50. Callahan, M. K. *et al.* Phosphatidylserine on HIV Envelope Is a Cofactor for Infection of Monocytic Cells. *J. Immunol.* **170**, 4840–4845 (2003).
51. Zaitseva, E. *et al.* Fusion Stage of HIV-1 Entry Depends on Virus-Induced Cell Surface Exposure of Phosphatidylserine. *Cell Host Microbe* **22**, 99-110.e7 (2017).
52. Sali, A. & Blundell, T. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1994).
53. Shen, M., Devos, D., Melo, F. & Sali, A. A composite score for predicting errors in protein structure models. *Protein Sci.* **15**, 1653–1666 (2006).
54. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry

of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).

55. Monticelli, L. *et al.* The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008).
56. 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).
57. DeLano, W. The PyMOL molecular graphics system. (2002).
58. 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).
59. 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).
60. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
61. Parrinello, M. & Rahman, A. Polymorphic Transitions in Single Crystals: a New Molecular Dynamics Method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
62. 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).
63. 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).
64. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
65. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
66. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577 (1995).
67. 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).
68. Wassenaar, T. A., Pluhackova, K., Böckmann, R. A., Marrink, S. J. & Tieleman, D. P. Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J. Chem. Theory*

Comput. **10**, 676–690 (2014).

69. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

70. 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).