1   **HIV-1 capsids possess dynamic pores that import nucleotides with kinetic**
2   **perfection**
3
4   **During the early stages of infection, the HIV-1 capsid protects viral components from cytosolic**
5   **sensors, such as cGAS, and nucleases, such as TREX, while allowing access to nucleotides for**
6   **efficient reverse transcription[1]. Here we show that each capsid hexamer has a size-selective pore**
7   **bounded by a ring of six arginine residues and a 'molecular iris' formed by the N-terminal β-**
8   **hairpin. The arginine ring creates a strongly positively charged channel that recruits the four**
9   **nucleotides with on-rates that near diffusion limits. Progressive removal of pore arginines results**
10 **in a dose-dependent and concomitant decrease in nucleotide affinity, reverse transcription and**
11 **infectivity. This positively charged channel is universally conserved in lentiviral capsids despite**
12 **the fact that it is strongly destabilising without nucleotides to counteract charge repulsion. We**
13 **also describe a channel inhibitor, hexacarboxybenzene, which competes for nucleotide binding**
14 **and efficiently blocks encapsidated reverse transcription demonstrating the tractability of the**
15 **pore as a novel drug target.**
16
17 There is increasing evidence that the HIV-1 capsid remains intact as it traverses the cytoplasm of a
18 newly-infected cell. Prematurely-uncoated viruses trigger innate immune sensing[2], assembled capsid
19 proteins are required to properly engage the nuclear pore complex[3], and intact capsids have been
20 observed at the nuclear envelope[4]. Reverse transcription has been postulated to occur within the HIV-1
21 virion during cytoplasmic transit, yet structural analyses of the HIV-1 capsid have not defined a pore
22 through which small molecules such as dNTP's might pass. One possible location for a pore would be
23 the 6-fold axis at the centre of each capsid protein (CA) hexamer but this is not evident from existing
24 hexamer structures as it is obscured by the N-terminal β-hairpin. By comparing all the available CA
25 crystal structures with resolved β-hairpins[5-11], including monomeric crystal forms, we observed that the
26 β-hairpin can adopt alternate conformations that differ by up to 15 Å (as measured by the displacement
27 of Q7) (Fig. 1a).   When reconstructed in the context of a hexamer, several of these β-hairpin
28 conformations result in a pore about the six-fold axis (Fig. 1b). The different β-hairpin conformations
29 are the result of a pivoting movement of up to 37.5° about the N-terminal proline, an essential capsid
30 residue that forms a salt-bridge with D51[12] (Fig. 1a, Supplementary Video 1). In structures where the
31 pore would be open, D51 also participates in a second salt-bridge interaction with H12. Conversely, in
32 structures where the pore would be closed, including all previously solved disulfide-stabilised
33 hexamers (CA$_{Hexamer}$), a water molecule has displaced the H12 side-chain and coordinates a tetrahedral
34 hydrogen-bond network between H12, T48, Q50 and D51. We hypothesized that the protonation state
35 of H12 may be crucial in determining which arrangement is favored and therefore that the
36 conformation of the β-hairpin in published structures will have been influenced by the pH at which
37 they were solved. Remarkably, when the relative displacement of the β-hairpin (Q7 Cα) is plotted
38 against crystallization pH the structures resolve into two groups; at pH < 7 an open pore β-hairpin
39 conformation is observed whereas at pH > 7 a closed pore conformation is favored (Fig. 1c, Extended
40 Data Table 1). The same correlation is observed when the distance between D51 and H12 is plotted

41 against pH, confirming the importance of H12 in determining β-hairpin conformation. Structures

42 solved at pH 7 display the greatest β-hairpin variability, consistent with maximum pore flexibility

43 occurring under physiological conditions. The likely reason why a pore has not been detected in

44 published $CA_{Hexamer}$ structures is because they were solved at a basic pH where H12 is deprotonated

45 and a closed pore is favored. To test this hypothesis and demonstrate that the pore can open in the

46 context of an assembled hexamer, we sought to crystallise $CA_{Hexamer}$ under acidic conditions. We

47 obtained a previously unreported crystal form at pH 5.5, the structure of which contains a β-hairpin in

48 the open conformation and an exposed pore (Fig. 1b).

49

50 Using all available CA structures to define a range of movement for the β-hairpin, we observed that the

51 pivoting about P1 results in an iris-like motion, which creates an aperture on the outer surface of the

52 capsid (Fig. 2a). In the open state, a chamber 25 Å deep and 3240 $Å^3$ in volume is revealed, which

53 culminates in a ring of six arginine side-chains from residue 18 (Fig. 2b, Supplementary Video 2). This

54 cluster of basic residues in close proximity results in highly electropositive foci at the centre of each

55 hexamer. The R18 residues adopt multiple conformations (Extended Data Fig. 1a) to give a maximum

56 pore diameter of 8 Å, sufficient to allow transit of a dNTP molecule. We therefore reasoned that this

57 feature might provide an efficient means to recruit dNTP's into the capsid interior (Fig. 2a) whilst

58 excluding larger molecules. We therefore tested whether $CA_{Hexamer}$ can interact with dNTP's by

59 fluorescence anisotropy and found that all four nucleotides bind with remarkably high affinity of

60 between 6-40 nM (Fig. 2c). All biophysical measurements were undertaken in an 'intracellular buffer'

61 (see Methods) that is designed to match salt concentrations in the cell. We also observed that

62 physiological concentrations of inorganic phosphate had little effect on dNTP binding and that the pore

63 could not distinguish between dNTP's and rNTP's (Extended Data Fig. 2) consistent with the

64 observation that rNMP's are often incorporated into newly-synthesised viral DNA[13]. Analyzing the

65 kinetics of interaction by stopped-flow revealed that binding is driven by an extremely rapid on-rate of

66 $> 2x10^8$ M^{-1} s^{-1}, although this is likely an underestimate as the reaction becomes immeasurably fast at

67 increasing reactant concentrations (Fig. 2d&e). Separate dissociation experiments in which fluorescent

68 dCTP was displaced with excess unlabeled dCTP determined that the off-rate is also fast at >12 s^{-1}

69 (Fig. 2f), equivalent to a half life of 58 ms. Calculation of on-rates for all four nucleotides based on

70 their steady-state affinities and off-rates confirms that HIV-1 hexamers achieve association rates

71 between 10^8-10^9 M^{-1} s^{-1}. These are unusually rapid association kinetics, typically found in enzymes that

72 have achieved so-called kinetic perfection. Such ultra-rapid enzymes are rare because of the strong

73 fitness advantage needed for their selection over merely very fast equivalents[14]. The rapid on-rate of

74 dNTP recruitment that HIV achieves may be the result of an electrostatically assisted association

75 binding mechanism, as has been described for barnase/barstar[15]. Importantly, the combination of fast

76 on and off rates suggests that while the HIV-1 capsid may recruit dNTP's extremely efficiently, these

77 nucleotides quickly dissociate to become available as substrates for reverse transcription.

78

79 To test our hypothesis that it is the ring of arginine residues that is responsible for nucleotide

80 recruitment, we solved the structure of HIV-1 $CA_{Hexamer}$ in complex with dATP and found that it binds

81 as predicted in the center of the arginine ring via its phosphate groups (Fig. 2g). While there is electron
82 density for the phosphates, the position of the base can only be modeled, likely because hexamer
83 rotational symmetry allows the dATP base to occupy six equivalent positions, averaging its density
84 over a large volume (Extended Data Fig. 1). Comparison with a structure solved under identical
85 conditions, but in the absence of dATP, confirms that the observed density corresponds to the
86 nucleotide. To further investigate the importance of R18 in the recruitment of dNTP's, we produced
87 R18G and R18A $CA_{Hexamer}$ mutants. Neither R18G nor R18A affected the overall structure of the
88 protein and neither displayed measurable nucleotide binding (Fig. 3a&b). To determine whether
89 formation of the arginines into a ring is required, we performed binding experiments on wild-type
90 protein in the presence of DTT, which reduces the disulphide bonds that stabilise the hexameric
91 construct, resulting in monomeric CA. No binding was observed to monomeric CA demonstrating that
92 once the pore is disassembled, the capsid can no longer recruit dNTP's (Fig. 3b).
93
94 The concentration of positive charge provided by the R18 ring is an unusual feature and might be
95 expected to exert a destabilizing influence on the capsid lattice. Conversely, it has been calculated that
96 arginine pairs can stabilize protein interfaces[16], while arginine clusters have been postulated to have a
97 stabilizing effect[17]. We performed differential scanning fluorimetry (DSF) to compare the relative
98 stability of $CA_{Hexamer}$ with and without the electropositive pore. We observed a remarkable stability of
99 the R18G hexamer relative to wild-type complex, corresponding to a surprisingly large increase in T_m
100 of 4 °C (Fig. 3c, Extended Data Fig. 3). A similar increase in stability was observed in wild-type
101 hexamer in the presence of dNTP's, while no stabilization was observed when dNTP's were added to
102 R18G hexamer (Fig. 3d). Taken together, these results suggest that the pore is indeed a destabilising
103 feature that is tolerated by the capsid lattice in order to facilitate nucleotide binding. An alignment of
104 capsid sequences predicts an electropositive pore to be conserved across retrovirus genera, with the
105 exception of the gammaretroviruses (Extended Data Fig. 4). While there appears to be no R18-
106 equivalent in gammaretroviruses, analysis of the published MLV capsid structure reveals a large
107 channel running down the six-fold axis with an inward facing Arg residue at position 3 (Mortuza *et al.*
108 2008).   This residue may have a role in attracting dNTP's, but it is unlikely that the MLV capsid has
109 the same size-selectivity as the HIV capsid as the pore is much wider.
110
111 The observation that HIV-1 has evolved the fastest possible rate constant for nucleotide recruitment
112 suggests that dNTP import may be a limiting factor in reverse transcription and infectivity. To
113 determine how the efficiency of nucleotide recruitment impacts on these measures of viral fitness, we
114 constructed a matched set of chimeric WT:R18G $CA_{Hexamer}$ and viruses (see Extended Data Fig. 5 for
115 chimera controls). R18G was chosen as its capsid morphology has previously been demonstrated to be
116 indistinguishable from WT[19]. Furthermore, R18G was able to saturate the activity of capsid-binding
117 restriction factor TRIM5α, confirming that assembled capsids enter the cytoplasm[20] (Extended Data
118 Fig. 6). We tested our chimeric hexamers for dNTP binding and found that an incorporation ratio of 5:1
119 (5 arginines to 1 glycine) had a minimal impact on affinity (Fig. 3e). However, as the proportion of
120 glycine residues was increased, there was a dose-dependent decrease in dNTP binding. Testing HIV-1

121 GFP VSV-G pseudotyped chimeric viruses for infectivity revealed a similar pattern of R18
122 dependence, in which there was little change in infectivity at a ratio of 5:1 but a dominant negative
123 effect at higher G18 ratios (Fig. 3f). Assuming a binomial distribution of arginines and glycines in viral
124 hexamers, the data fit a model in which removal of two or more arginines from the pore is detrimental
125 to the virus (Extended Data Fig. 7), which is consistent with the observation that removal of one
126 arginine has little impact on dNTP affinity. Importantly, there is close correlation between chimera
127 infectivity and nucleotide affinity, consistent with the recruitment of dNTP's impacting directly on
128 viral infection (Fig. 3g). Such a mechanism would be expected to influence infection at the level of
129 reverse transcription and indeed a similarly close correlation is observed between chimera affinity and
130 the production of early reverse transcripts (Fig 3h).

131

132 We propose that reverse transcription takes place within the protected environment of the capsid by
133 recruiting nucleotides through a strongly electropositive pore at the center of each capsid hexamer. In
134 order to explore this further we performed endogenous reverse transcription (ERT) assays in which
135 HIV-1 capsid cores were purified from virions[21] and their reverse transcriptase activity quantified *in*
136 *vitro* (Extended Data Fig. 8). Efficient strong-stop reverse transcription (the first DNA synthesis step)
137 was observed upon incubation of cores with dNTP's. Moreover, addition of DNase I, RNase A or the
138 promiscuous nuclease Benzonase failed to prevent encapsidated reverse transcription (Fig. 4a,
139 Extended Data Fig. 8e). This demonstrates that dNTP's can access the interior of the capsid but larger
140 nucleases cannot, supporting the notion of a size selective pore. Processivity beyond strong-stop was
141 observed but at lower efficiency, in agreement with published data[22].

142

143 If the R18 pore is responsible for dNTP import, it is conceivable that capsid mutations that affect the
144 movement of the β-hairpin may also affect the efficiency of reverse transcription. Of the residues
145 primarily responsible for the hairpin movement (Fig 1a), P1 and D51 are invariant, with mutations at
146 these positions resulting in non-infectious particles due to defective capsid assembly[12]. His12 is also
147 highly conserved, however, in ~2% of sequences it has been replaced by a tyrosine. As tyrosine is not
148 titratable over physiological pH's, we hypothesised that the H12Y mutation would result in the β-
149 hairpin favouring one conformation. Solving the crystal structure of this mutant revealed that under the
150 high-pH condition, Y12 displaces the bound water molecule and makes a hydrogen-bond contact with
151 Asp51 (Fig. 4b). Despite contacting D51 directly, the larger sidechain of Y12 relative to H12 causes
152 the β-hairpin to favour the closed conformation (Fig. 4c). Importantly H12Y does not completely shut
153 the pore because residues 4-9 do not occupy a single defined state (Extended Data Fig. 9). The β-
154 hairpin therefore retains a degree of flexibility despite rigidification about the P1-D51 'hinge'.
155 Nevertheless we predicted that favouring the closed conformation should result in H12Y having
156 reduced RT kinetics while retaining some infectivity and this is what we observe (Fig. 4d).

157

158 To provide further evidence that nucleotides are being recruited through the R18 pore to allow ERT,
159 we sought a small molecule inhibitor that would block the pore. Small polyanionic compounds have
160 been used previously to block analogous arginine-rich pores[23]. We found that the hexacarboxybenzene

161 series (which are polyanionic at physiological pH) bound to $CA_{Hexamer}$ and competed for nucleotide
162 binding as measured by fluorescence anisotropy and DSF, respectively (Fig. 4e&f, Extended Data Fig.
163 3). Activity broadly increased with the number of negative charges present within the compound, with
164 hexa- or pentacarboxybenzene being the most effective. DSF indicated that the compounds did not
165 bind in the absence of R18 or when hexamers were reduced by DTT, and the crystal structure of the
166 hexacarboxybenzene-bound $CA_{Hexamer}$ confirmed that the compound was co-ordinated by R18 within
167 the central pore (Fig. 4f&g). At sufficiently high concentration, tetra-, penta- and hexacarboxybenzene
168 fully inhibit reverse transcriptase, presumably by competing with dNTP's (Extended Data Fig. 8f).
169 However, in ERT assays, when reverse transcriptase is enclosed within an intact viral capsid,
170 tetracarboxybenzene has no effect on reverse transcription, with only a small effect observed for
171 pentacarboxybenzene (Fig. 4h). In contrast, hexacarboxybenzene inhibited ERT almost completely.
172 The failure of tetracarboxybenzene to inhibit ERT demonstrates that a compound sufficiently small to
173 pass through the channel is still efficiently excluded from the capsid interior if it cannot bind the pore.
174 This result emphasises the chemical selectivity of the pore and its role in dNTP import during reverse
175 transcription.

176

177 As a semi-permeable reaction chamber, the HIV-1 capsid is reminiscent of bacterial
178 microcompartments – primitive 'organelles' that utilize a protein coat to isolate toxic reaction
179 intermediates from the cytoplasm[24]. Microcompartments import substrates through a size-selective
180 pore to be consumed by enzymes located inside a chamber[25]. Similarly, dNTP's translocated inside the
181 HIV capsid will be hydrolysed by encapsidated reverse transcriptase. Coupling import with hydrolysis
182 may create a local chemical gradient, promoting interior movement despite the release of captured
183 dNTP's on either side of the pore. Appositely, a subset of microcompartments, carboxysomes, contain
184 positively charged amino acids that are thought to selectively transport bicarbonate and ribulose-1,5-
185 bisphosphate over uncharged CO_2 and O_2[26]. Some microcompartment structures have gated channels
186 located at the six-fold axes in their protein lattice, to control substrate entry and product release. The
187 fact that a similar 'gate' potentially exists in the HIV-1 capsid, provided by the 'molecular iris' of the
188 β-hairpin, suggests that the virus could use this as a mechanism to regulate reverse transcription. In
189 addition, the regulation of capsid stability through dNTP recruitment, and possibly DNA synthesis,
190 provides a model whereby HIV-1 may co-regulate DNA synthesis and uncoating to facilitate
191 cytoplasmic DNA synthesis that remains invisible to cytoplasmic DNA sensing. Finally, the high
192 degree of conservation of R18 coupled with the fact that the pore can be obstructed chemically
193 identifies the pore as a novel target for drug development.

194
195
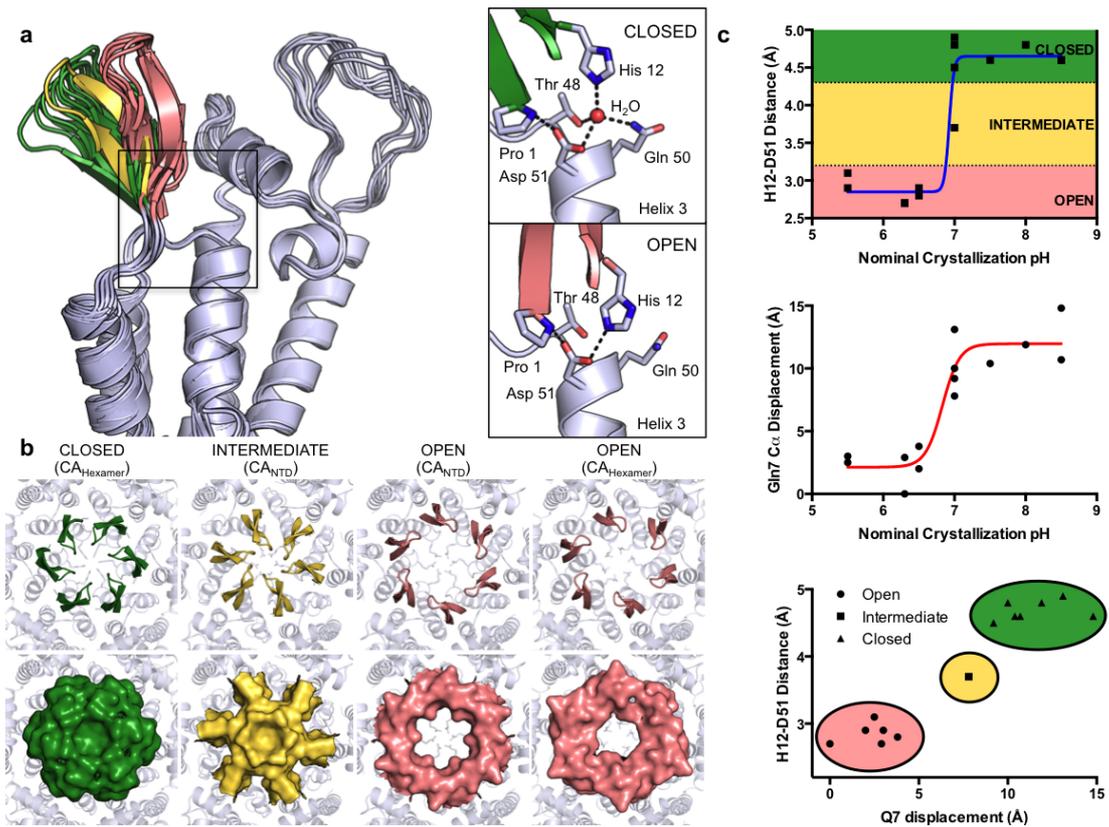
Figures

198 **Figure 1: HIV-1 capsid hexamers have a pore at the center of their 6-fold symmetry axis. a,**
199 Superposition of N-terminal domains from solved capsid structures (PDB IDs: 1AK4, 2GON, 2X83,
200 3H4E, 3NTE, 3P05, 4B4N and previously unpublished entries). A detailed view of the boxed region
201 shows that the β-hairpin toggles between closed (green) and open (pink) states as a result of the
202 hydrogen-bond network about P1, H12, and D51. **b**, β-hairpin (coloured) conformations dictate the
203 presence of a pore at the 6-fold axis. Hexamers of CA N-terminal domain (CA_{NTD}) structures have
204 been assembled using symmetry operators from $CA_{Hexamer}$ structures. **c**, Analysis of the salt-bridge
205 between H12 and D51 and the movement of the β-hairpin as a function of crystallization pH. Acidic
206 crystallization conditions promote an open β-hairpin conformation whilst basic conditions favor the
207 closed state.

208
209
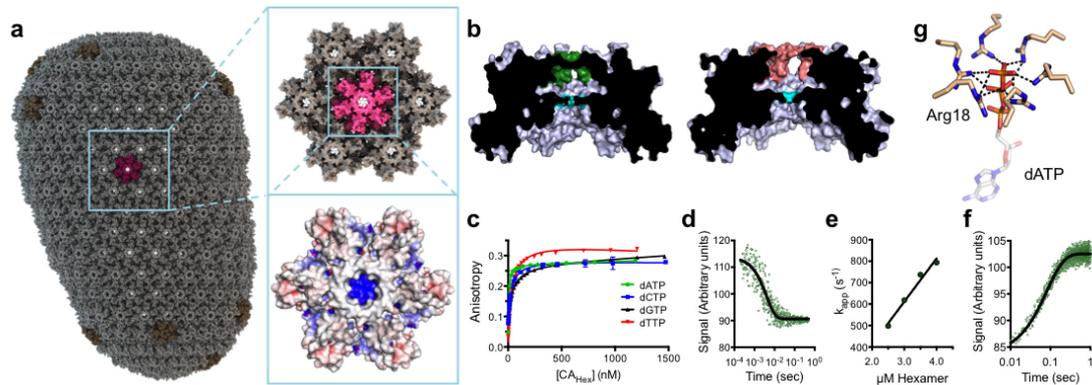
Figure 2: The HIV-1 capsid pore is strongly electropositive and recruits dNTP's with rapid association and dissociation kinetics. a, Model of an HIV-1 virion with hexamers in an open conformation reveals that the capsid is porous. Mapping the surface electrostatic potential shows that the pores are highly electropositive. **b**, Cross sections through the closed (β-hairpin green) and open (β-hairpin pink) CA$_{Hexamer}$ showing a central chamber that is accessible in the open state. R18 (cyan) creates a bottleneck at the base of the chamber underneath the β-hairpin. **c**, Fluorescence anisotropy measurements of dNTP's binding to CA$_{Hexamer}$. **d**, Example of pre-steady state association kinetics of dCTP with CA$_{Hexamer}$. **e**, Apparent rate constant (k$_{app}$) at increasing CA$_{Hexamer}$ concentrations. **f**, Dissociation of unlabeled dCTP:CA$_{Hexamer}$ by excess fluorescent-dCTP. **g**, R18 co-ordinates the phosphates in a dATP-bound CA$_{Hexamer}$ structure.
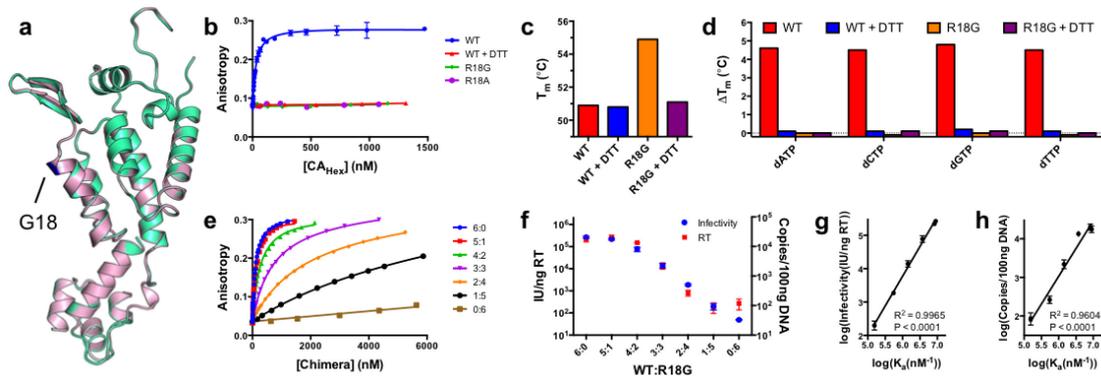
223

Figure 3: Conserved capsid residue R18 is crucial for nucleotide recruitment, reverse transcription and infectivity. a, Superposed single monomers of R18G (light-pink) and wild-type (light-green) $CA_{Hexamer}$. **b**, Binding of capsid variants to dCTP as measured by fluorescence anisotropy. **c**, DSF stability measurements expressed as T_m for WT and R18G ± DTT. **d**, DSF measurements of the effect of dNTP's on the stability of WT and R18G expressed as ΔT_m relative to unbound. **e**, Fluorescence anisotropy titrations of dTTP-binding by chimeric $CA_{Hexamers}$ with different R:G ratios at position 18. **f**, Comparison of infectivity and reverse transcription of chimeric viruses. **g**,**h**, Correlation between HIV-1 capsid dTTP affinity, viral infectivity **g** and reverse transcription **h**.

232

233

234
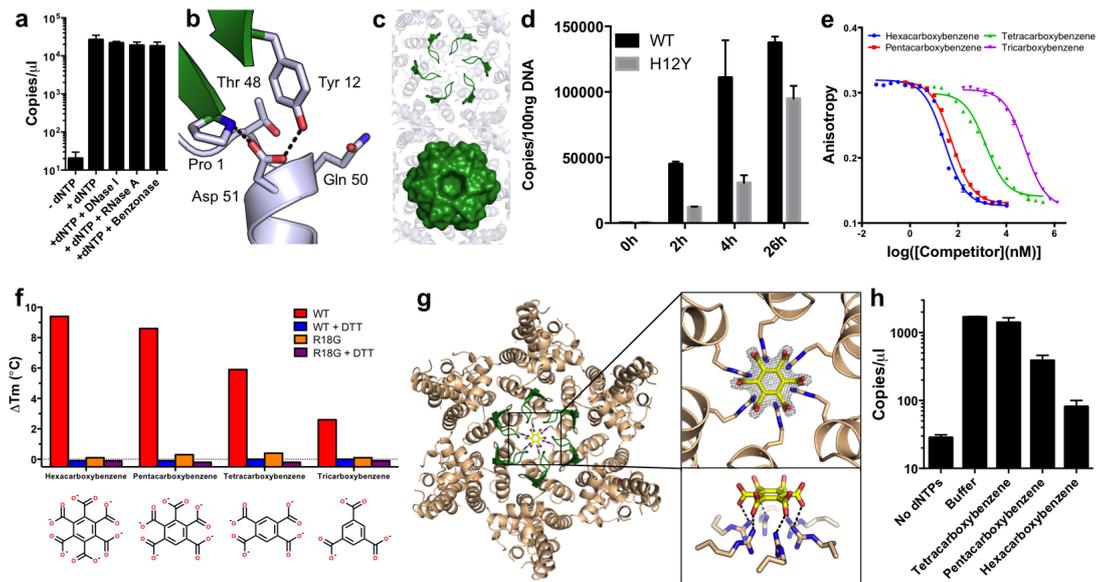
235   **Figure 4: HIV-1 reverse transcription is inhibited by blockade of the capsid pore a**, In vitro
236   endogenous reverse transcription measuring strong-stop transcripts. **b**, Residues surrounding Y12 in
237   the H12Y hexamer structure. **c**, Cartoon and surface representations of the β-hairpin in the H12Y
238   hexamer. **d**, WT and H12Y reverse transcription kinetics. **e**, Competition binding of carboxybenzene
239   compounds to $CA_{Hexamer}$. **f**, Change in wild type and R18G $CA_{Hexamer}$ T_m as measured by DSF in the
240   presence of carboxybenzene compounds. **g**, $CA_{Hexamer}$ crystal structure in complex with
241   hexacarboxybenzene, which is co-ordinated by R18. **h**, Effect of carboxybenzene compounds on
242   endogenous reverse transcription.

243
244

References

1   Campbell, E. M. & Hope, T. J. HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat Rev Microbiol* **13**, 471-483 (2015).

2   Rasaiyaah, J. *et al.* HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature* **503**, 402-405 (2013).

3   Price, A. J. *et al.* Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly. *PLoS pathogens* **10**, e1004459 (2014).

4   Arhel, N. J. *et al.* HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. *Embo J* **26**, 3025-3037 (2007).

5   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).

6   Kelly, B. N. *et al.* Implications for viral capsid assembly from crystal structures of HIV-1 Gag(1-278) and CA(133-278)(N). *Biochemistry-Us* **45**, 11257-11266 (2006).

7   Ylinen, L. M. J. *et al.* Conformational Adaptation of Asian Macaque TRIMCyp Directs Lineage Specific Antiviral Activity. *PLoS pathogens* **6** (2010).

8   Pornillos, O. *et al.* X-Ray Structures of the Hexameric Building Block of the HIV Capsid. *Cell* **137**, 1282-1292, doi:10.1016/j.cell.2009.04.0634 (2009).

9   Du, S. C. *et al.* Structure of the HIV-1 Full-Length Capsid Protein in a Conformationally Trapped Unassembled State Induced by Small-Molecule Binding. *Journal of molecular biology* **406**, 371-386 (2011).

10   Pornillos, O., Ganser-Pornillos, B. K. & Yeager, M. Atomic-level modelling of the HIV capsid. *Nature* **469** (2011).

11   Price, A. J. *et al.* CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. *PLoS pathogens* **8**, e1002896 (2012).

12   von Schwedler, U. K. *et al.* Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *Embo J* **17**, 1555-1568 (1998).

13   Kennedy, E. M., Amie, S. M., Bambara, R. A. & Kim, B. Frequent incorporation of ribonucleotides during HIV-1 reverse transcription and their attenuated repair in macrophages. *The Journal of biological chemistry* **287**, 14280-14288 (2012).

14   Bar-Even, A., Milo, R., Noor, E. & Tawfik, D. S. The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness. *Biochemistry-Us* **54**, 4969-4977 (2015).

15   Schreiber, G. & Fersht, A. R. Rapid, electrostatically assisted association of proteins. *Nature structural biology* **3**, 427-431 (1996).

16   Magalhaes, A., Maigret, B., Hoflack, J., Gomes, J. N. & Scheraga, H. A. Contribution of unusual arginine-arginine short-range interactions to stabilization and recognition in proteins. *Journal of protein chemistry* **13**, 195-215 (1994).

17   Neves, M. A., Yeager, M. & Abagyan, R. Unusual arginine formations in protein function and assembly: rings, strings, and stacks. *The journal of physical chemistry. B* **116**, 7006-7013 (2012).

18   Mortuza, G. B. *et al.* Structure of B-MLV capsid amino-terminal domain reveals key features of viral tropism, gag assembly and core formation. *Journal of molecular biology* **376**, 1493-1508 (2008).

19   Rihn, S. J. *et al.* Extreme genetic fragility of the HIV-1 capsid. *PLoS pathogens* **9**, e1003461 (2013).

20   Keckesova, Z., Ylinen, L. M. & Towers, G. J. The human and African green monkey TRIM5alpha genes encode Ref1 and Lv1 retroviral restriction factor activities. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10780-10785 (2004).

21   Shah, V. B. & Aiken, C. In vitro uncoating of HIV-1 cores. *Journal of visualized experiments : JoVE*, doi:10.3791/3384 (2011).

22   Warrilow, D., Warren, K. & Harrich, D. Strand transfer and elongation of HIV-1 reverse transcription is facilitated by cell factors in vitro. *PLoS one* **5**, e13229 (2010).

23   Cheley, S., Gu, L. Q. & Bayley, H. Stochastic sensing of nanomolar inositol 1,4,5-trisphosphate with an engineered pore. *Chemistry & biology* **9**, 829-838 (2002).

24   Tanaka, S., Sawaya, M. R. & Yeates, T. O. Structure and mechanisms of a protein-based organelle in Escherichia coli. *Science* **327**, 81-84 (2010).

25   Chowdhury, C. *et al.* Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2990-2995 (2015).

304   26   Kerfeld, C. A. *et al.* Protein structures forming the shell of primitive bacterial organelles.
305   *Science* **309**, 936-938 (2005).
306
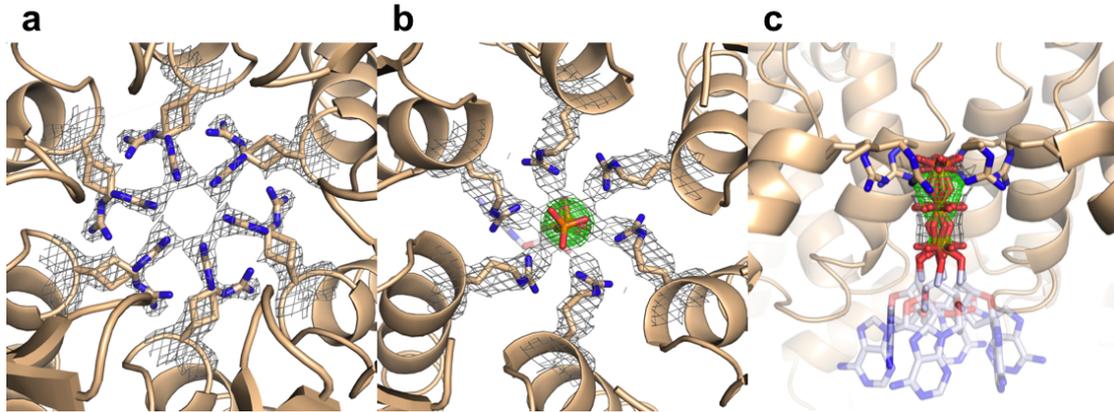
307

308 **Extended Data Figure 1: dATP binds to the R18 pore at the centre of the capsid hexamer.** 2Fo-

309 Fc density (grey mesh) contoured at 1.0σ about R18 for the unbound (**a**) and dATP-bound (**b**)

310 $CA_{Hexamer}$ structures. Fo-Fc omit density (green mesh) contoured at 3.0σ is shown for the dATP-bound

311 structure. **c**, dATP lies on the crystallographic 6-fold axis and significant rotationally-averaged density

312 is observed only for the triphosphate group.

313
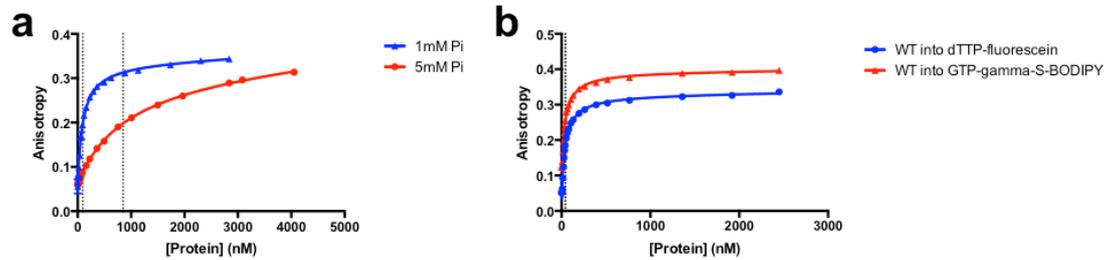
314
315 **Extended Data Figure 2: Controls for dNTP-binding experiment.** **a**, Titration of $CA_{Hexamer}$ into
316 2nM fluorescein-labelled dTTP in the presence of 1mM (physiological) or 5mM inorganic phosphate.
317 Under the 1 mM conditions, there is no significant effect on hexamer binding to dTTP. At 5 mM
318 apparent affinity is decreased to 851 nM, demonstrating that inorganic phosphate can compete for the
319 pore. However, given that the intracellular [dNTP] is approximately 100 μM, under intracellular
320 conditions dNTP binding would dominate. **b**, Titration of $CA_{Hexamer}$ into BODIPY-labelled rGTP-γ-S
321 and fluorescein-labelled dTTP. Each binds with the same affinity suggesting that the R18 pore is
322 unable to discriminate between ribose and deoxyribose nucleoside triphosphates. The difference in the
323 magnitude of the fluorescence anisotropy signals is due to differences in fluorophore excited state
324 lifetimes. K_D values are indicated by a dotted line.
325

326
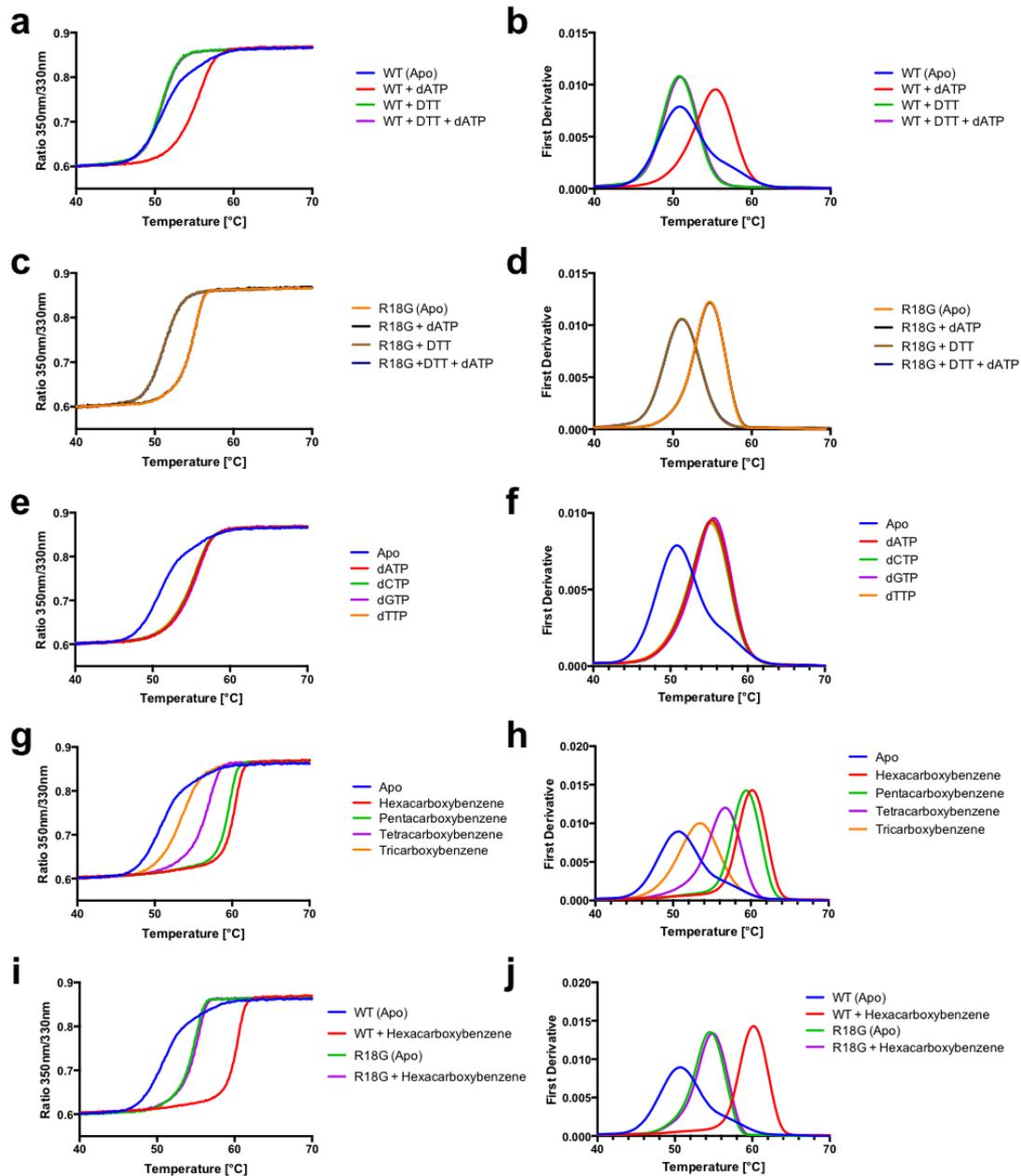
327 **Extended Data Figure 3: DSF melt curves.** The left-hand panels report the ratio of tryptophan
328 fluorescence emission at 350nm and 330nm as a function of temperature. The right-hand panels report
329 the first derivative of the same data, the peak of which is used to determine the T_m value. **a**, **b**, Effect of
330 dATP and DTT on WT $CA_{Hexamer}$. **c**, **d**, Effect of dATP and DTT on R18G $CA_{Hexamer}$. **e**, **f**, Effect of
331 each dNTP on WT $CA_{Hexamer}$. **g**, **h**, Comparison of the effects of carboxybenzene compounds on WT
332 $CA_{Hexamer}$. **i**, **j**, Comparison of the effects of hexacarboxybenzene on WT and R18G $CA_{Hexamer}$.

333

```
ALV {alpharetrovirus}    - - - P A V I K T E G - - P A W T P L E P K L I T R L A D T V R A K G L R S P - I T M A E V E A L M S S P L L P H D V T
RSV {alpharetrovirus}    - - - P V V I K T E G - - P A W T P L E P K L I T R L A D T V R T K G L R S P - I T M A E V E A L M S S P L L P H D V T
JSRV {betaretrovirus}    - - - P V F E N N N Q - - R Y Y E S L P F K Q L K E L K I A C S Q Y G P T A P - F T I A M I E S L G T Q A L P P N D W K
MMTV {betaretrovirus}    P V V F M G E S D D D D T P V W E P L P L K T L K E L Q S A V R T M G P S A P - Y T L Q V V D M V A S Q W L T P S D W H
FeLV {gammaretrovirus}   - - - P L R E G P N N R P Q Y W P F S A S D L Y N W K S H N P P F S Q D P V A - L T N L I E S I L V T H Q P T W D D C Q
MLV {gammaretrovirus}    - - - P L R L G G N G Q W Q Y W P F S S S D L Y N W K N N N P S F S E D P G K - L T A L I E S V L T H Q P T W D D C Q
BLV {deltaretrovirus}    - - - - - P I I S E G N R N R H R A W A L R E L Q D I K K E I E N K A P G S Q V W I Q T L R L A I L Q A D P T P A D L E
HTLV_1 {deltaretrovirus} - - - - P V M H P H G A P P N H R P W Q M K D L Q A I K Q E V S Q A A P G S P Q F M Q T I R L A V Q Q F D P T A K D L Q
HIV-1 {lentivirus}       - - - - P I V Q N L Q G Q M V H Q A I S P R T L N A W V K V V E E K A F S P E - - - V I P M F S A L S D G A T P Q D L N
EIAV {lentivirus}        - - - - P I M I D G A G N R N F R P L T P R G Y T T W V N T I Q T N G L L N E - - A S Q N L F G I L S V D C T S E E M N
```

Extended Data Figure 4: Alignment of selected retrovirus capsid sequences bordering the electropositive pore. The position equivalent to R18 in HIV is marked with an arrow.
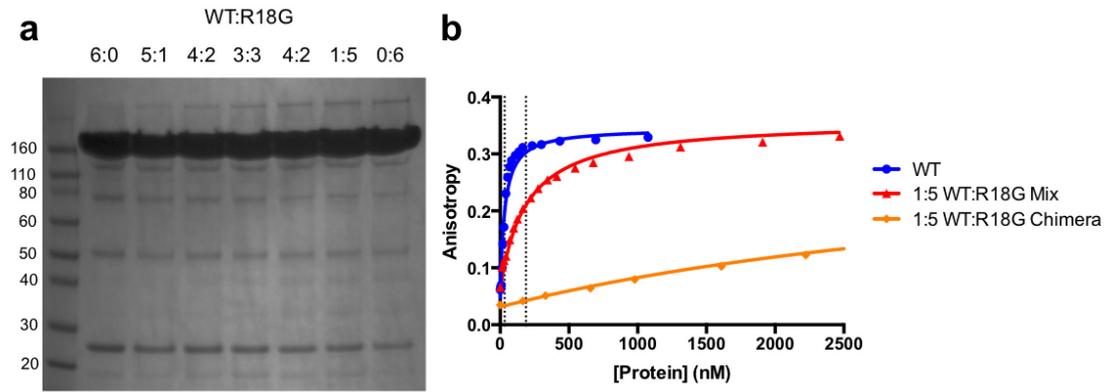
338

339 **Extended Data Figure 5: Confirmation of CA$_{\text{Hexamer}}$ Chimera Assemblies.** **a**, Non-reducing SDS-
340 PAGE of CA$_{\text{Hexamer}}$ WT:R18G chimera samples demonstrates that the recombinant proteins had
341 reassembled into hexamers. Molecular weight standards (kDa) are presented in the first lane. **b**,
342 Comparison of 1:5 homohexamer mix and the equivalent chimera. The 1:5 WT:R18G mix experiences
343 a six-fold loss of apparent K$_d$, as expected for a 6-fold dilution of WT with a non-binding mutant. In
344 contrast, the 1:5 chimera chimera has a 58-fold decrease in K$_d$, demonstrating that chimeric hexamers
345 had indeed formed.

346

347
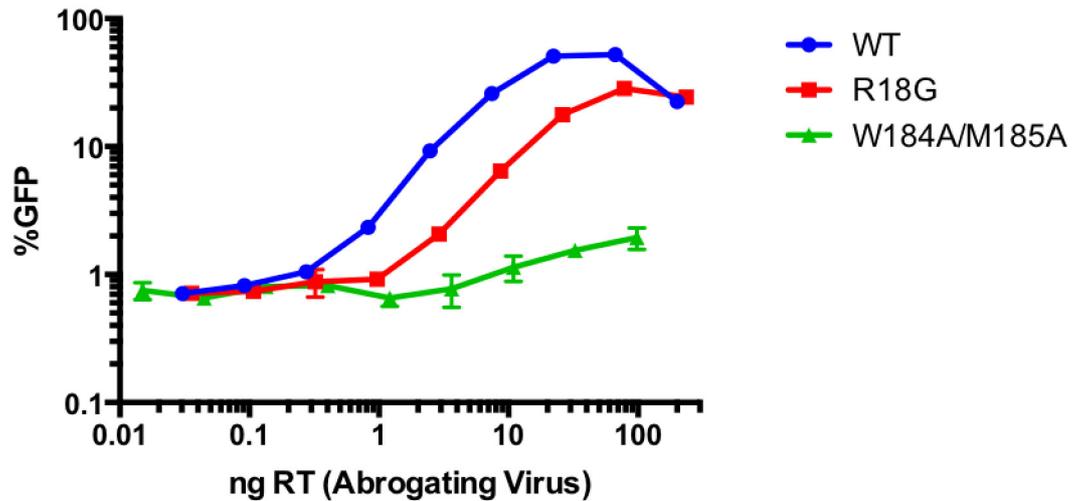
Extended Data Figure 6: R18G is capable of abrogating TRIM5α-mediated restriction. Rhesus TRIM5α provides a potent block to infection of HIV in FRhK-4 cells. Titration of a non-GFP-expressing virus can compete for TRIM5α-binding and relieve the restriction of a GFP-expressing virus only if it delivers an assembled capsid into the cytoplasm. R18G abrogates restriction but W184A/M185A, which is incapable of forming assembled capsids due to loss of the CTD-CTD dimerization interface, does not.

a

b
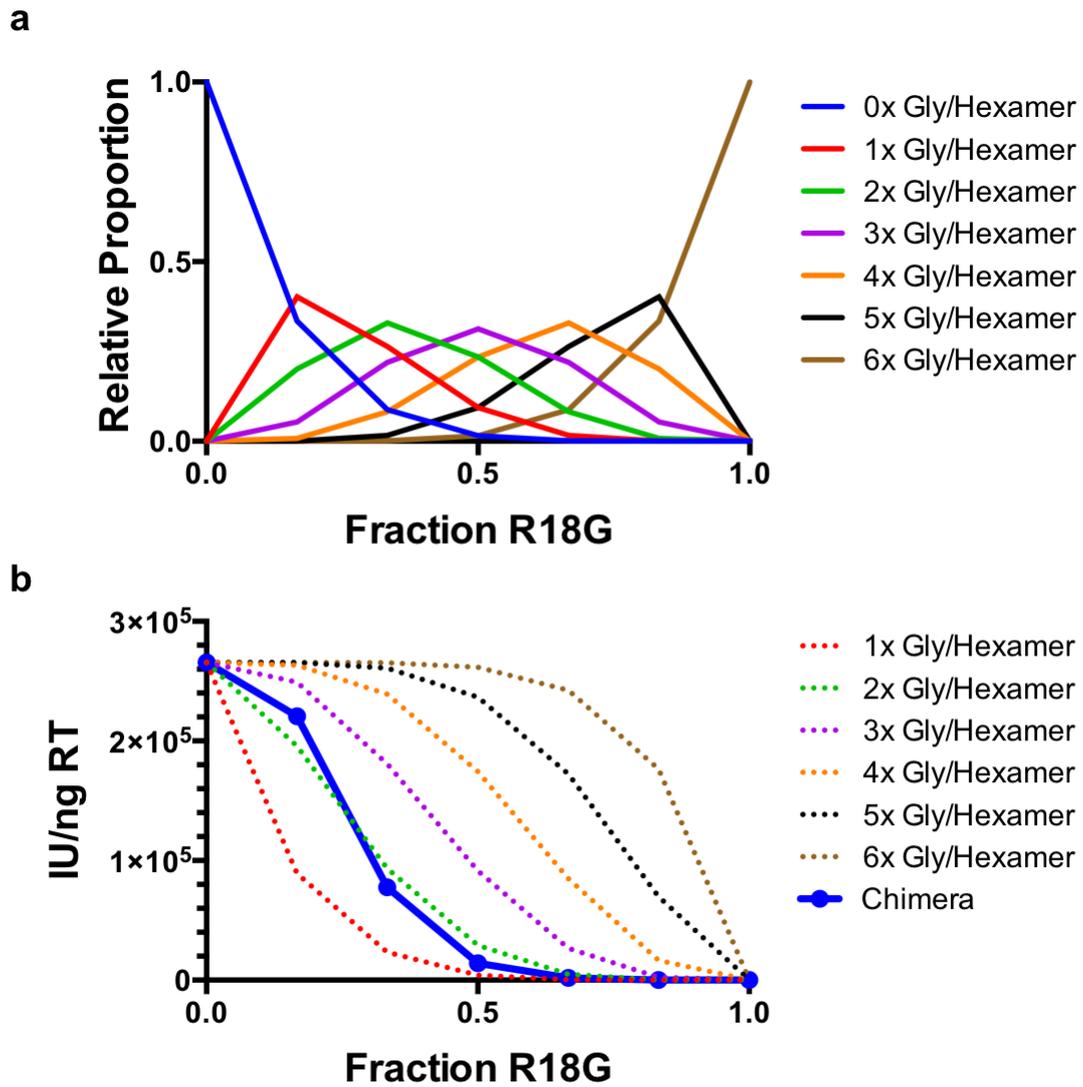
356

357 **Extended Data Figure 7: Modelling the effect of progressive removal of arginines from the pore.**

358 **a**, Binomial distribution model for the relative proportion of capsid hexamers carrying a discrete

359 number of glycines at position 18 at defined bulk ratios of WT:R18G. **b**, Six models (dotted lines)

360 predicting the effect of replacing arginine 18 with glycines. Each model assumes a different number of

361 glycines is required to render the pore defective. The data from WT:R18G chimeric virus

362 measurements (solid line) is consistent with a model in which four or more arginines (i.e. 2 or fewer

363 glycines, green) are required to maintain a functional pore.

364

365

366
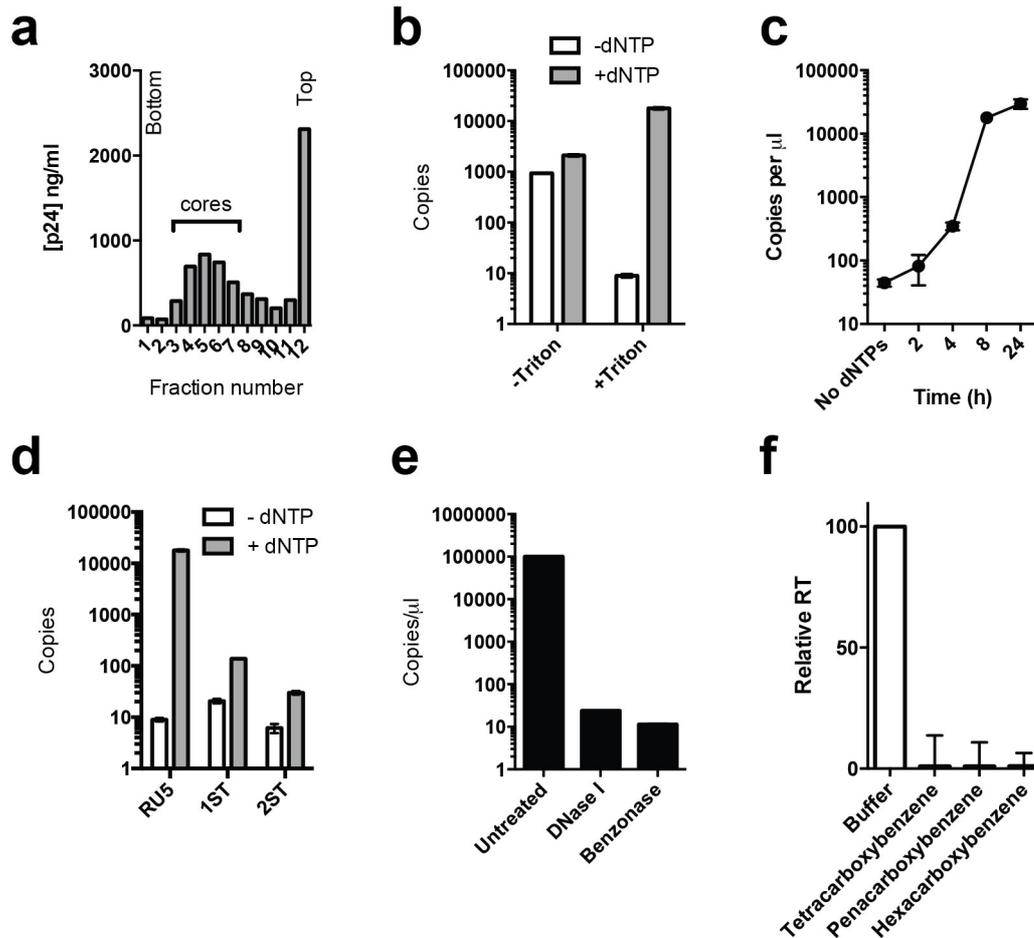
Extended Data Figure 8: ERT assay. a, HIV-1 cores were prepared by ultracentrifugation through a Triton X-100 layer over a sucrose gradient. Resulting fractions were subjected to ELISA for p24 and fractions 3 – 7 were pooled for further experiments. **b**, Endogenous RT activity for strong stop in the presence of DNase I using HIV-1 fractions that were prepared with or without the Triton X-100 spin-through layer. Input levels of p24 were normalized between reactions. **c**, dNTP's were added to HIV-1 cores prepared by Triton X-100 spin-through in the presence of DNase I. Reactions were stopped at the indicated time point by shifting to -80° C and levels of strong stop were quantified. **d**, Levels of strong-stop (RU5), first-strand transfer (1ST) and second-strand transfer (2ST) DNA after overnight incubation of HIV-1 cores with or without dNTP's in the presence of DNase I. **e**, Levels of naked HIV-1 DNA genomes untreated or incubated overnight with DNase I or Benzonase. **f**, Effect of carboxybenzene compounds on recombinant reverse transcriptase activity.

378

379
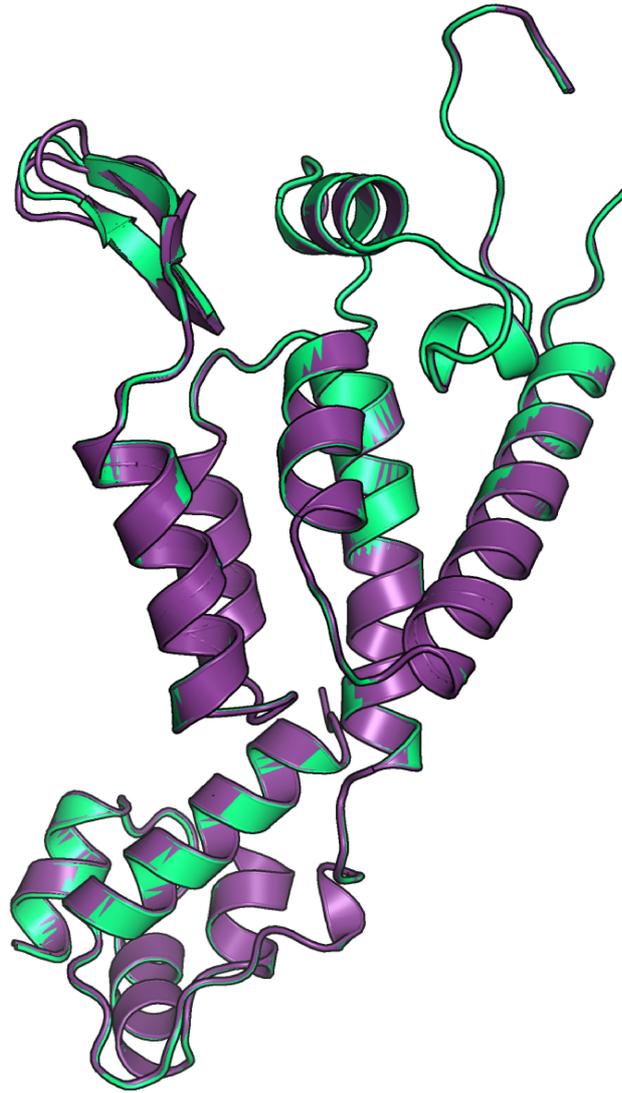
Extended Data Figure 9: Comparison of WT and H12Y crystal structures. The H12Y monomer (in the context of the hexamer, purple) superposes on the WT (green) with RMSD = 0.2471 Å. Residues 4-9 of the H12Y structure have been modeled in two alternate conformations owing to flexibility towards the tip of the hairpin.

385 **Extended Data Table 1: Details of structures used for β-hairpin analysis**
386
387

Molecule	Chain	Resolution	Crystallization pH	Q7 Displacement	H12-D51 distance
5HGL	A	3.1	5.5	3	2.9
5HGL	B	3.1	5.5	2.5	3.1
5HGK	A	1.76	6.3	0	2.7
5HGK	B	1.76	6.3	2.9	2.7
3NTE	A	1.95	6.5	2	2.9
3NTE	B	1.95	6.5	3.8	2.8
2X83	A	1.7	7	7.8	3.7
1AK4	C	2.36	7	10	4.8
2X83	B	1.7	7	9.2	4.5
4B4N	A	1.81	7	13.1	4.9
3H4E	A	2.7	7.5	10.4	4.6
2GON	C	1.9	8	11.9	4.8
3P05	B	2.5	8.5	14.8	4.6
3P05	C	2.5	8.5	10.7	4.6

Extended Data Table 2: Crystallographic data collection and refinement statistics

	CA$_{NTD}$ (OPEN)	CA$_{Hexamer}$ (OPEN)	CA$_{Hexamer}$ + dATP	CA$_{Hexamer}$ (APO, CLOSED)	CA$_{Hexamer}$ (R18G)	CA$_{Hexamer}$ + Hexacarboxy-benzene	CA$_{Hexamer}$ (H12Y)
Data collection							
Space group	P2$_1$	C222$_1$	P6	P6	P6	P6	P6
Cell dimensions							
a, b, c (Å)	43.72, 23.85, 129.55	89.69, 159.27, 249.40	90.81, 90.81, 56.68	90.73, 90,73, 56.75	90.81, 90.81, 56.88	90.76, 90.76, 56.76	90.60, 90.60, 56.93
α, β, γ (°)	90, 96.31, 90	90, 90, 90	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	39.87-1.76 (1.86-1.76)	19.99-3.10 (3.27-3.10)	45.98-2.03 (2.08-2.03)	26.69-1.90 (1.94-1.90)	46.09-2.00 (2.05-2.00)	45.38-1.95 (2.00-1.95)	56.93-1.70 (1.73-1.70)
R_{merge}	0.065 (0.227)	0.094 (0.552)	0.112 (0.551)	0.121 (0.748)	0.166 (0.831)	0.105 (0.862)	0.095 (0.813)
I / σI	13.3 (5.0)	7.1 (1.9)	8.2 (1.9)	8.3 (2.1)	6.5 (1.9)	11.7 (2.1)	9.0 (2.4)
Completeness (%)	91.9 (80.6)	97.4 (92.0)	94.9 (69.1)	99.8 (100.0)	100.0 (100.0)	97.9 (99.5)	95.1 (94.7)
Redundancy	4.6 (4.6)	2.4 (2.4)	2.9 (2.4)	5.1 (4.8)	6.0 (5.9)	6.8 (6.5)	4.8 (4.9)
Refinement							
Resolution (Å)	39.87-1.76 (1.81-1.76)	19.99-3.10 (3.18-3.10)	45.98-2.04 (2.10-2.04)	26.69-1.90 (1.95-1.90)	39.32-2.00 (2.05-2.00)	45.38-1.95 (2.00-1.95)	56.93-1.70 (1.74-1.70)
No. reflections	23837 (1552)	30312 (2044)	15488 (1084)	19999 (1468)	17285 (1246)	18188 (1347)	26370 (1949)
R_{work} / R_{free}	0.190/0.224 (0.233/ 0.286)	0.250/0.281 (0.368/ 0.399)	0.236/0.263 (0.316/ 0.384)	0.200/0.225 (0.266/ 0.342)	0.205/0.221 (0.216/ 0.255)	0.194/0.221 (0.266/ 0.299)	0.197/0.232 (0.265/ 0.285)
No. atoms							
Protein	2284	9358	1558	1623	1605	1612	1688
Ligand/ion	1	193	30	-	-	24	-
Water	338	-	82	124	105	119	123
B-factors							
Protein	27.928	54.173	26.577	30.265	27.732	30.878	27.523
Ligand/ion	38.516	83.117	109.040	-	-	95.409	-
Water	24.730	-	30.031	35.719	28.567	34.088	33.604
R.m.s. deviations							
Bond lengths (Å)	0.007	0.008	0.007	0.007	0.007	0.008	0.006
Bond angles (°)	1.222	1.156	0.991	0.998	1.046	1.281	1.009

390 **Methods**

391

392 **Protein production and purification**

393 The CA N-terminal domain and the disulfide-stabilised $CA_{Hexamer}$ were expressed and purified as

394 previously described[8,27]. The R18G mutation was introduced by QuikChange site-directed

395 mutagenesis. Chimeric $CA_{Hexamers}$ were produced by mixing the desired ratio of pre-assembled WT and

396 R18G $CA_{Hexamer}$ (16 mg/ml) followed by a four-step dialysis: i) Disassembly in TRIS (pH 8.0, 50mM),

397 NaCl (40mM), β-mercaptoethanol (20mM); ii) Reassembly in TRIS (pH 8.0, 50mM), NaCl (1 M), β-

398 mercaptoethanol (20mM); iii) Oxidation in TRIS (pH 8.0, 50mM), NaCl (1 M); Redispersion in TRIS

399 (pH 8.0, 20mM), NaCl (40mM). In the context of the chimera experiments, WT and R18G were also

400 subjected to this process so that samples were matched with the other ratios. Reassembled hexamers

401 were observed by non-reducing SDS-PAGE. Chimeric hexamers were compared with mixes of

402 homohexamers by fluorescence anisotropy (see below and Extended Data Fig. 5) in order to

403 demonstrate that chimeras had indeed formed.

404

405 **Crystallisation, structure solution and analysis**

406 All crystals were grown at 17 °C by sitting-drop vapour diffusion in which 100 nL protein was mixed

407 with 100 nL precipitant and suspended above 80 uL precipitant. The CA N-terminal domain

408 (15mg/ml) 'open' conformation was crystallised from PEG3350 (20%), Ammonium chloride (0.2M,

409 pH 6.3). Crystals were cryoprotected in precipitant supplemented with 25% glycerol. The $CA_{Hexamer}$

410 (15 mg/ml) 'open conformation' was crystallised from PEG4000 (12%), NaCl (0.1M), $MgCl_2$ (0.1M),

411 sodium citrate (0.1M, pH 5.5). Crystals were cryoprotected in precipitant supplemented with 20%

412 MPD. The remaining $CA_{Hexamer}$ structures (apo, dATP-bound, R18G and hexacarboxybenzene-bound)

413 were all obtained from 10-12mg/ml protein mixed with PEG550MME (13-14%), KSCN (0.15M),

414 TRIS (0.1M, pH 8.5) and cryoprotected with precipitant supplemented with 20% MPD. For the dATP-

415 bound structure, the protein was supplemented with 10 mM dATP immediately prior to crystallisation;

416 while for the hexacarboxybenzene structure, the protein was likewise supplemented with 1 mM

417 hexacarboxybenzene (TRIS-buffered to pH 8.0). All crystals were flash-cooled in liquid nitrogen and

418 data collected either in-house using Cu *Kα* X-rays produced by a Rigaku FR-E rotating anode generator

419 with diffraction recorded on a mar345 image plate detector (marXperts), or at beamline I02 at Diamond

420 Light Source. The datasets were processed using the CCP4 program suite[28]. Data were indexed and

421 integrated with IMOSFLM[29] and scaled and merged with either POINTLESS and SCALA[30] or

422 AIMLESS[31]. Structures were solved by molecular replacement using PHASER[32] and refined using

423 REFMAC5[33]. Between rounds of refinement, the model was manually checked and corrected against

424 the corresponding electron-density maps in COOT[34]. Solvent molecules and bound ligands were added

425 as the refinement progressed either manually or automatically within COOT and were routinely

426 checked for correct stereo-chemistry, for sufficient supporting density above a 2Fo-Fc threshold of

427 1.0σ and for a reasonable thermal factor. The quality of the model was regularly checked for steric

428 clashes, incorrect stereochemistry and rotamer outliers using MOLPROBITY[35]. Final figures were

429 rendered in The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC. Surface

430 electrostatics were calculated using the APBS PyMOL plugin[36] and cavity volume measurements with

431 3V[37]. Data collection and refinement statistics are presented in Extended Data Table 2. The 'fullerene

432 cone' model of an HIV-1 virion is based on 3J3Q[38] but using the inter-hexamer packing from 4XFY[39]

433 and an open β-hairpin conformation.

434

435 **Fluorescence Anisotropy**

436 Fluorescence anisotropy measurements were performed at 22 °C on a Cary Eclipse Fluorescence

437 Spectrophotometer (Agilent). Fluorescein-labelled dNTP's were obtained from Perkin Elmer and used

438 for saturation binding experiments at a concentration of 2nM prepared in 'Intracellular Buffer':

439 potassium gluconate (110mM), KCl (25mM), NaCl (5mM), $MgCl_2$ (2mM), HEPES (10mM), final pH

440 7.2. $CA_{Hexamer}$ disassembly was achieved by the addition of DTT (4mM), and was performed routinely

441 at the conclusion of each saturation binding experiment to confirm the absence of non-specific binding.

442 It was found that the triphosphate was not stable over the timescale of the competition binding

443 experiments; so fluorescein-labelled dNTP's were substituted for a non-hydrolysable BODIPY-labelled

444 GTP-γ-S (ThermoFisher Scientific). Saturation binding experiments determined that this non-

445 hydrolysable analogue bound with unchanged affinity to the $CA_{Hexamer}$ (K_d = 14nM). 200 mM stock

446 solutions of hexacarboxybenzene (Sigma), pentacarboxybenzene (MP Biomedicals), 1,2,4,5-

447 tetracarboxybezene (Sigma), and 1,3,5-tricarboxybenzene (Fluka) were prepared in 50mM TRIS and

448 adjusted to pH 8.0. For competition binding experiments, the competitor was titrated into a mix of

449 $CA_{Hexamer}$ (28nM) and BODIPY-GTP-γ-S (2nM). All fluorescence anisotropy measurements are

450 representative of at least two experiments. Each point is measured in quadruplicate and plotted as

451 mean ± standard deviation. In many cases error bars lie within the datapoint. Saturation binding and

452 competition binding curves were fit using GraphPad Prism (GraphPad Software, Inc.).

453

454 **Rapid reaction kinetics**

455 Experiments were carried out using a dual-channel fluorescence TgK single-mix SF-61SX2 stopped-

456 flow spectrometer. All samples were prepared in Intracellular Buffer. Mixing was performed 1:1, using

457 an excitation wavelength of 488 nm and a 520 nm cutoff filter. Association experiments were carried

458 out at 0.25 μM dCTP and a range of μM $CA_{Hexamer}$ concentrations. Dissociation experiments were

459 carried out using 20 μM unlabeled dCTP and a pre-formed fluorescein-labeled 1μM dCTP: $CA_{Hexamer}$

460 complex. Relaxation rates were determined using a single exponential model: $F = \Delta F \exp(-k_{obs}t) + F_e$,

461 where F is the observed fluorescence, ΔF is the fluorescence amplitude, k_{obs} is the observed pseudo

462 first-order rate constant, and F_e is the end-point fluorescence. The bimolecular association rate constant

463 (k_{on}) was determined by fitting the linear relationship between k_{obs} and the increasing pseudo-first order

464 concentrations of $CA_{Hexamer}$ to: $k_{obs} = k_{on}[CA_{Hexamer}] + k_{reverse}$. For stopped flow experiments, every 0.5 s

465 measurement included >2000 datapoints, each of which was oversampled 99 times. At least three

466 independent mixing experiments were averaged for each ligand concentration.

467

468 **Differential Scanning Fluorimetry**

469 DSF measurements were performed using a Prometheus NT.48 (NanoTemper Technologies) over a
470 temperature range of 20-95 °C using a ramp rate of 2.5 °C/min. $CA_{Hexamer}$ samples were prepared at a
471 final concentration of 1mg/ml in Intracellular Buffer (+/- DTT (4mM)). dNTP's or competitors were
472 added at 200 μM. DSF scans are single reads. Consistency between like points yields an uncertainty in
473 T_m of no greater than 0.2 °C.

474

475 **Cells and Viruses**
476 Replication deficient VSV-G pseudotyped HIV GFP vectors were produced in HEK293T cells as
477 described previously[3]. Site-directed mutagenesis of CA was performed using the QuikChange method
478 (Stratagene) against the Gag-Pol expression plasmid, pCRV-1. Chimeric viruses were produced by
479 mixing the appropriate ratio of WT or mutant pCRV-1 prior to transfection. Reverse transcriptase
480 activity was quantified using a colorimetric ELISA assay (Roche) and was found not to vary
481 significantly between viruses. Production of mature particles was confirmed by western blot for p24
482 from pelleted virus, with no observable difference between chimeras.

483

484 **Infection Experiments**
485 Infections of HeLa cells were performed in the presence of 5 μg/ml polybrene. GFP expressing cells
486 were enumerated on a BD LSRII Flow Cytometer (BD Biosciences) 2 days post-transfection after
487 fixation of cells in 4% paraformaldehyde. Chimera infectivity was determined by a 6-point titration of
488 each chimera onto HeLa cells. Values are the mean ± standard deviation calculated from all points for
489 which the proportion of infected cells after 48 h was between 1% and 50%.

490

491 **TRIM5α Abrogation Assay**
492 'Abrogating virus' (VSV-G pseudotyped HIV Puromycin vectors) was produced as described above,
493 with the exception that the *gfp* gene was replaced with the *pac* gene to ensure the virus did not confer
494 fluorescence upon infection. Virus was concentrated by ultracentrifugation with an SW28 rotor at
495 25,000 rpm for 2 h. The abrogating virus capsids were WT, R18G or W184A/M185A (a mutant with a
496 known assembly defect that cannot compete for TRIM5α). VSV-G pseudotyped HIV GFP vectors
497 were titrated on FRhK-4 cells in the presence of 5 μg/ml polybrene to determine the volume of virus
498 required to achieve 1% infection. In a separate experiment, cells were then coinfected with that
499 amount HIV-GFP vector and a titration of VSV-G pseudotyped HIV Puromycin vectors (the
500 abrogating virus). GFP expressing cells were measured in duplicate and enumerated as above. Results
501 are representative of three experiments and are presented as mean ± standard deviation. For many
502 points the error bars lie within the datapoint.

503

504 **Quantitative PCR**
505 For analysis of reverse transcription products, viral supernatant was treated with 250 U/ml DNase
506 (Millipore) for 1 h prior to infection. Cells were harvested 6h post infection. DNA was extracted using
507 DNeasy Blood and Tissue Kit (Qiagen). GFP copies were quantified using primers GFPF
508 (CAACAGCCACAACGTCTATATCAT), GFPR (ATGTTGTGGCGGATCTTGAAG) and probe

509   GFPP (FAM-CCGACAAGCAGAAGAACGGCATCAA-TAMRA) against a standard curve of CSGW

510   on an ABI StepOnePlus Real Time PCR System (Life Technologies). Chimera reverse transcription

511   measurements are representative of 3 experiments with each point measured in triplicate. Results are

512   presented as mean ± standard deviation. For H12Y, a timecourse was also performed, in which each

513   time point was measured in triplicate and presented as above.

514

515   **Preparation of HIV-1 cores**

516   HIV-1 capsid cores were prepared using a protocol based on[21] with modifications. 90ml HEK293T

517   supernatant containing VSV-G pseudotyped HIV-1 GFP was pelleted over 20% sucrose dissolved in

518   core prep buffer (CPB; 20 mM Tris pH 7.4, 20 mM NaCl, 1 mM $MgCl_2$) in an SW28 rotor (Beckman)

519   at 25,000 rpm at 4° C. Pellets were gently resuspended at 4° C in CPB for 1 h with occasional agitation.

520   Resuspended pellets were treated with DNase I from bovine pancreas (Sigma Aldrich) for 1 h at 200

521   µg/ml at room temperature to remove contaminating extra-viral DNA. Virus was subjected to spin-

522   through detergent stripping of the viral membrane as follows. A gradient at 80-30% sucrose was

523   prepared in SW40Ti ultracentrifuge tubes and overlaid with 250 µl 1% Triton X-100 in 15% sucrose,

524   followed by 250 µl 7.5% sucrose. All solutions were prepared in CPB. 750 µl DNase-treated,

525   concentrated virus was layered on top of the gradient and subjected to 32,500 rpm at 4° C for 16 h. The

526   preparation was fractionated and the location of cores was determined by ELISA for p24 (Perkin

527   Elmer). Core-containing fractions were pooled and snap frozen before storage at -80° C.

528

529   **Endogenous reverse transcription assays**

530   Viral cores were diluted to 400 µg/ml p24 with 60% sucrose in CPB and pre-treated with nucleases for

531   1h before addition of dNTPs. Final concentrations of dNTP's were 100 µM each, DNase I and RNase

532   A were at 100 µg/ml and Benzonase was at 250 U/ml. 20 µl reactions were incubated at room

533   temperature for 16 h unless indicated otherwise and were stopped by shifting to -80° C. DNA was

534   prepared using DNeasy Blood and Tissue kit (Qiagen) after addition of 200 µl PBS with of 50 µg/ml

535   salmon sperm carrier DNA to each sample. Reverse transcript products were detected using TaqMan

536   Fast Universal PCR Mix (ABI) and RU5 primers to detect strong stop DNA[40] (RU5 fwd

537   TCTGGCTAACTAGGGAACCCA, RU5 rev CTGACTAAAAGGGTCTGAGG and RU5 probe FAM-

538   TTAAGCCTCAATAAAGCTTGCCTTGAGTGC-TAMRA), GFP primers to detect first strand

539   transfer products (described above) and primers for second strand transfer products[40] (2ST fwd

540   TTTTAGTCAGTGTGGAAAATCTGTAGC, 2ST rev TACTCACCAGTCGCCGCC and 2ST probe

541   FAM-TCGACGCAGGACTCGGCTTGCT-TAMRA). Where used, carboxybenzene compounds were

542   dissolved in CPB, pH adjusted with NaOH and added to reactions at a final concentration of 20mM. In

543   order for dNTP concentration to be limiting, these reactions were performed in the presence of 1 µM

544   each dNTP and reactions were stopped 5 h after their addition. ERT experiments were performed in

545   experimental triplicate and are representative of several experimental replicates. Data are represented

546   as mean ± sem

547

548 **Supplementary Video 1: Structural morph between the closed and open states of CA$_{Hexamer}$.** On
549 the left the protein is represented in cartoon format, coloured according to secondary structure. The
550 sidechain of L6 is shown as sticks to emphasise that it is this residue that results in pore closure. On
551 the right P1, H12, T48, Q50, and D51 are represented as sticks to show that the movement of the β-
552 hairpin is driven by the formation of a salt-bridge between H12 and D51. Distances shown are in
553 Ångstroms.
554
555 **Supplementary Video 2: Pore opening exposes R18.** Surface representation of the morph depicted
556 in Supplementary Video 1. The β-hairpin and R18 are coloured yellow and blue, respectively.
557

558 **Extended Data References**
559
560 27 Price, A. J. *et al.* Active site remodeling switches HIV specificity of antiretroviral TRIMCyp.
561   *Nature structural & molecular biology* **16**, 1036-1042 (2009).
562 28 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D*
563   **67**, 235-242 (2011).
564 29 Leslie, A. G. W. & Powell, H. R. Processing diffraction data with MOSFLM. *Nato Sci Ser Ii*
565   *Math* **245**, 41-51 (2007).
566 30 Evans, P. R. An introduction to data reduction: space-group determination, scaling and
567   intensity statistics. *Acta Crystallogr D* **67**, 282-292 (2011).
568 31 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta*
569   *Crystallogr D* **69**, 1204-1214 (2013).
570 32 Mccoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
571 33 Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by
572   the maximum-likelihood method. *Acta Crystallogr D* **53**, 240-255 (1997).
573 34 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr*
574   *D* **60**, 2126-2132 (2004).
575 35 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular
576   crystallography. *Acta Crystallogr D* **66**, 12-21 (2010).
577 36 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of
578   nanosystems: Application to microtubules and the ribosome. *Proceedings of the National*
579   *Academy of Sciences of the United States of America* **98**, 10037-10041 (2001).
580 37 Voss, N. R. & Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor.
581   *Nucleic Acids Res* **38**, W555-W562 (2010).
582 38 Zhao, G. *et al.* Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom
583   molecular dynamics. *Nature* **497**, 643-646 (2013).
584 39 Gres, A. T. *et al.* STRUCTURAL VIROLOGY. X-ray crystal structures of native HIV-1
585   capsid protein reveal conformational variability. *Science* **349**, 99-103 (2015).
586 40 Julias, J. G., Ferris, A. L., Boyer, P. L. & Hughes, S. H. Replication of phenotypically mixed
587   human immunodeficiency virus type 1 virions containing catalytically active and catalytically
588   inactive reverse transcriptase. *Journal of virology* **75**, 6537-6546 (2001).
589
590
591