

Virology

Mechanisms of HIV-1 integrase resistance to dolutegravir and potent inhibition of drug-resistant variants

Min Li^{1†}, Dario Oliveira Passos^{2†}, Zelin Shan², Steven J. Smith³, Qinfang Sun⁴, Avik Biswas^{2,5}, Indrani Choudhuri⁴, Timothy S. Strutzenberg², Allan Haldane⁵, Nanjie Deng⁶, Zhaoyang Li¹, Xue Zhi Zhao³, Lorenzo Briganti⁷, Mamuka Kvaratskhelia⁷, Terrence R. Burke Jr.³, Ronald M. Levy⁴, Stephen H. Hughes³, Robert Craigie^{1*}, Dmitry Lyumkis^{2,8,9*}

Copyright © 2023
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
License 4.0 (CC BY).

HIV-1 infection depends on the integration of viral DNA into host chromatin. Integration is mediated by the viral enzyme integrase and is blocked by integrase strand transfer inhibitors (INSTIs), first-line antiretroviral therapeutics widely used in the clinic. Resistance to even the best INSTIs is a problem, and the mechanisms of resistance are poorly understood. Here, we analyze combinations of the mutations E138K, G140A/S, and Q148H/K/R, which confer resistance to INSTIs. The investigational drug 4d more effectively inhibited the mutants compared with the approved drug Dolutegravir (DTG). We present 11 new cryo-EM structures of drug-resistant HIV-1 intasomes bound to DTG or 4d, with better than 3-Å resolution. These structures, complemented with free energy simulations, virology, and enzymology, explain the mechanisms of DTG resistance involving E138K + G140A/S + Q148H/K/R and show why 4d maintains potency better than DTG. These data establish a foundation for further development of INSTIs that potentially inhibit resistant forms in integrase.

INTRODUCTION

HIV-1 and other retroviruses integrate a DNA copy of their viral RNA genome into the genome of the host cell, a process that is mediated by the virally encoded enzyme integrase (IN) (1, 2). The integrated proviruses serve as templates for transcription of new viral RNA genomes and the mRNAs required for production of progeny virions. The catalytic integration reaction that inserts a DNA copy of the viral genome in the host genome presents a target for treating people living with HIV-1 (PLWH) and for preventing new infections.

Two chemical steps are required for integration: (i) 3'-end processing, in which IN cleaves two nucleotides from each 3'-end of the viral DNA, and (ii) DNA strand transfer, in which IN inserts the ends of the viral DNA into target DNA in a one-step transesterification reaction (3). Completion of integration requires removal of two nucleotides from the 5'-ends of the viral DNA, filling in of the single strand gaps between viral and target DNA, and ligation. These latter steps are mediated by cellular enzymes. IN is both necessary and sufficient to carry out both 3'-processing and DNA

strand transfer in vitro. Each step of DNA integration, up to formation of the integration intermediate, occurs within a series of stable nucleoprotein complexes (intasomes), involving a multimer of IN and a pair of viral DNA ends. The atomic structures have now been determined for a variety of retroviral intasomes (4–13), including HIV-1 (14, 15), revealing differences in the oligomeric assemblies and how they engage target DNA. The formation of intasome species and the intasome-mediated insertion of the viral DNA (vDNA) into the host genome are essential steps in the viral replication cycle [see (1, 2) for recent reviews].

Because integration is an essential step in viral replication, intasomes are targeted by an important class of antiretroviral drugs, the IN strand transfer inhibitors (INSTIs). INSTIs selectively interact with both the bound viral DNA and IN and, as a consequence, do not bind tightly to free IN (4, 16, 17). INSTIs work by chelating the two Mg²⁺ ions, blocking the active site and inhibiting the strand transfer activity of IN. Five INSTIs have been clinically approved by the U.S. Food and Drug Administration and antiretroviral regimens that include INSTIs are standard of care. These include first-generation INSTIs raltegravir (RAL, approved 2007) and elvitegravir (EVG, approved 2011), as well as second-generation INSTIs dolutegravir (DTG, approved 2013), bictegravir (BIC, approved 2018), and cabotegravir (CAB, approved 2021). INSTIs are widely considered to be among the best therapeutic options for use in combination therapies to treat HIV-infected patients (18).

The major difference between first- and second-generation INSTIs lies in their ability to inhibit mutant forms of IN. Whereas both RAL and EVG can potentially inhibit the wild-type (WT) enzyme, drug-resistant mutations (DRMs) quickly develop (19). The second-generation INSTIs can potentially inhibit many of the mutant forms of IN that arise during the course of treatment with the first-generation inhibitors (20–22). However, clinical data indicate that resistance to second-generation INSTIs is now a

¹National Institute of Diabetes and Digestive Diseases, National Institutes of Health, Bethesda, MD, 20892, USA. ²The Salk Institute for Biological Studies, La Jolla, CA, 92037, USA. ³Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD, 21702, USA. ⁴Center for Biophysics and Computational Biology, and Department of Chemistry, Temple University, Philadelphia, PA 19122, USA. ⁵Center for Biophysics and Computational Biology and Department of Physics, Temple University, Philadelphia, PA 19122, USA. ⁶Department of Chemistry and Physical Sciences, Pace University, New York, NY, 10038, USA. ⁷Division of Infectious Diseases, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA. ⁸Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, 92037, USA. ⁹Graduate School of Biological Sciences, Section of Molecular Biology, University of California San Diego, La Jolla, CA 92093, USA.

*Corresponding author. Email: dlyumkis@salk.edu (D.L.); robertc@niddk.nih.gov (R.C.)

†These authors contributed equally to this work.

growing problem (23–25). Unfortunately, because HIV-1 replicates rapidly and the viral load is high, DRMs can arise, and viral sequences within an infected individual can differ by 10% or more (26). These data highlight the challenges to developing effective treatments. They also point to the fact that, while resistance is a problem, it can be addressed when the mechanisms are understood and by using this information to develop compounds that inhibit both the WT and the DRM enzymes.

The most frequently encountered DRMs to INSTI therapy include mutations at position Q148H/K/R. The Q148 mutations are commonly found in combination with the G140A/S mutations (23–25, 27) [and, rarely, G140C (27)]. In addition to the pair of DRMs at positions 140/148, numerous other DRMs arise in and around the active site, frequently including E138K (27). Recent structural work using the simian immunodeficiency virus (SIV_{rcm}) IN as a model for HIV-1 IN led to a proposed mechanism for the resistance of the G140S/Q148H double mutant (28). The mutation Q148H arises first, expelling a water molecule from the secondary coordinating shell of the Mg²⁺ ions and introducing a bulky electropositive residue underneath the bound ligand. The subsequent mutation of the neighboring residue G140S acidifies the Nε2 of H148. Coupling of the residues S140–H148–E152 increases the partial positive charge near the Mg²⁺ ligand. Collectively, these changes weaken chelation of the bound INSTI, leading to a reduction in drug binding. Because there are many complex DRMs that have a change at Q148 (H/K/R), Mg²⁺ chelation of Mg²⁺ was suggested to be the Achilles' heel of this drug class. However, SIV_{rcm} serves as a proxy for HIV-1; their active sites, including residues that are encountered in drug-resistant viral clones, are not identical. Moreover, only the double mutant G140S/Q148H was structurally defined, but there are many combinations of mutations at position 148H/K/R which can be paired with the G140A/S mutations and with other nearby mutations, including E138K. Thus, the full consequences of these and related mutations, and the structural basis of resistance in HIV-1 IN, which should help us understand how to develop INSTIs that can overcome resistance caused by mutations at these positions, remain unclear.

INSTIs containing a naphthyridine core have been developed into potent IN inhibitors that rival or even outperform clinical drugs in terms of their ability to inhibit a broad range of DRMs (20, 21, 29, 30). For example, the compound 4d was reported to inhibit certain DRMs, including variants with 140/148 mutations, with over an order of magnitude higher potency than the clinically used drug DTG (21). Structural biology data have shed light on how naphthyridine-containing compounds bind to intasomes from HIV-1 and prototype foamy virus (PFV) (15, 29), revealing insights into the small but important differences in the ways INSTIs bind to PFV and HIV-1 IN (15). However, how and why these compounds retain potency against DRMs remains unclear. Thus, there is a pressing need not only to explain the underlying mechanisms of viral resistance but also to understand why compounds like 4d can potentially inhibit resistant forms of IN.

Defining the underlying mechanisms of resistance lags behind clinical tabulation of emerging variants. This is primarily due to a lack of high-resolution structures of resistant variants and complementary tools to probe the resistance mechanisms. There have also been discrepancies in resistance profiles recorded for the 140/148 positions and their behavior in combination with other DRMs. Since mutations in HIV-1 IN that confer drug resistance occur in

several distinct combinations, structures of HIV-1 intasomes bound to inhibitors with and without such mutations, complemented by insights on inhibitory potencies and enzyme/viral fitness, are required to understand the details of how these mutations confer resistance and to guide the design of drugs with improved resistance profiles.

Here, we sought to understand the mechanism(s) of resistance caused by mutations at positions 138, 140, and 148. Our results reveal weaknesses in one of the leading clinically used drugs, DTG, and explain the mechanism(s) that underlie its loss of potency against DRMs. We extend these findings to explain how and why mutations at the three residue substitutions lead to drug resistance. Last, we provide insight into how a leading investigational compound, which is currently undergoing preclinical testing, retains considerable potency against the set of DTG-resistant variants we analyzed. Collectively, these data provide mechanistic insights into both drug resistance and its subversion and will enable more direct pathways to designing INSTIs with improved resistance profiles.

RESULTS

Inhibitory potencies of DTG and a leading developmental INSTI 4d against resistant mutants

Novel developmental INSTIs based on the naphthyridine scaffold maintain potency against drug-resistant variants with mutations at position Q148 of HIV-1 IN (15, 20, 21, 29–31). However, whether their potency is maintained across the Q148H/K/R family of DRMs, particularly in association with the E138K and G140A/S mutations, remains unclear. We determined the ability of a leading clinically used second-generation INSTI, DTG, to inhibit mutant viruses containing all possible combinations of E138K + G140A/S + Q148H/K/R DRMs using single-round viral infection assays and compared the data to the most potent preclinical developmental INSTI 4d (Fig. 1A).

We began by measuring the median effective concentration (EC₅₀) values against the single mutants in a single-round infection assays, which include E138K, G140A, G140S, Q148H, Q148K, and Q148R (Fig. 1B). Both DTG and 4d potently inhibited WT HIV-1 with EC₅₀ values of 2.6 ± 0.3 nM and 1.5 ± 0.3 nM, respectively. The compounds also retained full potency (<2.0 nM) against all the mutants E138K, G140A/S, and Q148H/K/R.

We next tested the IN double mutants E138K/G140A, E138K/G140S, E138K/Q148H, E138K/Q148K, E138K/Q148R, G140A/Q148H, G140A/Q148K, G140A/Q148R, G140S/Q148H, G140S/Q148K, and G140S/Q148R (Fig. 1C). To compare the relative effectiveness of the two drugs, we calculated the fold changes (FCs) in the EC₅₀ values, using the Monogram biological cutoffs for resistance for DTG as a guide (32). In general, both DTG and 4d retained potency against this panel of IN double mutants, but 4d has a moderately better antiviral profile compared to DTG. 4d potently inhibited the IN double mutant G140A/Q148K (4.6 ± 0.7 nM; FC = 3), whereas DTG did not (132.0 ± 9.9 nM; FC = 51, *P* value <0.001). This suggests that the G140A/Q148K mutant has specific properties that make it particularly susceptible to 4d but not to DTG.

Last, we tested the complex IN mutants containing three or more mutations, which include E138K/G140A/Q148H, E138K/G140A/Q148K, E138K/G140A/Q148R, E138K/G140S/Q148H, E138K/G140S/Q148K, and E138K/G140S/Q148R (Fig. 1D). 4d was

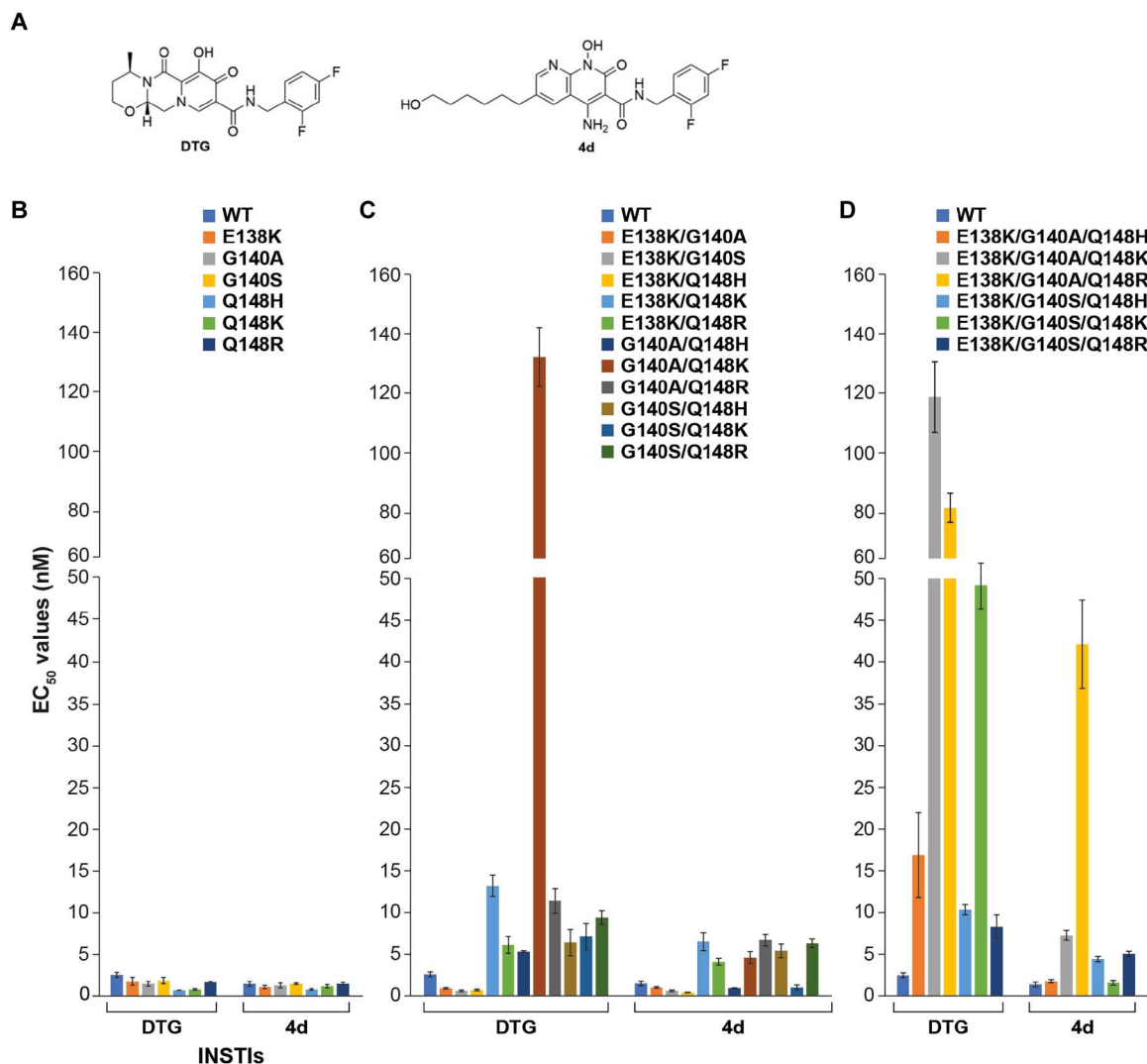


Fig. 1. Comparison of antiviral potencies of DTG and 4d against a panel of INSTI-resistant mutants. (A) Chemical structures of DTG and 4d, used for the assays. The EC_{50} values were determined for either of the two compounds using vectors that carry INSTI-resistant (B) single, (C) double, and (D) triple mutants in a single-round infection assay. The error bars represent SDs of independent experiments, $n = 3$, performed in triplicate.

considerably more effective than DTG against IN mutants E138K/G140A/Q148K (7.3 ± 0.6 nM; FC = 5 and 117.6 ± 11.6 nM; FC = 45, P values < 0.001, respectively) and E138K/G140S/Q148K (1.7 ± 0.3 nM; FC = 1 and 49.2 ± 2.7 nM; FC = 19, P values < 0.001, respectively). These results substantiate our observation that the double mutant G140A/Q148K is resistant to DTG and suggest that E138K can, under some circumstances, potentiate resistance. In accordance with the latter observation, the IN triple mutants formed by the addition of E138K to the IN double mutants G140A/Q148H, G140S/Q148H, G140S/Q148K, and G140A/Q148R all showed a decrease in susceptibility to DTG, although G140A/Q148K and G140S/Q148R remained largely unchanged. The only one of these triple mutants that led to a loss of susceptibility to 4d inhibition was E138K/G140A/Q148R. Although 4d did lose potency against E138K/G140A/Q148R (41.5 ± 5.2 nM; FC = 28), it retained more potency than DTG (P value < 0.001). Overall, as with the double

mutants, the antiviral profile of 4d was always better than DTG against this panel of triple mutants.

Collectively, we find that there are significant variations in potency among the Q148H/K/R variants, but combinations containing Q148K were consistently less susceptible to DTG inhibition. Mutations at position G140A/S typically potentiate the resistance of Q148H/K/R, especially variants containing Q148K. The mutation E138K can potentiate resistance in the Q148H/K/R series, sometimes decreasing the potency of an INSTI by up to an order of magnitude, although this effect was less consistent in comparison to G140A/S. The preclinical INSTI 4d retained potency better than DTG against all tested mutants. This difference is most apparent in the ~10- to 20-fold higher potency against the double mutant G140A/Q148K and the triple mutant E138K/G140A/Q148K.

Dissociation rates of 4d and DTG from intasomes

Variations in INSTI potencies have been previously attributed to differences in the dissociation rates of the first- and second-generation INSTIs from the intasome active site (28, 33). Here, we used surface plasmon resonance (SPR) to measure the dissociation rates of 4d versus DTG. For this, we used the WT PFV intasome, because accurate measurements of INSTI binding constants to HIV intasomes by SPR are challenging due to the large size differences between INSTIs and the HIV intasome and the fact that the HIV intasomes can be compositionally heterogeneous (~400 to 600 kDa) (14). We found that the dissociation rate of 4d from PFV intasome was substantially slower than the dissociation rate of DTG (fig. S1).

Enzymatic activities and viral replicative capacities

DRMs that arise from the effects of drug pressure can have pleiotropic effects on the fitness of mutant viruses, which affect viral replicative capacity and derive from numerous factors, including protein stability and in vitro integration activity. Because the effects of the mutations on viral fitness affect which mutations are selected in vivo, we wanted to understand the effects of introducing DRMs on the fitness of the mutant viruses.

We first tested the ability of mutant INs to perform two-ended integration in vitro using double-stranded oligonucleotide mimics of the U5 vDNA end in a strand transfer assay (Fig. 2, A and B). In comparison to the WT enzyme, mutations Q148H/K/R reduced integration activity of IN ~5- to 10-fold and mutations at G140A/S reduced integration activity ~3- to 5-fold. When G140A/S mutations were paired with Q148H/K/R mutations, the reduction in integration typically matched the lowest integration activity for the single mutants, with the sole exception being the double mutant pair G140S/Q148H; when paired with Q148H, but not with Q148K/R, the G140S mutation restored integration by ~3- to 5-fold. We also noted the compensatory behavior of E138K. By itself, this mutation has a minimal effect on integration, but when other changes are present, E138K consistently restores integration activity (e.g., compare Q148K with E138K/Q148K or G140A/Q148K with E138K/G140A/Q148K). The compensatory behavior of E138K is observed across experiments, leading to an increase in the integration activity of all of the triple mutant INs in comparison to their double or single mutant counterparts.

To determine the relative fitness of WT HIV-1 or the IN single, double, and triple mutant viruses in the absence of drugs, we used a single round infection assay (Fig. 2C). Among the IN single mutants, the relative infectivities of E138K, G140A, and G140S were the highest, whereas Q148H/K/R mutants reduced the replicative capacities to greater extents. Our data reproduce prior reports indicating that G140S acts as a compensatory mutation in the context of Q148H (28, 34), and we additionally found a modest (~2-fold) compensation for loss of replicative capacity when G140S was added to Q148K and when G140A was added to Q148R. The addition of E138K as a third mutation to the IN double mutants G140A/S + Q148H/K/R had varying effects in single round replication assays. Whereas the addition of E138K played a compensatory role for the IN triple mutants E138K/G140S/Q148K and E138K/G140A/Q148K, its effect in the context of other mutants was variable. These data suggest that the G140S mutation can have, in some cases, a compensatory effect when combined with a Q148H/K mutation, and the E138K mutation can also

have, in some cases, a compensatory effect when combined with the G140A/S + Q148H/K/R mutations.

In general, we observed a similar decrease in integration activity in mutant INs when the values are compared with replicative capacities of mutant viruses, although these different measures of the effects on IN mutations did not exactly align for all of the DRMs. When the values do not align, we typically observed a decrease in integration but not replicative capacity. It is possible that there is a modest excess of IN enzymatic activity in the virion, so that a small loss of enzymatic activity would not have a significant effect in an infectivity assay. In addition, mutations in IN can quantitatively affect the efficiency of in vitro reactions by mechanisms that might not be relevant to integration in cells. For example, IN tends to aggregate, and the intasome forms nonphysiologically relevant stacks under in vitro reaction conditions (14, 35). Both the propensity of IN to aggregate and form stacks is exquisitely sensitive to reaction conditions and mutations in the IN protein. Thus, the readout from in vitro integration assays does not necessarily quantitatively mirror the intrinsic enzyme activity of IN, which would, in turn, explain the unusually low integration activity in some cases.

Collectively, these data provide insights into the fitness costs associated with single, double, and triple IN mutants and show which mutations compensate for a loss of fitness caused by primary resistance mutations and which potentiate resistance.

Pairwise residue couplings derived from Potts model analyses

We next extended the in vitro measurements, conducted with either purified proteins or in cell culture, to patient sequence data. To this end, we took advantage of the vast amount of IN sequences derived from drug-experienced patients available in the Stanford HIV Drug Resistance Database (36, 37). We use a co-evolutionary (Potts) Hamiltonian model to define the underlying relationships between protein structure, function, and fitness. The Potts Hamiltonian provides a mapping of sequences to statistical scores in which lower Potts statistical energies are more probable, generating a "prevalence landscape" (38, 39). Recent studies have demonstrated that experimental measurements of fitness differences are well correlated with the change in Potts statistical energy for sequences from large Eukaryotic protein families (40, 41). Potts models have also been used to interrogate the fitness landscape of viral systems, most notably HIV-1 proteins escaping immune pressure (38, 42) and HIV-1 enzymes evolving under drug selection pressure (43, 44).

Here, we used the Potts models to provide insights into coupled (paired) mutations, which may arise due to the proximity of the mutations within the three-dimensional (3D) protein structure or their participation in a functional role mediated by IN. The output of the data is a plot of double mutant cycles, which provide a biophysical means to interrogate epistasis (45) (see note S1). We find that the vast majority of mutation pairs ($>10^5$) appear close to the center of the distribution ($\Delta\Delta E = 0$), i.e., these pairs are not epistatically interacting. In contrast, most of the pairs of mutations associated with pairwise combinations of E138K + G140A/S + Q148H/K/R are indicated by the Potts model to be among the most strongly coupled mutation pairs observed for IN (Fig. 2D). The most strongly coupled pair of IN mutations is G140S/Q148H, followed by G140A/Q148K which is also among the very top pairs. In general, the mutations (S and A) at position 140 interact very strongly with the primary drug resistance mutations (H, K, and R) at position 148.

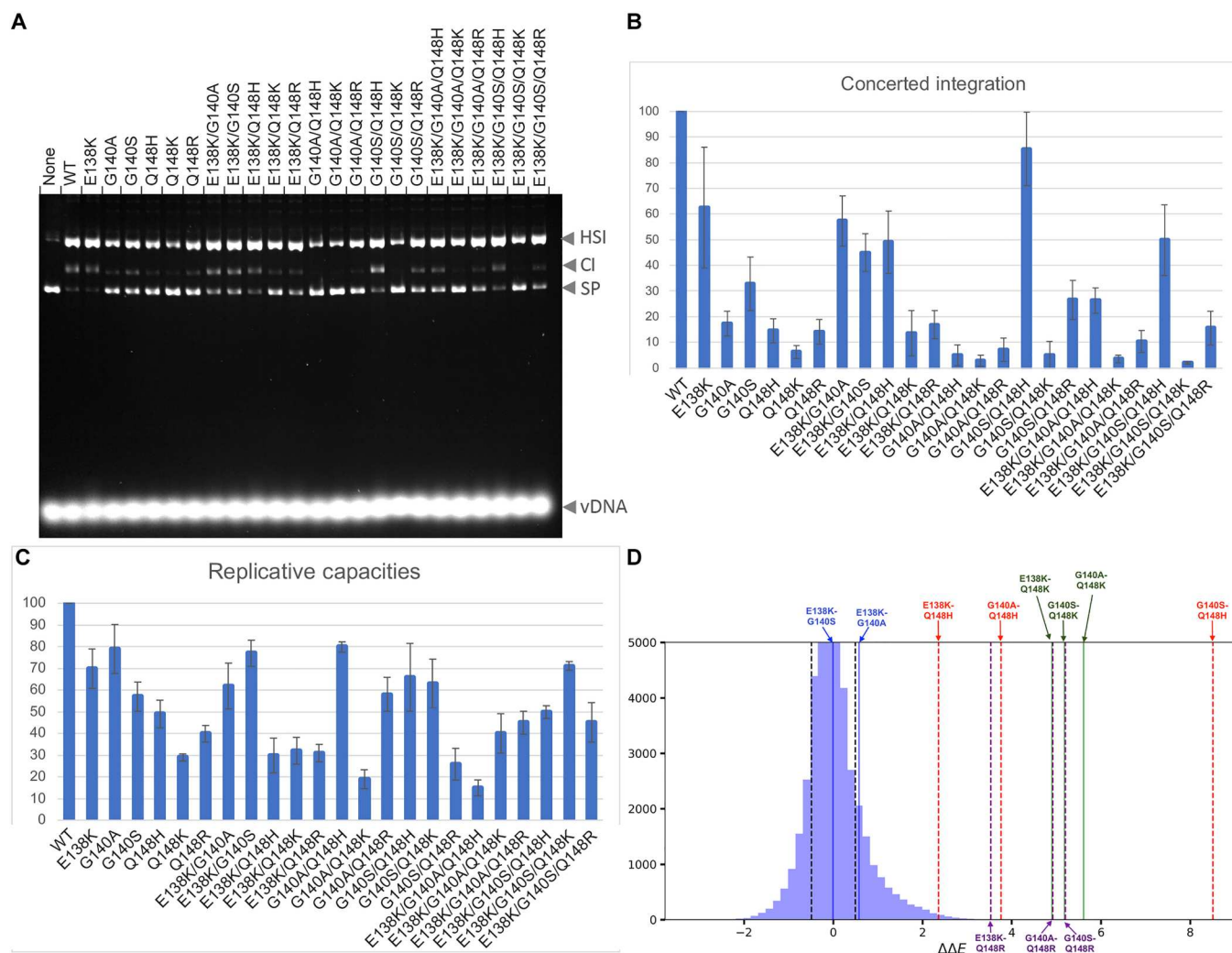


Fig. 2. Enzymatic activities, viral replicative capacities, and coupled residue pairs. (A) Enzymatic catalytic activities, half-site (HS) and concerted integration (CI), for all combinations of single, double, and triple IN mutants corresponding to changes at positions E138K + G140A/S + Q148H/K/R. The bands corresponding to supercoiled plasmid DNA (SP) and viral DNA are also indicated. (B) Bar graph of the two-end integration data in (A). The integration activity of WT IN was set to 100%, and the activity of mutant IN was presented as a percentage of wt IN. The error bars represent SDs of independent experiments, $n = 3$, performed in triplicate. (C) Replicative capacities measured in cell culture of the INSTI-resistant mutants using a single-round infection assay. The replicative capacity of wt IN was set to 100%, and the activity of mutant IN was presented as a percentage of wt IN. (D) Double mutant cycle data derived from Potts statistical analyses. As fitness is inversely proportional to the Potts energy, $\Delta\Delta E > 0$ implies that the mutations are compensatory. The plot shows that many of the residue pairs, most notably G140S + Q148H and G140A + Q148K, are coupled to one another, indicating that compensatory effects between these pairs of positions within the protein drive the observed residue covariation. Residue pairs close to the center of the distribution are not coupled.

Furthermore, the 138K mutation interacts more strongly with the Q148K variant than the Q148H or Q148R variants at position 148, suggesting that E138K plays a stronger role in the E138K/G140A/Q148K pathway than in the E138K/G140S/Q148H pathway. In contrast, the mutation pairs at 138 and 140 are not strongly coupled in any of the pathways. Coupled mutations can arise for different reasons, which may include compensatory effects or cooperative effects of mutations on protein stability, catalytic activity, or response to drug. The double mutant cycle analysis only identifies the strongest epistatic coupling. Such sets of sites are generally more likely to be associated with resistance to the drugs with which the patients were treated, and coupled mutations will

frequently lead to mutational pathways selected for pathogen escape (43).

Structural insights into mechanisms of resistance to DTG for the E138K/G140A/Q148K pathway mutants

Measurements done in cells infected with HIV-1 vectors showed that there is a difference in the potency of DTG against the IN variants G140A/Q148K and E138K/G140A/Q148K in comparison to the rest. The relevance of these mutations is underscored by their appearance in cell culture (27, 46) and in clinical studies (23, 47, 48). We sought to understand the mechanisms underlying these DRMs and the difference in the loss of DTG potency using structural biology. We determined high-resolution structures of the

HIV-1 intasome constructed from all the possible combinations of single and double mutants that underlie E138K/G140A/Q148K, and the triple mutant, with DTG bound. These structures were compared to the structure of the WT HIV-1 intasome with DTG bound. We cloned, expressed, and purified all seven mutant INs (E138K, G140A, Q148K, E138K/G140A, E138K/Q148K, G140A/Q148K, and E138K/G140A/Q148K) and assembled intasomes with DTG bound. Pooled fractions containing intasomes with DTG bound (fig. S2) were vitrified on UltraAuFoil gold grids. We collected cryo-electron microscopy (cryo-EM) datasets for all seven mutant assemblies (table S1). Each cryo-EM dataset was subjected to iterative 2D and 3D classification and refinement, yielding structures resolved to ~2- to 3-Å global resolution, with the highest resolution around the active site, where DTG is bound (fig. S3). The cryo-EM maps demonstrated clear density for all side chains within the intasome core, the ordered water molecules in and around the active site, and, in important cases, the rotameric configurations of the amino acid side chains. The corresponding atomic models for the complete set of mutants with DTG bound was used to develop a detailed understanding the mechanisms of resistance and the contributions of the individual mutations.

The active sites of all experimentally derived mutant structures with DTG bound show the substitutions at positions 138, 140, 148, and the drug DTG (Fig. 3, A to H). The structures also show the position of the nearby residue H114, which is not mutated in these structures but participates in the resistance mechanism (Fig. 3A). For all structures, pink arrows point to the mutated residues, and asterisks highlight the induced structural changes, including new interactions (red asterisks) or alterations to the structure (yellow asterisks).

We began by analyzing the structures of the intasomes with single-residue substitutions, E138K, G140A, and Q148K (Fig. 3, B to D). In the E138K_DTG structure, the ζ -nitrogen of K138 side chain approaches within ~3 Å of the backbone phosphate of the 5'-end of the vDNA, participating in a salt bridge with the phosphate (Fig. 3B). There are no other additional significant changes in the structure. In the G140A_DTG structure, the introduction of a methyl group affects the side-chain orientation of the nearby residue Q148, inducing a metastable preference for an altered rotameric configuration in which there is a change in the rotation of +120° about the γ -carbon (Fig. 3C). The presence of density for the alternative side-chain configuration of Q148 can be clearly seen in the cryo-EM map, which is resolved to ~2 Å in this region (fig. S4). If the position of the side chain of Q148 was not altered, the methyl group of A140 would approach the amide moiety of Q148 to within ~2.5 Å, introducing a weak clash that would destabilize the rotamer that is preferred in the WT intasome bound to DTG. The most marked changes are observed in the structure of Q148K_DTG, in which the side chain of K148 affects multiple nearby residues, leading to: (i) an outward shift of the backbone spanning residues 139 to 141 by up to 1.5 Å, with an root mean square deviation (RMSD) of 1.3 Å between residues 139 and 141 compared to WT; (ii) an induced change in the rotameric configuration of the side chain of the neighboring residue H114 (fig. S5 and Fig. 3D). Collectively, in the presence of bound drug, the most significant changes in the structure are mediated by Q148K. In contrast, G140A and E138K cause less marked changes.

We next analyzed the structures of the double mutants E138K/G140A, E138K/Q148K, and G140A/Q148K (Fig. 3, E to G). In the structure of E138K/G140A_DTG, the changes are largely additive to those found in the corresponding single mutants. The E138K mutation induced salt bridge with the vDNA 5'-phosphate, and G140A induced a rotameric reconfiguration of the side chain of Q148; there was density showing an alternative configuration of the side chain of Q148 (fig. S4). In the E138K/Q148K_DTG structure, the E138K mutation again leads to the salt bridge with vDNA, but the other changes induced by Q148K were less prominent, specifically: (i) the backbone spanning residues 139 to 141 largely reverts to the position seen in the WT intasome, with only minor differences in orientation with respect to WT (RMSD 0.3 Å between residues 139 and 141); (ii) the rotameric configuration of H114 is similar to what was seen in the WT, albeit with a minor rotation of the imidazole ring evident in the double mutant. In turn, the side chain of K148 is oriented slightly toward H114 to avoid clashes (fig. S5 and Fig. 3F). Last, in the G140A/Q148K_DTG structure, there are changes that would directly affect the binding of DTG. The G140A substitution, as has already been discussed, influences the configuration of the residue at position 148. However, because the residue at position 148 is now Lys, steric effects cause the long side chain to reorient +240° about the γ -carbon, positioning the positively charged head group within 2.7 Å of catalytic Glu¹⁵², sufficiently close to form a strong salt bridge interaction, and within 3.6 Å from Mg²⁺ at position A and 3.8 Å from Mg²⁺ at position B (Fig. 3G and fig. S6). This reorientation, in turn, redistributes the charge around the catalytic metal ions and would, collectively, weaken the chelation by the bound DTG (28).

Last, we analyzed the structure E138K/G140A/Q148K_DTG containing all three substitutions (Fig. 3H). As with both the single and double mutants, E138K makes a salt bridge with vDNA. The remaining substitutions, G140A and Q148K, led to changes in the structure that were similar to what was seen for the G140A/Q148K double mutant but with a small change in the orientation of the K148 side chain (fig. S6) and an increase in the distance of the Lys to the two Mg²⁺ ions. In the E138K/G140A/Q148K_DTG structure, the positively charged head group is 4.1 Å from Mg²⁺ at position A and 4.4 Å from Mg²⁺ at position B.

The structures suggest that the G140A mutation potentiates resistance through its effect on Q148K. To gain further insight into this interaction and probe the role of G140A in the mechanism of drug resistance, we performed free energy perturbation (FEP) simulations focusing on the effect of the G140A mutation (E138K/Q148K → E138K/G140A/Q148K) on the potency of the INSTI DTG using an all-atom molecular dynamics (MD) free energy model. In the simulations, we observe that, when the residue at 140 is Gly (i.e., E138K/Q148K), the positively charged amino head group of Lys¹⁴⁸ fluctuates between ~7.6 and ~8.5 Å from the two Mg²⁺ ions. In contrast, when the residue at 140 is Ala (i.e., E138K/G140A/Q148K), there is a corresponding change in the rotamer state of the K148 side chain, and the head group approaches the two Mg²⁺ ions, where its position is stable for the duration of the simulation (Fig. 3I). This approach, which is predicted from simulation and without underlying knowledge of the experimental conformation of Q148K in the structure of E138K/G140A/Q148K (see note S2 and Materials and Methods), supports our structural biology findings and would, in turn, affect the chelation of the Mg²⁺ ions by DTG. Furthermore, on the basis of the FEP

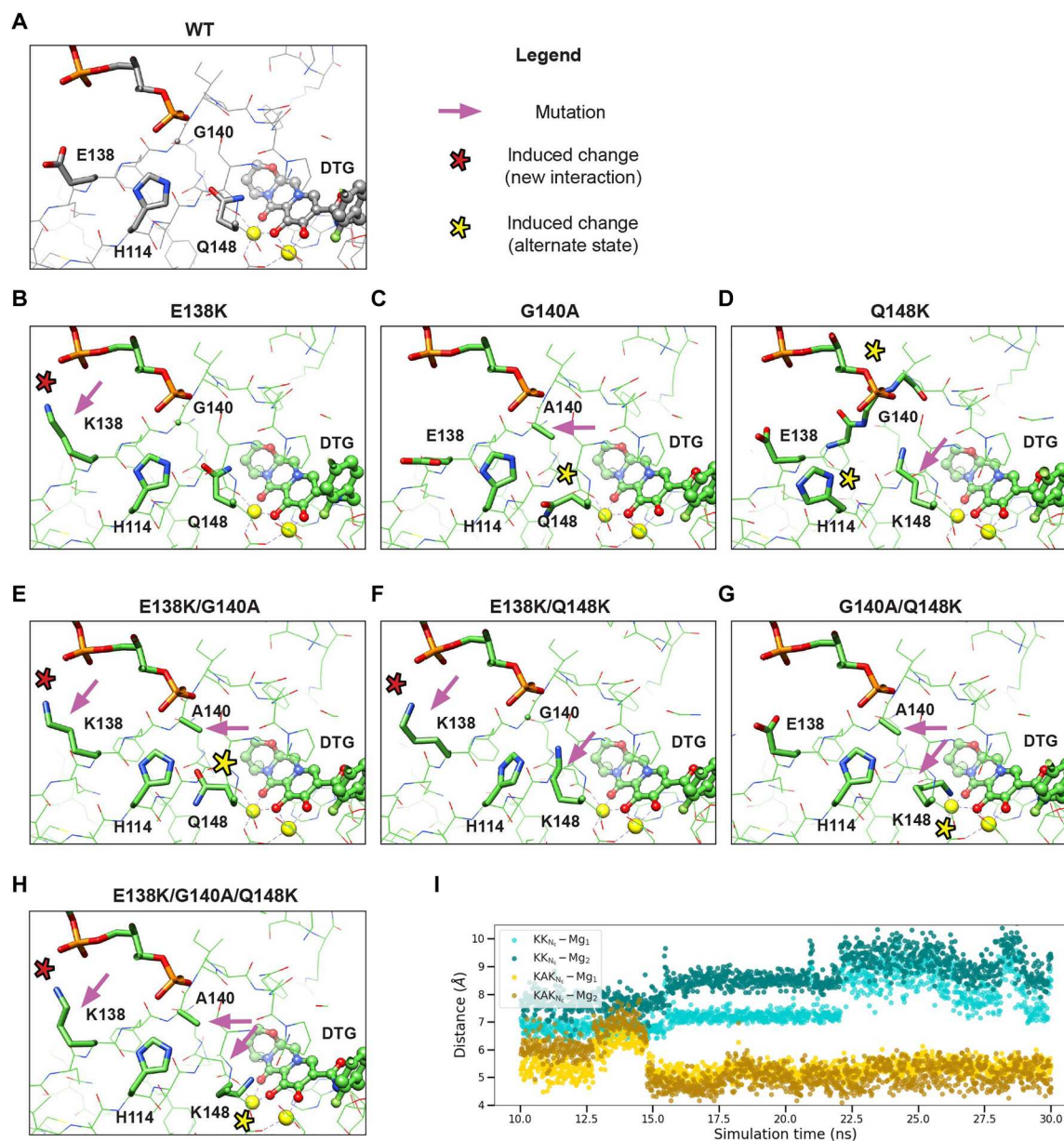


Fig. 3. Drug-resistant mutations E138K/G140A/Q148K induce altered residue configurations in the active site around bound DTG ligand. Atomic models of the active site from high-resolution intasome (WT and mutant) structures bound to DTG are displayed for (A) the WT intasome, (B to D) the single-mutant variants (B) E138K, (C) G140A, (D) Q148K, (E to G) the double-mutant variants (E) E138K/G140A, (F) E138K/Q148K, (G) G140A/Q148K, and (H) the triple-mutant variant E138K/G140A/Q148K. In the legend, pink arrows indicate the positions of the mutations, while the asterisks refer to the induced changes, red indicating a novel interaction and yellow indicating an alteration of the configuration of the residue. (I) MD simulation of E138K/Q148K (KK) and E138K/G140A/Q148K (KAK) intasomes with DTG bound. The atomic distances, measured from the positively charged amino head of Lys¹⁴⁸ to either of the Mg²⁺ ions, are indicated for the duration of the simulation.

calculations, the introduction of the mutation G140A in the background of E138K/Q148K is predicted to reduce the binding affinity of DTG by ~1.0 kcal/mol (fig. S7 and table S2), which would translate to a ~10-fold reduction in the EC₅₀ of DTG when comparing E138K/Q148K with E138K/G140A/Q148K (49); this is exactly the reduction in EC₅₀ that was observed experimentally in virology assays in Fig. 1B. In contrast, if the ligand is replaced with 4d, then the change in free energy is within the noise limit (fig. S7 and table S2), which is again consistent with experimentally measured EC₅₀ values in Fig. 1B. Collectively, the all-atom free energy

model simulations successfully account for the effect of the G140A mutation on the change in the binding of DTG to intasomes and highlight how G140A mediates the conformation of Q148K, leading to polarization of the Mg²⁺ ions and weakening chelation by the bound ligand.

Together, the structures and the simulations provide a framework for understanding, at an atomic level, the changes that are observed in the mutations within the E138K/G140A/Q148K pathway when DTG is bound. The E138K mutation induces a salt bridge with the 5'-end of the vDNA, an interaction that is observed in all

structures. This interaction stabilizes the position of the 5'-end of the vDNA and presumably the conformation of the active site, which explains its mild compensatory behavior observed in enzymology and virology assays (Figs. 1 and 2). The Q148K mutation can adapt distinct configurations in the active site because the Lys side chain is flexible and dynamic, and the exact configuration of the Lys side chain is strongly affected by the neighboring residues. By itself, Q148K points toward the backbone residues 139 to 141 and induces multiple, presumably deleterious, changes to the local structure. The accessory mutations E138K or G140A largely (but not completely) restore the WT configuration of the protein backbone at residues 139 to 141 and the rotameric configuration of the side chain of H114. However, they do so in distinct ways: The mutation E138K, through its interaction with vDNA, pulls the peptide "back into position," leading to the reconfiguration of both the backbone and the H114 side chain. In contrast, A140 relieves steric stress by pushing K148 toward the two Mg^{2+} ions. Notably, it is the G140A mutation that most affects the orientation of the side chain at position 148 and thus the mechanism of drug resistance. If residue 148 is Lys, the mutation G140A leads to a repositioning of the side chain by $+240^\circ$ about the γ -carbon, bringing the ζ -nitrogen to ~ 3.6 to 4.4 Å of the two Mg^{2+} ions. This reconfiguration, in turn, polarizes the charge distribution around the metal ions and weakens the chelation with the bound DTG, reduces its binding, and, consequently, leads to drug resistance.

Mechanism of resistance of the Q148H/K/R mutants

We next asked how each of the three different amino acid mutations Q148H/K/R affect the binding of INSTIs. In prior work, the SIV_{rcm} IN was used as a model for HIV-1 IN, and the analysis was limited to the double mutant G140S/Q148H with BIC bound (28). Here, we determined structures of the HIV-1 intasomes E138K/G140A/Q148R and E138K/G140S/Q148H, with DTG bound, resolved to 2.5 and 2.2 Å, respectively (table S1). These structures were compared to the structure of E138K/G140A/Q148K intasome with DTG bound and to the structure of the G140S/Q148H SIV_{rcm} intasome with BIC bound.

The most notable observation in these structures is the relative position of the side chain at position 148 and its relationship to the two Mg^{2+} ions. In the structure of the E138K/G140A/Q148K intasome with DTG bound, K148 approaches the Mg^{2+} ions to within ~ 4.1 to 4.4 Å, as discussed above (Fig. 4A). In the structure of E138K/G140A/Q148R intasome with DTG bound, the guanidium group resides in a similar overall configuration relative to the Mg^{2+} ions, but the side chain is approximately ~ 2.5 Å longer, which positions the head group underneath the bound INSTI and further from the Mg^{2+} ions (5.2 Å), as measured from the ζ -carbon where the positive charge is localized (Fig. 4B). The longer side chain affects the position of residues 144 to 147, leading to a small local displacement of this segment (the RMSD is 0.3 Å relative to WT; fig. S8). The displacement of residues 144 to 147 is similar when the E138K/G140A/Q148R intasome is compared to both E138K/G140A/Q148K (RMSD 0.3 Å) and E138K/G140S/Q148H (RMSD 0.4 Å). The Q148R mutation causes a weakening of the chelation by DTG, which, when combined with other mutations, also weakens the binding of the drug. In contrast to the two structures described above, the side chain of His is shorter in the E138K/G140S/Q148H intasome, and accordingly, the His side chain is also furthest from the Mg^{2+} ions (Fig. 4C).

To compare the distinct systems used to examine how INSTIs bind to intasomes, we next compared the structure of the E138K/G140S/Q148H HIV-1 intasome with DTG bound to the structure of G140S/Q148H SIV_{rcm} intasome with BIC bound, which was solved previously (28). When the inner catalytic core domains are aligned to one another, there are small differences in the two structures (fig. S9, A to C), and nine residues within 10 Å of the bound drugs differ between HIV-1 IN (NL4-3) and SIV_{rcm} IN (fig. S9D). Because certain DRMs may be better represented by either SIV_{rcm} or by the NL4-3 laboratory isolate of HIV-1, these distinct structures thus provide complementary model systems to study drug resistance. If the E138K mutation is present in the HIV-1 intasome, the side chain of H114 rotates toward the S140/H148 pair, inducing an alternate rotamer in S140 compared the SIV_{rcm} intasome. This causes the Ne2 atom of H114 to approach closer to the Nδ1 atom of H148 by 1.0 Å in the HIV-1 intasome. The repositioning of H114 strongly implicates this residue in the resistance mechanism through its interaction with H148, which would facilitate proton transfer to stabilize the buildup of positive charge around the Mg^{2+} ions. The relative positions of the side chains of the mutant residues and their proximity to the Mg^{2+} ions can be seen in an overlay of the three structures (Fig. 4D).

Collectively, these data provide mechanistic insights into how the distinct residue changes Q148H/K/R affect DTG binding. The proton-donating capabilities of the three side chains are Arg \rightarrow Lys \rightarrow His. Whereas Arg is the most potent proton donor, it is also more distant from the Mg^{2+} ions in comparison to Lys, which explains why its effect on resistance, at least in the context of accessory mutations E138K/G140A, is less than that of Lys at the same position. The His-mediated polarization of Mg^{2+} ions is expected to be weakest, and its effect on the binding of INSTIs is also expected to be less. Thus, it is both the position and the charge of the side chains at residue 148 relative to the Mg^{2+} ions that determines the effects of the mutation on drug binding. Given that the chemical scaffolds of DTG and BIC are similar, as are the inhibitory potencies, our findings can reasonably be extended to BIC. We also expect similar rules to apply to CAB, although the effects will be more pronounced, because this drug is less potent against many DRMs (22). Furthermore, as changes to residues G140A/S + Q148H/K/R are some of the most prevalent in INSTI-resistant HIV-1 IN (23), the current structures should also provide a valuable foundation for future drug design efforts aimed at combatting resistance.

Stacking of the terminal adenine base of vDNA on the INSTI contributes to the potency of compound 4d against the E138K/G140A/Q148K family of DRMs

We next wished to determine why the naphthyridine-containing compound 4d maintains potency against most of the DRMs we tested, particularly the triple mutant E138K/G140A/Q148K, when DTG does not (Fig. 1). We determined the structure of the E138K/G140A/Q148K mutant intasome from HIV-1 with bound 4d at 2.75 -Å resolution and compared it with the structure of the same mutant with DTG bound (table S1).

We first compared the configuration of the ligand when either DTG or 4d is bound to E138K/G140A/Q148K HIV-1 intasomes. Both compounds contain a di-fluorinated benzene that makes a π - π stacking interface with the base of the penultimate cytosine of the vDNA. The fluorinated benzene is connected to the INSTI core through an amide-containing linker that is identical in both DTG

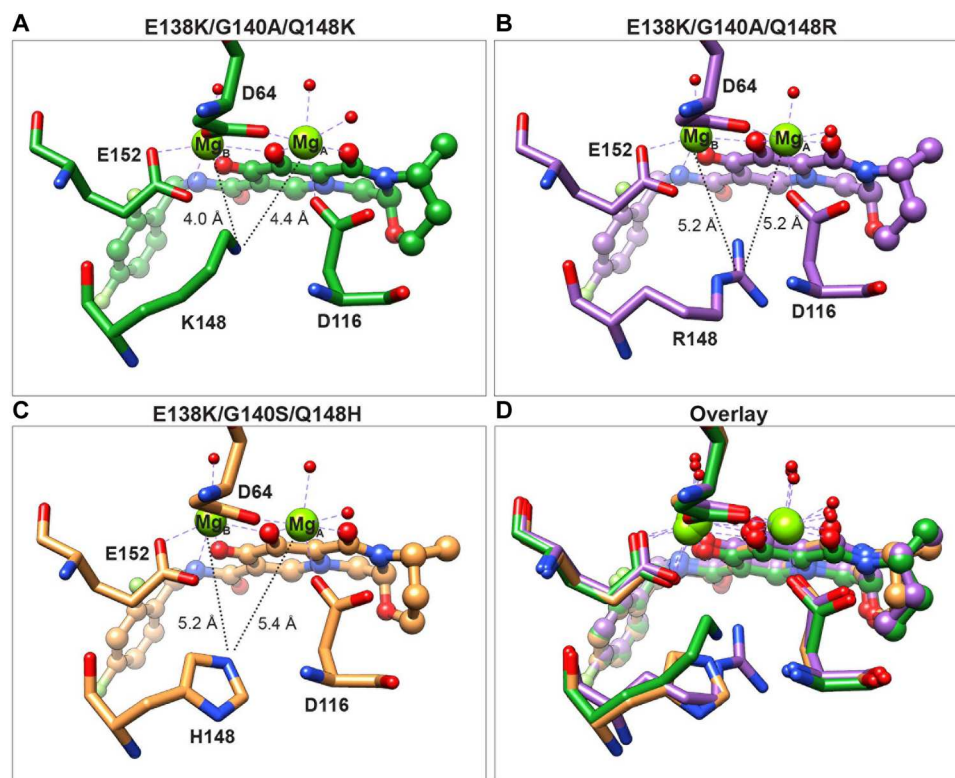


Fig. 4. Structural insights from comparisons of Lys, Arg, and His mutations at position 148. Atomic models of the region surrounding the bound DTG from experimental structures of (A) E138K/G140A/Q148K, (B) E138K/G140A/Q148R, and (C) E138K/G140S/Q148H intasomes with DTG bound. (D) Overlay of the three experimentally determined structures. Relevant atomic distances are indicated in the snapshots.

and 4d, and contacts between the first ring of both ligands to the intasome are similar. The major differences between the compounds are in the way they chelate Mg^{2+} ions and the nature of the extension at the solvent-end of the active site (Fig. 5, A to C). Whereas DTG uses three oxygen atoms—all external to the ligand core—to coordinate the two Mg^{2+} ions (Fig. 5A), 4d has a nitrogen embedded into its naphthyridine core (Fig. 5B). This leads to an overall $\sim 10^\circ$ rotation of 4d toward the Mg^{2+} ions (Fig. 5C). 4d has a hexanol moiety appended to the 6-position of the naphthyridine scaffold instead of the 4-methyl-1,3-oxazinanone ring of DTG. This hexanol displaces weakly bound waters in the apo structure and extends out toward the solvent-exposed cleft, sandwiched between N117-G118 and P142-Y143, a configuration that is expected to stabilize the binding of the ligand to HIV-1 intasomes containing DRMs. Furthermore, when 4d is bound to E138K/G140A/Q148K intasomes, relative to the structure with DTG bound, Tyr¹⁴³ extends outward by 1.5 Å, a result of the hexanol extension making van der Waals contacts in this region (Fig. 5C).

We next examined the structural configurations within the intasome with either DTG or 4d bound. There are no significant differences in the position of the mutated amino acids E138K, G140A, or Q148K, and the free amine on the side chain of K148 occupies a similar position (<0.2 -Å difference) relative to the Mg^{2+} ions (fig. S10A). There are minor variations in the two models in the positions of S119, R228, R231, and other small amino acids, but the significance of these ancillary changes is unclear (Fig. 5C). There is some rearrangement of the hydration shell around the ligands,

while the three core bound waters underneath the ligand and the fourth water, which is displaced if the mutation Q148K is present, are entirely conserved (fig. S10B). The major difference, at the level of the intasome, is in the configuration of the terminal adenosine of the vDNA. This adenosine, which is universally conserved in the genome of retroviruses, can adopt two distinct conformations in the active site: (i) one that stacks the adenine base on the bound INSTI, and (ii) one in which the base extends freely into the space between the IN subunits. Quantification of the relative occupancies of each conformation revealed a strong preference for the stacked conformation when 4d is bound (Fig. 5D). A similar stacked conformation is a minority population in all mutant structures with DTG bound (Fig. 5E), suggesting that the high fractional occupancy of the stacked conformation when 4d is bound is specific to this INSTI and helps maintain the drug bound in its preferred position in the presence of the DRMs.

Stacking of the terminal adenine establishes a π - π bond reminiscent of the one between the halogenated benzene and the base of the penultimate cytosine (Fig. 5F). The stronger stacking of 4d with the terminal adenosine is expected to stabilize the binding of 4d relative to DTG (fig. S11). We thus wished to estimate how stacking affects the potency of 4d. To this end, we can use a simple thermodynamic cycle as shown in Fig. 5G. In this schematic, the ligand is either DTG or 4d, and the target is the adenine base; the ligand binding free energies for the stacked configuration ΔG^S , the nonstacked configuration ΔG^{NS} , and the transition $\Delta G^{S \rightarrow NS}$ complete a thermodynamic cycle that describes the relationships between the distinct

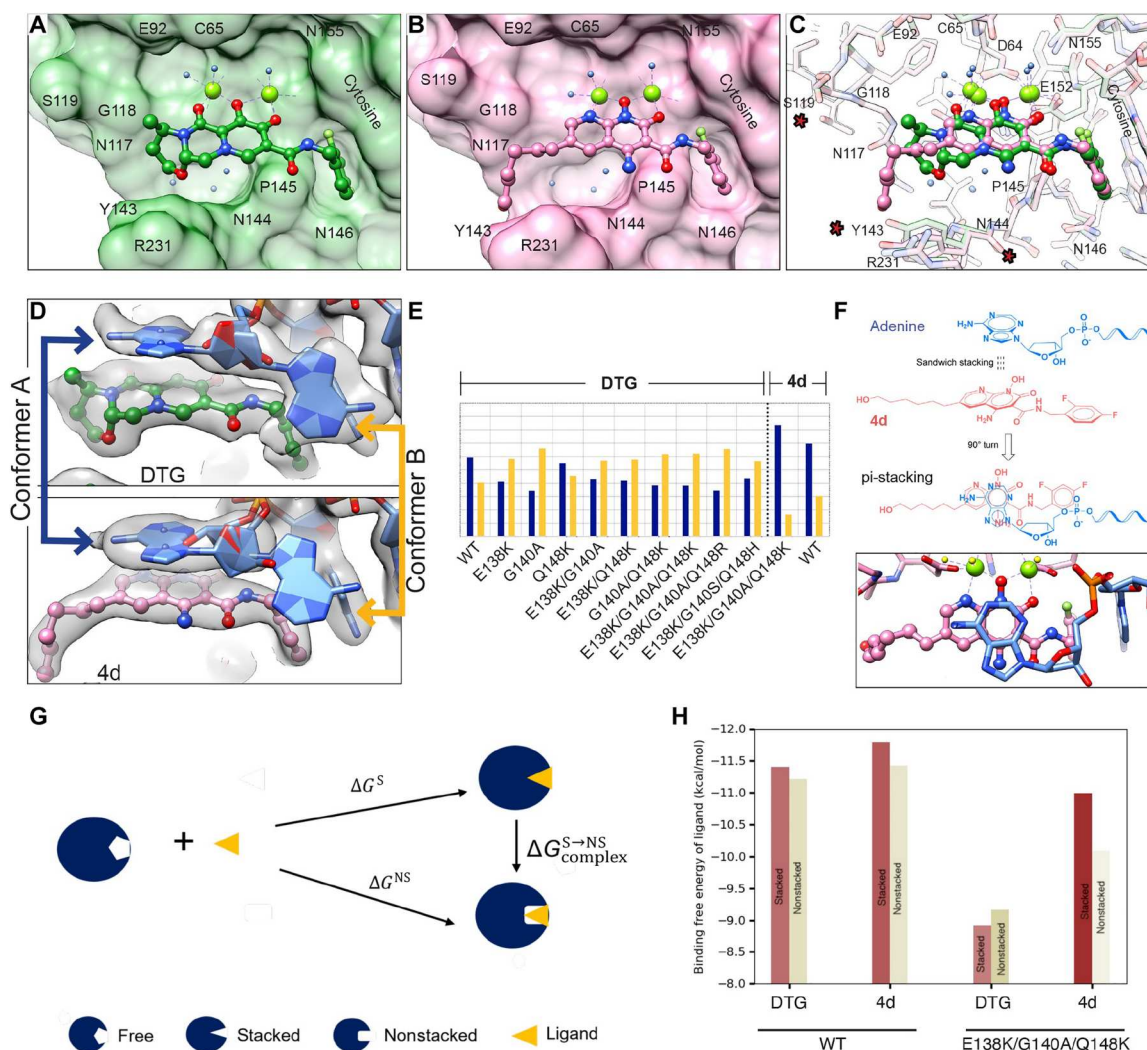


Fig. 5. Structure-based explanation for 4d potency. (A and B) Atomic models of (A) DTG or (B) 4d from high-resolution structures of the E138K/G140A/Q148K intasome bound to either of the two drugs, respectively. The intasome is displayed as a surface view, while the drugs are displayed in ball-and-stick. The terminal adenine base is removed, for clarity. (C) Overlay of the two structures in (A) and (B), with all atoms displayed as ball-and-stick. Residues whose positions vary between the two structures are highlighted with a red asterisk. (D) Close-up of the cryo-EM density of the terminal adenosine and its interaction with the bound drug. When the intasome contains the E138K/G140A/Q148K mutation, experimental density is weak for conformer A, which stacks on top of the bound INSTI, when the ligand is DTG, but is pronounced when the ligand is 4d. (E) Quantitative comparison of the rotameric occupancies of the two conformers, A and B, in all structures derived in the current work. Only the structures in which 4d is bound clearly show pronounced conformer A occupancy. (F) Chemical schematic of the π - π stacking between the terminal adenine base and 4d, which helps to maintain the bound drug in the active site of the intasome. A top-down view of the π - π stacking within the structure is shown below. (G) Thermodynamic cycle used for estimating the ligand binding free energies to the target in the two conformers. (H) Estimates of the free energies for the stacked (brown) and the non-stacked (tan) configurations of the terminal adenine base, for WT (left) and E138K/G140A/Q148K (right) DRM intasomes.

conformational states. In the context of this thermodynamic cycle, the stacking probabilities of 4d and DTG, determined from the cryo-EM maps, and the EC_{50} values, determined using virology assays, provide experimental constraints that can be used to solve a system of equations to yield binding affinities in the stacked and in the nonstacked conformations for 4d and DTG to the E138K/G140A/Q148K triple mutant (see note S3). The inputs and the resulting binding affinities of 4d and DTG to the stacked and nonstacked forms of the E138K/G140A/Q148K intasome complex are listed in table S3. We find that, while the binding free energies of the adenine-DTG conformers to the E138K/G140A/Q148K intasome are similar (both stacked and nonstacked are ~ -9 kcal/mol), the

stacked adenine-4d conformer shows a more favorable binding free energy of ~ -11 kcal/mol, and this conformer makes the dominant contribution to the binding strength (Fig. 5H). Furthermore, the difference in binding free energy of ligand between 4d and DTG for the E138K/G140A/Q148K intasome complex is higher in the stacked conformation (~ -2.1 kcal/mol) versus the nonstacked conformation (~ -0.9 kcal/mol), indicating that stacking is especially relevant in the context of the E138K/G140A/Q148K DRM (table S3). This observation is consistent with structural biology, wherein the conformation of 4d, in the context of the E138K/G140A/Q148K intasome, is displaced toward the adenine base, which is caused by the Q148K mutation “underneath” the ligand;

this same displacement is not observed when DTG is bound (fig. S11). Thus, adenine stacking onto 4d is a contributing factor to the increased potency of 4d as compared with DTG and should be considered in future design strategies to optimize INSTI binding to DRM intasomes.

DISCUSSION

Mechanistic insights into loss of DTG potency revealed through mutations containing G140A/Q148K

Our efforts to understand the effects of the mutations E138K + G140A/S + Q148H/K/R revealed specific weaknesses in the ability of DTG to inhibit certain complex DRMs. Specifically, if the G140A mutation is present in combination with Q148K, as in the double mutant G140A/Q148K or the triple mutant E138K/G140A/Q148K, then the inhibitory potency of DTG drops by ≥ 1 order of magnitude (Fig. 1). Thus, we wished to define the mechanism by which viruses containing DRMs G140A + Q148K become resistant to DTG. Our structures explain why high-level resistance to DTG requires the G140A mutation, and its specific pairing with Q148K. Loss of DTG potency can be primarily attributed to the introduction of positive charge around the Mg^{2+} ligands by K148. Although E138K can potentiate resistance, its effects are modest. In contrast, G140A strongly potentiates resistance. This reduction in potency stems from the effect that G140A has on inducing a reorientation of K148 toward the Mg^{2+} ions and charge redistribution in the surrounding region (Fig. 3). Thus, the proximity of the charged ζ -amine on K148 to Mg^{2+} is directly related to the effect of A140. On the basis of inhibitory potencies, the same logic can be extended to the mutation G140S.

The structural changes induced by K148 potentiate resistance, but they also incur a fitness cost. For a resistant virus to be selected by a drug, the gain from the development of resistance must be greater than the loss incurred by the reduction in replicative capacity. Both the infectivity of the virus and enzymatic activity of IN are substantially reduced in viruses containing the Q148K mutation. This is readily explained; the bulky charged Lys at position 148 intrudes into the active site of IN, impairing its activity (Fig. 2, A to C). Despite their detrimental effect on fitness, DRMs containing G140A/Q148K combinations appear in patients (23, 47, 48) demonstrating their clinical relevance. Together, the changes induced in the structure of the intasome make viruses containing G140A/Q148K or E138K/G140A/Q148K mutations less susceptible to inhibition by DTG, but the mutant INs are still able to mediate integration of vDNA into host DNA, explaining why they are found in both cell culture and PLWH.

The proximity of the amino acid side chain to the Mg^{2+} ions and the strength of the proton-donating capabilities of the head group defines the potency of resistance for Q148H/K/R variants

There are three DRMs frequently encountered at position 148: His, Lys, or Arg. Variants containing Lys, including the double mutant G140A/Q148K and the complex mutant E138K/G140A/Q148K, are the least susceptible to inhibition by DTG (Fig. 1). Structural biology shows that the drop in susceptibility to DTG can be primarily explained by the proximity of the NH_2 moiety of the flexible Lys side chain to the two Mg^{2+} ions. In the presence of bound DTG, Lys is positioned closer than both the Arg and the His variants by ~ 1.0

Å (Fig. 4), which explains the ~ 10 -fold loss of DTG potency in the presence of G140A/Q148K mutations. Although Arg is further from the Mg^{2+} ions, it is a stronger proton donor, and thus, the Arg would also be expected to have a significant effect on the polarization around the Mg^{2+} ions, weakening the chelation by the bound drug. We observe loss of potency of DTG against the complex mutation E138K/G140A/Q148R (Fig. 1). We do not observe loss of potency of DTG against G140A/Q148R intasomes, which implies that the Arg is slightly reoriented in the double mutant in comparison to the triple mutant. If the INSTIs under consideration are DTG and 4d, then the least effective substitution at position 148 is His. Both compounds potentially inhibit all combinations of DRMs (up to three changes) containing the Q148H mutation (Fig. 1). This is consistent with the larger distance from the side chain of His to the Mg^{2+} ions. In sum, it is the specific configuration of the side chain of the residues relative at position 148 to the two Mg^{2+} ions that determines the effects of the mutation on ligand binding.

Despite their relatively mild effect on ligand binding, variants containing Q148H arise and are prevalent in patients being administered either first or second-generation INSTIs (23, 47, 48). The broad persistence of Q148H variants may be due, at least in part, to the fact that they are readily selected by the first-generation INSTIs, which have been used clinically since 2007. The propensity to select mutants that carry the G140S/Q148H mutations may be due to the relatively high replicative capacity and enzymatic activity of the G140S/Q148H variants in the absence of any drugs, supported here by *in vitro* enzymatic and single cycle replication assays (Fig. 2, A to C). Thus, while the strength of the positive charge and the proximity to the Mg^{2+} ions can explain the potency of resistance for the Q148H/K/R group variants, their appearance in the clinical setting and persistence in patient populations depend on which INSTIs are used to select the mutants, on the accompanying mutations at positions 138 and 140 (among other positions), and on how the different mutants affect the complex interplay of these two factors with the collection of viruses present in the infected individual. Given that Q148K/R mutants are less susceptible to inhibition by DTG, we would expect mutants containing Q148K/R to begin appearing more frequently in patients being administered DTG and, by extension, other second-generation INSTIs.

Preferences for G140A/S are influenced by the mutational context of Q148H/K/R

The residue variant at position 140, either Ser or Ala, plays an important role in modulating resistance. On the basis of "intrinsic" fitness constraints identified using Potts double mutant cycle values, which reflect internal constraints in the enzyme that apply in any environment in the Potts framework, the most strongly interacting pair of IN mutations is G140S/Q148H, which is followed by G140A/Q148K, and then the cluster of residue pairs that include G140S/Q148R, G140S/Q148K, G140A/Q148R, and G140A/Q148H (Fig. 2D). Our data, generated using multiple experimental systems to measure inhibitory potencies, fitness, enzyme activity, and atomic structure, help to explain why S140 appears to arise in conjunction with Q148H, whereas both A140 and S140 can arise in conjunction with Q148K/R. In the case of the double mutant G140S/Q148H, based on structure, S140 directly interacts with H148 via the polar hydroxyl group of Ser and the N δ 1 atom of His. As previously suggested (28), this interaction may potentiate

resistance by contributing to the proton-donating capability of this pair of histidines, although our data suggest that this effect is weak (Fig. 1). The presence of G140S restores enzymatic activity and viral fitness (Fig. 2, A to C), supporting its previously described role as a compensatory mutation (34). The compensatory effect appears to be caused by S140 forming hydrogen bonds with both H148 and T115, which creates a favorable energetic environment around the active site when the mutation Q148H is present in conjunction with G140S. In contrast to G140S, the mutation G140A does not have free protons to donate, and it should not contribute favorably to the energetic environment. Consistently, G140A, by itself or in combination with other DRMs, generally does not restore enzymatic activity or viral infectivity (Fig. 2, A to C). Instead, it modulates the conformation of residue 148 when a drug is bound, positioning the residue closer to the Mg^{2+} ions. This is seen in the structures of Q148K-containing intasomes (Fig. 3) and, based on comparisons of inhibitory potencies, is implied for Q148R-containing intasomes. Because the mutation pairs G140S/Q148K, G140A/Q148R, and G140S/Q148R arise nearly as frequently as the mutation pair G148A/Q148K (Fig. 2D), and there are no distinct groupings that can be discerned based on measurements of inhibitory potencies, enzyme activity, or viral replicative capacities, we conclude that, in the presence mutations Q148K/R, the variant G140S should function analogously to the variant G140A. Collectively, these data indicate that the selection of G140A/S depends on the substituent at position 148 and that preferences for certain combinations of mutations can, to some extent, be explained by the structures. However, a complete understanding of exactly how the various combinations of mutations interact and how such interactions will affect the structure and function of IN would require considering the complex interplay between the virus, the specific INSTIs used to treat PLWH, including the background mutational landscape of the virus and the relative fitness of the viruses in the presence of the different drugs.

Mechanism of action of E138K in drug resistance and functional compensation

Both the E138K and E138T mutations occur in INSTI-resistant HIV-1 strains. The E138K mutation can reduce the potency of DTG and, in at least some mutant backgrounds, plays a compensatory role, restoring enzyme activity and viral fitness. The E138T mutation is known to reduce INSTI potency when present in conjunction with G140S/Q148H (28). Prior work suggested that the mutation E138T in SIV_{rcm}, which resides in the same general region as E138K/T in the NL4-3 isolate of HIV-1 (fig. S9), potentiates resistance by serving as an H-bond donor to His114 within an extended network that reinforces the positive charge around DTG, weakening its chelation of Mg^{2+} (28). The residues Thr and Lys have different chemical properties, and the role of the E138K mutation in potentiating resistance appears to be different from E138T, based on multiple observations. First, the distance between Lys¹³⁸ and His¹¹⁴ always ranges from ~6 to 8 Å, which is too large to form an H-bond network. Second, the mutation G140A, which can arise in conjunction with E138K, lacks an H-bond acceptor, but the aliphatic side chain of Ala, like Ser, strongly reduces the potency of DTG (Fig. 1). Both observations would preclude the establishment of an H-bond network. Instead, we suggest that the mutation E138K, which invariably establishes a salt bridge with the vDNA, induces micro-rearrangements of the protein and solvent which, collectively, lead to

small but significant changes around the active site that can reposition the DRM at position 148, thereby leading to drug displacement. The compensatory behavior of E138K may stem from stabilizing the binding and/or proper positioning of the vDNA. The strand transfer step of integration is a dynamic process that requires careful coordination of multiple substrates and the sequestration of the 3'-OH moiety on the vDNA for S_N2-mediated nucleophilic attack (5, 50). Many mutations are known to affect the strand transfer reaction and, presumably, the conformation and dynamics of the vDNA. A salt bridge between Lys¹³⁸ and the vDNA phosphate backbone should help stabilize the position of the 5'-end of vDNA, especially when other DRMs are present. Our data show that the E138K mutation could be selected in the presence of several different mutant backgrounds, suggesting that this is a general compensatory mechanism that may be encountered when multiple DRMs accumulate.

Potency of 4d and the naphthyridine compounds

Loss of potency of INSTIs that is caused by mutations at positions E138K + G140A/S + Q148H/K/R can be addressed through the development of better compounds. Select chemical modifications on the ligand, such as the primary amine at the 4-position of the naphthyridine scaffold (51), allow 4d to retain better binding and potency than DTG (21). Such prior observations suggested that introducing specific chemical groups could enhance interactions with the intasome to minimize resistance mediated by the Lys or Arg variants at position 148. In this regard, the consistent observation of stacking was mediated by the terminal adenine base of the vDNA, and its predominance when the INSTI 4d is bound was encouraging. The fact that the stacking is more pronounced when 4d is bound than when DTG is bound (and, by extension, when BIC and CAB are bound) is chemically consistent with the fact that the π - π interactions are expected to be stronger against the naphthyridine moiety of 4d than against the tricyclic moiety of DTG. Although stacking is not the only contributor to potency, it appears to be an important one. Likely, there are additional means by which to exploit, and perhaps enhance, the stabilization of the terminal adenine base, to generate even tighter binders. It is also relevant to identify novel means to counteract loss of INSTI potency against the prevalent group of Q148H/K/R mutations that affect Mg^{2+} chelation. Future efforts should explore these ideas and aim to design novel compounds that are broadly effective against drug-resistant HIV.

Targeting DRMs for anti-HIV therapy

INSTIs constitute an important arsenal of clinically used drugs to treat HIV. All INSTIs potentially inhibit the WT virus. However, because the virus can rapidly evolve resistance to therapy, INSTIs frequently fail to inhibit DRMs. More generally, drug resistance is one of the major issues afflicting HIV treatment, and the challenges extend beyond INSTIs. Although resistance to second-generation INSTIs is less problematic compared to other antiretroviral drug classes, DRMs still develop and can lead to substantial loss of susceptibility to therapy. One of our objectives is to understand why certain compounds, such as 4d, perform better than the clinically approved drugs against key DRMs. Multidisciplinary analyses, which include structure, biochemistry, virology, and computation, should help dissect the principles that lead to efficacy of some compounds but not others. Having such an understanding would be an important step in the design of novel broadly potent INSTIs, while

the principles could be extended to other drug targets. The need to address drug resistance necessitates explaining the mechanisms through which resistance emerges. These insights could then be harnessed to develop next-generation therapeutics to specifically target DRMs.

MATERIALS AND METHODS

Vector constructs

To make the IN mutants, the vector pNLN_goMIVR-ΔENV.LUC, which has been described previously (22), was used to produce the new IN mutants analyzed in this study. The IN open reading frame was removed from pNLN_goMIVR-ΔENV.LUC by digestion with Kpn I and Sal I, and the resulting fragment was inserted between the Kpn I and Sal I sites of pBluescript KS+. Using that construct as the WT template, the following HIV-1 IN mutants were prepared using the QuikChange II XL site directed mutagenesis kit (Agilent Technologies, Santa Clara, CA): E138K, G140A, G140S, Q148H, Q148K, Q148R, E138K/G140A, E138K/G140S, E138K/Q148H, E138K/Q148K, E138K/Q148R, G140A/Q148H, G140A/Q148K, G140A/Q148R, G140S/Q148H, G140S/Q148K, G140S/Q148R, E138K/G140A/Q148H, E138K/G140A/Q148K, E138K/G140A/Q148R, E138K/G140S/Q148H, E138K/G140S/Q148K, and E138K/G140S/Q148R. The following sense oligonucleotides were used with matching cognate antisense oligonucleotides (Integrated DNA Technologies, Coralville, IA) in the mutagenesis:

E138K, 5'-TGGGCGGGGATCAAGCAGAAATTTGGCATTCCTACAAT-3'; G140A, 5'-GGGATCAAGCAGGAATTTGCTATCCCTACAATCCCCAA-3'; G140S, 5'-GGGATCAAGCAGGAATTTCCATTCCTACAATCCCCAA-3'; Q148H, 5'-CCCTACAATCCCCAAAGTCATGGAGTAATAGAATCTATG-3'; Q148K, 5'-CCCTACAATCCCCAAAGTAAAGGAGTAATAG AATCTATG-3'; Q148R, 5'-CCCTACAATCCCCAAAGTCGA GGAGTAATAGAATCTATG-3'; E138K for E138K/G140A, 5'-GGGATCAAGCAGAAATTTGCCATTCCTACAATCCCCAA-3'; and E138K for E138K/G140S, 5'-GGGATCAAGCAGAAATTTAG CATTCCCTACAATCCCCAA-3'.

The IN mutants E138K, G140A, G140S, Q148H, Q148K, and Q148R were made using the WT template and the respective oligonucleotide listed above. Because of the complexity of the IN mutants (some IN mutants may have two or more IN mutations), certain IN mutants in the KS construct will be used as templates for additional rounds of mutations. The IN mutants E138K/G140A, E138K/G140S, E138K/Q148H, E138K/Q148K, E138K/Q148R, G140A/Q148H, G140A/Q148K, G140A/Q148R, G140S/Q148H, G140S/Q148K, G140S/Q148R, E138K/G140A/Q148H, E138K/G140A/Q148K, E138K/G140A/Q148R, E138K/G140S/Q148H, E138K/G140S/Q148K, and E138K/G140S/Q148R were made as follows. The IN mutant E138K/G140A was made using the previously constructed IN mutant G140A as the template, while the proper E138K oligonucleotide listed above was used to make the second amino acid substitution. The IN mutant E138K/G140S was made using the previously constructed IN mutant G140S as the template, while the proper E138K oligonucleotide aforementioned was used to make the second amino acid substitution. To make the IN mutants E138K/Q148H, E138K/Q148K, and E138K/Q148R, the previously constructed IN mutant E138K was used as the template, and correct oligonucleotide for either Q148H, Q148K, or Q148R was used to make the second amino acid substitution. To make

the IN mutants G140A/Q148H, G140A/Q148K, and G140A/Q148R, the previously made IN mutant G140A was used as a template, and the proper oligonucleotide for either Q148H, Q148K, or Q148R was used to make the second amino acid substitution. For the IN mutants G140S/Q148H, G140S/Q148K, and G140S/Q148R, the previously constructed IN mutant G140S was used as a template, and the correct oligonucleotide for either Q148H, Q148K, or Q148R was used to make the second mutation. For the IN mutants E138K/G140A/Q148H, E138K/G140A/Q148K, and E138K/G140A/Q148R, the previously constructed IN double mutant E138K/G140A was used as a template, and the proper oligonucleotide for either Q148H, Q148K, or Q148R was used to make the third amino acid substitution. For the IN mutants E138K/G140S/Q148H, E138K/G140S/Q148K, and E138K/G140S/Q148R, the previously made IN double mutant E138K/G140S was used as the template, and either the proper oligonucleotide for Q148H, Q148K, or Q148R was used to make the third mutation.

After the mutagenesis, the DNA sequence of each construct was determined. The mutated IN coding sequences from pBluescript KS+ will then be subcloned into pNLN_goMIVR-ΔEnv.LUC (between the Kpn I and Sal I sites) to produce mutant HIV-1 constructs, which were also checked by DNA sequencing.

Cell-based replication assays

Cell-based assays

HIV-based viral vectors with either WT or mutant IN were used in single-round infectivity assays to determine the antiviral activity (EC₅₀ values) of the compounds as previously described (52). Briefly, VSV-g-pseudotyped HIV-1 was produced by transfections of 293 cells with pNLN_goMIVR^ΔLUC and pHCMV-g (obtained from Dr. Jane Burns, University of California, San Diego) using the calcium phosphate method. At approximately 6 hours after the calcium phosphate precipitate was added, 293 cells were washed twice with phosphate-buffered saline and incubated with fresh media for 48 hours. The virus-containing supernatants were then harvested, clarified by low-speed centrifugation, filtrated, and diluted for preparation in antiviral infection assays. On the day before the screen, HOS cells were seeded in a 96-well luminescence cell culture plate at a density of 4000 cells in 100 μl per well. On the day of the screen for antiviral activity infection assays, cells were treated with compounds from a concentration range of 5 to 0.0001 μM using 11 serial dilutions and then incubated at 37°C for 3 hours. After compound incorporation, 100 μl of virus-stock diluted to achieve a luciferase signal between 0.2 and 2.0 relative luciferase units was added to each well and further incubated at 37°C for 48 hours. Infectivity was measured by using the Steady-lite plus luminescence reporter gene assay system (PerkinElmer, Waltham, MA). Luciferase activity was measured by adding 100 μl of Steady-lite plus buffer (PerkinElmer) to the cells, incubating at room temperature for 20 min and measuring luminescence using a microplate reader. Antiviral activities were normalized to the infectivity in cells that featured the absence of target compounds. KaleidaGraph (Synergy Software, Reading, PA) was used to perform nonlinear regression analysis on the data. EC₅₀ values were determined from the fit model.

Single-round infection assays

A modified version of the single-round infectivity assay was used to determine the replication of the INSTI-resistant mutant vectors. Briefly, 200 ng of a WT or INSTI-resistant mutant HIV-1 based

vector was added to each well in 96-well plates, incubated for 48 hours, and luciferase activity was measured as mentioned above. The luciferase activity of the WT virus was set to 100%, and the infectivity of the mutant viruses was measured as a percentage of WT (20).

Surface plasmon resonance

SPR binding experiments were performed using the Reichert 4-SPR. WT PFV intasome containing biotinylated DNA (5' - /5Biosg/TGC GAA ATT CCA TGA CA -3') (Integrated DNA Technologies) was immobilized on the streptavidin sensor chip via biotin-streptavidin interaction. The running buffer contained 0.01 M Hepes (pH 7.4), 0.32 M NaCl, 25 mM MgCl₂, 0.05% (v/v) Tween 20, and 5% dimethyl sulfoxide (DMSO). DTG and 4d were prepared by serially diluting these compounds in 100% DMSO and then by adding running buffer (without DMSO) to reach a final DMSO concentration of 5% and drug concentrations from 15.62 to 250 nM. For each interaction, background binding and drift were subtracted via a reference. Data were analyzed using Clamp XP and fit with a kinetic titration model.

IN protein expression and purification

His-tagged Sso7d-IN and mutants were expressed and purified essentially as described before (14, 15, 35). Briefly, the IN was expressed in *Escherichia coli* BL21(DE3), and the cells were lysed in buffer containing 20 mM Hepes (pH 7.5), 10% glycerol, 2 mM 2-mercaptoethanol, 20 mM imidazole, and 1 M NaCl. The protein was purified by nickel-affinity chromatography, and the His-tag was removed with thrombin. Aggregated protein was removed by gel filtration on a HiLoad 26/60 Superdex-200 column (Cytiva) equilibrated with 20 mM Hepes (pH 7.5), 10% glycerol, 2 mM dithiothreitol (DTT), 1 mM EDTA, and 0.5 M NaCl. The protein was concentrated using an Amicon centrifugal concentrator (EMD Millipore) as necessary, flash-frozen in liquid nitrogen, and stored at -80°C. His-tagged Lens epithelium-derived growth factor-Integrase Binding Domain (LEDGF-IBD) (307-460) was expressed and purified in the similar way except His-tag was not removed from IBD.

Intasome assembly and purification

The double-stranded vDNA substrate was prepared by annealing the following oligonucleotides: 5'-AGCGTGGGCGGGAAAATCTCTAGCA-3' and 5'-ACTGCTAGAGATTTTCCCCGCCACGCT-3'.

Sso7d-IN Cleaved Synaptic Complex (CSC) intasomes were assembled and purified similar as previously described (15). Briefly, 3.0 μM IN and 1.0 μM viral DNA substrate were mixed in the buffer containing 20 mM Hepes (pH 7.5), 25% (w/v) glycerol, 50 mM NDSB-256, 5 mM beta-mercaptoethanol (b-ME), 4 μM ZnCl₂, 100 mM NaCl, 5 mM CaCl₂, or 5 mM MgCl₂ with 50 μM dolutegravir (DTG) or 25 μM 4d. The reaction was carried at 37°C for 60 min and stopped by transferring onto ice, and then 500 mM NaCl was added to assembly reaction mixtures to solubilize intasomes. LEDGF-IBD (10 μM) was added into the supernatant of intasomes and incubated at 4°C overnight. The reaction mix was loaded onto a HisTrap column (Cytiva) equilibrated with 20 mM Tris-Cl (pH 8.0), 5 mM 2-mercaptoethanol, 0.5 M NaCl, and 20% (w/v) glycerol, and the intasomes were then eluted with a linear gradient of 0 to 500 mM imidazole in the same buffer. The fractions

containing intasomes were combined and concentrated using a concentrator (Ultracel-100K, Millipore) and loaded onto Superose 6 10/300 GL (Cytiva) gel filtration column equilibrated with 20 mM bis-tris (pH 6.2), 0.5 mM Tris (2-carboxyethyl) phosphine (TCEP), 600 mM NaCl, 5.0 mM MgCl₂, and 10% (w/v) glycerol. The intasomes were further purified with TSKgel UltraSW in the buffer of 20 mM bis-tris (pH 6.2), 0.5 mM TCEP, 550 mM NaCl, 5.0 mM MgCl₂, and 6% (w/v) glycerol. INSTI (25 μM) was added to the intasomes after every step of purification. CSCs were concentrated to 0.4 mg/ml with spin concentrator for cryo-EM grid preparation. Apo CSCs were prepared in essentially the same way except MgCl₂ and INSTI were substituted by 5.0 mM CaCl₂, and MgCl₂ was omitted in all purification buffers.

IN activity assays

IN (3 μM; unless otherwise noted) and 1.0 μM viral DNA substrate were preincubated on ice in 20 mM Hepes (pH 7.5), 25% glycerol, 10 mM DTT, 5 mM MgCl₂, 4 μM ZnCl₂, and 100 mM NaCl in a 20-μl reaction volume. For the vDNA, fluorescent DNA was purchased from Integrated DNA Technologies and contained a 6-carboxy-fluorescein (6-FAM) fluorophore attached to the 5'-end of the precut oligonucleotide, yielding Fam-AGCGTGGGCGGGAAAA TCTCTAGCA vDNAs. Three hundred nanograms of target plasmid DNA pGEM-9zf was then added, the reaction was initiated by transfer to 37°C, and incubation was continued for 1 hour. For integration product analysis, the reactions were stopped by addition of SDS and EDTA to 0.2% and 10 mM, respectively, together with 5 μg of proteinase K. Incubation was continued at 37°C for a further 1 hour. The DNA was then recovered by ethanol precipitation and subjected to electrophoresis in a 1.5% agarose gel in 1× Tris-boric acid-EDTA buffer. DNA was visualized either by ethidium bromide staining or by fluorescence using a Typhoon 8600 fluorescence scanner (GE Healthcare).

Modeling of double mutant cycles using Potts statistical energy models

The Potts model is a probabilistic model of sequence covariation built on the single and pairwise site amino acid frequencies of a protein multiple sequence alignment (MSA) and is aimed to describe the observation probabilities of the various states of the system (i.e., sequences in the MSA). The Potts statistical energy of a sequence provides an estimate of the probability of observing the sequence, giving a proxy for the fitness of the sequence, and for a sequence *S* of length *L* is given by $E(S) = \sum_{i < j} J_{si sj}^{ij} + \sum_i h_{si}^i$, where the model parameters h_{si}^i called "fields" represent the statistical energy of a residue *Si* at position *i* in sequence *S*, and $J_{si sj}^{ij}$ are "couplings" representing the energy contribution of a position pair *i, j*. The probability of observing the sequence *S* according to the model is given by $P(S) \propto e^{-E(S)}$. The change in Potts energy due to mutating a residue *α* at position *i* in *S* to *β* is then given by

$$\Delta E(S_{\alpha \rightarrow \beta}^i) = \Delta E(S_{\alpha}^i) - \Delta E(S_{\beta}^i) = h_{\alpha}^i - h_{\beta}^i + \sum_{j \neq i}^L J_{\alpha sj}^{ij} - J_{\beta sj}^{ij}$$

In this form, $\Delta E(S_{\alpha \rightarrow \beta}^i) > 0$ implies that the residue *β* is more favorable than residue *α* at position *i* in the given sequence *S*. For the pair of mutations *α* and *β* at positions *i, j* in the protein, the double mutant cycle captures the non-additivity (strength of epistatic

interaction) of the effects of mutations and is given by the difference between the sum of the independent mutational effects, $\Delta E(S_i^a) + \Delta E(S_j^b)$, and the effect of the corresponding double mutation $\Delta E(S_i^a, S_j^b)$: $\Delta\Delta E = \Delta E(S_i^a, S_j^b) - [\Delta E(S_i^a) + \Delta E(S_j^b)]$. If $\Delta\Delta E = 0$, then the mutations are mutually independent, whereas if $\Delta\Delta E \neq 0$, then the two mutations are epistatically coupled; $\Delta\Delta E > 0$ implies that the mutations are beneficial/cooperative to each other and vice versa.

The Potts statistical energy models probe the underlying relationships between protein structure, function, and fitness, and here, they are used to provide insights into the coupled nature of the effects of primary and compensatory residue pairs (see note S1). For a more detailed description of the Potts model in protein biophysics, we refer the reader to (53–57) and, for applications of the Potts model to HIV including description of data processing and model inference, to (39, 43, 44, 58–61).

Cryo-EM sample preparation and data collection

For cryo-EM sample preparation, 3.0 μ l of purified CSC was applied onto freshly plasma cleaned (60s, Pelco easyGlow plasma cleaner) holey grids (UltrAuFoil R 1.2/1.3 300 mesh, quantifoil), adsorbed for 10 s, blotted for 4 s, and then plunged into liquid ethane using Vitrobot V plunge freezer (Thermo Fisher Scientific). Data collections were performed at different locations using a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operating at 300 keV, equipped with either a K2 or K3 direct electron detector (Gatan). Data collection was performed either using Leginon (62, 63) or SerialEM (64), depending on the facility being used. Most of the data collections were performed using stage tilt, respectively. All imaging parameters across the 11 datasets are comprehensively summarized in table S1.

Cryo-EM image analysis

Cryo-EM datasets were processed using a similar workflow. The movie frames were imported to Relion 4.0-beta-2 (65) for dose-weighted motion correction (66) on 5 by 5 patch squares using a B-factor of 150 \AA^2 . The gain reference used for Relion motion correction was generated by using the Sum_all_tifs program, which is part of the cisTEM image processing suite (67). The motion-corrected micrographs were then imported into Warp 1.0.9 (68) to perform contrast transfer function (CTF) estimation and particle selection. The particles that scored above 0.6 were selected in Warp with a retrained BoxNet model using constraint settings of particle diameter of 180 \AA and a minimum distance between particles of 20 \AA . Particles were initially extracted in Warp using a box size of 384 pixels. The motion-corrected micrographs in Relion and particle star file generated in Warp were then imported into cryoSPARC V3.3.2 (69) to continue data processing. The particle coordinates indicated in the particle star file were reextracted in cryoSPARC with the same box size of 384 pixels and were used to perform 2D classification. The classes containing the best particles were selected after several rounds of 2D classification based on particle features and visual inspection of the 2D class averages. Following 2D classification and selection of the best classes, the remaining particles were subjected to heterogeneous refinement with C2 symmetry to further clean the dataset and using a previously published map of the HIV-1 intasome bound to the INST1 4d as a reference (EMD_20484) (15). Multiple rounds of heterogeneous refinement

were performed until no further improvements in resolution were observed, as determined by the Fourier shell correlation (FSC). At each round, particles corresponding to the high-resolution map were selected, whereas the remaining particles were removed from processing. Following heterogeneous refinement, the remaining particles were used for homogeneous refinement in cryoSPARC, which generally produced reconstructions resolved to $<3 \text{ \AA}$, as determined using the FSC criterion (70) using a fixed threshold of 0.143. At this point, the remaining particles were imported back to Relion for Bayesian polishing using default parameters. Subsequently, the polished particles were reimported back to cryoSPARC to perform one round of nonuniform refinement (71), which was followed by one round of global CTF refinement and one round of per-particle CTF refinement. We repeated the procedure iteratively—Bayesian polishing in Relion followed by nonuniform refinement and global and per-particle CTF refinement in cryoSPARC iteratively, until no further improvements in resolution were observed, as determined using the FSC. In general, the optics parameters, per-particle defoci, and the map resolutions converged after several iterations of polishing and nonuniform and CTF refinement. The final map was generated by the last nonuniform refinement. The directional resolution of the map was evaluated using the 3D FSC server (3dfsc.salk.edu) (72), and the quality of the orientation distribution was evaluated using the sampling compensation function (73, 74). All images were generated using UCSF Chimera (75).

Atomic model building and refinement

Modeling of HIV-1 intasomes containing drug-resistant variants was initially focused on deriving a single accurate unhydrated atomic model. The initial reference used to derive an accurate model of the nucleoprotein components was the structure of the cleaved synaptic complex intasome bound with magnesium and the INST1 4d, resolved to 2.8 \AA [Protein Data Bank (PDB) 6PUY]. A baseline atomic model was derived for the structure of the G140A intasome bound to DTG, which corresponded to one of the highest resolution maps, at 2.2 \AA . Mutations corresponding to the drug-resistant variants were generated in silico, the coordinates of the ligand (DTG) were added to the model, restraint files were generated using phenix.elbow (76), and the model was manually adjusted in Coot (77, 78) wherever discrepancies were apparent with the high-resolution map. The final model was derived through iterative model building using Coot and real-space refinement using phenix.real_space_refine (79). The same procedure was then adapted to all other map/model combinations, using the model of the G140A intasome bound to DTG as a baseline.

Atomic models were individually hydrated to account for variances in their global resolution. We determined an estimate for the number of waters that we should observe based on an analysis of Carugo and Bordo (80). Before dousing, cryo-EM density reconstructions were resampled and resized such that the voxel size became roughly one quarter of the resolution estimate using software available in cisTEM (67). We next used phenix.douse, available from Phenix version 1.20.1-4487 (81) to programmatically add water molecules to the density that was not accounted by the protein or DNA models. During dousing, we used the same settings for each of the models – `dist_min = 2.5` `dist_max = 4.5` `keep_input_water = TRUE` `sphericity_filter = false` – and varied the `map_threshold` value to add the number of waters expected based

on the resolution of the map and the number of atoms in the model according to analysis of Carugo and Bordo. After dousing, the models were subjected to a round of real space refinement with phenix.real_space_refine. Each map was then manually inspected to evaluate agreement of water placement with cryo-EM densities. Problematic water molecules were identified using validation tools in Phenix (82) and MolProbity (UnDowser) (83) and removed if they (i) clashed with protein molecules, (ii) were placed in density that belong to other protomers of the dodecamer, or (iii) were placed errantly without corresponding density. Dousing was performed using an asymmetric unit of the intasome, and twofold symmetry was applied at the end. The geometry of the final symmetrized models and other validation statistics were reported by MolProbity. The relevant refinement statistics are summarized in table S1. High-resolution structural figures were prepared for publication using UCSF Chimera (75).

Molecular dynamics and FEP system setup and simulation details

For MD-based FEP simulations, the starting model was an experimentally determined structure of the E138K/Q148K HIV intasome with DTG bound. The structure of the E138K/Q148K-4d complex used in the FEP simulation is obtained from E138K/Q148K-DTG but replacing DTG with 4d from the WT intasome with 4d bound (PDB 6PUY) (15). The E138K/Q148K-DTG and E138K/Q148K-4d complexes were prepared with Protein Preparation Wizard. FEP simulations using the FEP+ program (84) from the Schrödinger suite 2021-2 was used to mutate Gly¹⁴⁰ into Ala. A series of FEP simulations are performed to alchemically “mutate” the Gly¹⁴⁰ side chain into Ala. The structure of the complex E138K/G140A/Q148K-ligand corresponds to the final state ($\lambda = 1$) of the FEP simulation. For every complex and the corresponding endpoint structures, the FEP protein mutation for ligand selectivity graphical user interface (84) in Maestro Suite is applied to build a perturbation map. The OPLS4e force field was used for modeling proteins and ligands (85). Torsion parameters were checked for all ligand fragments using Force Field Builder. A 10-Å cubic box filled with ~31,300 SPC water (86) was used for the complex and solvent perturbation leg. Additional Na⁺ and Cl⁻ ions were added to achieve 0.15 M NaCl in the simulation box. The number of alchemical λ windows is set to 12 by default to connect the WT and the mutant states. For each λ window, the production MD is run for 15 ns in the constant-temperature, constant-pressure (NPT) ensemble. The Bennett acceptance ratio method (87) was used to calculate the free energy differences. The mutational free energy calculations were performed in triplicate by initializing the MD with different random seeds. The reported relative binding free energies were calculated as the average of the three $\Delta\Delta G$ from independent simulations.

Supplementary Materials

This PDF file includes:

Supplementary Note S1 to S3
Figs. S1 to S12
Tables S2 and S3
Legend for table S1
References

Other Supplementary Material for this manuscript includes the following:

Table S1

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. D. O. Passos, M. Li, R. Craigie, D. Lyumkis, "Retroviral integrase: Structure, mechanism, and inhibition" in (Academic Press, 2021; <https://sciencedirect.com/science/article/pii/S187460472100007X>), vol. 50 of *The Enzymes*, pp. 249–300.
2. G. N. Maertens, A. N. Engelman, P. Cherepanov, Structure and function of retroviral integrase. *Nat. Rev. Microbiol.* **20**, 20–34 (2022).
3. A. Engelman, K. Mizuuchi, R. Craigie, HIV-1 DNA integration: Mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67**, 1211–1221 (1991).
4. S. Hare, S. S. Gupta, E. Valkov, A. Engelman, P. Cherepanov, Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236 (2010).
5. G. N. Maertens, S. Hare, P. Cherepanov, The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329 (2010).
6. A. Ballandras-Colas, M. Brown, N. J. Cook, T. G. Dewdney, B. Demeler, P. Cherepanov, D. Lyumkis, A. N. Engelman, Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* **530**, 358–361 (2016).
7. A. Ballandras-Colas, D. P. Maskell, E. Serrao, J. Locke, P. Swuc, S. R. Jonsson, A. Kotecha, N. J. Cook, V. E. Pye, I. A. Taylor, V. Andresdottir, A. N. Engelman, A. Costa, P. Cherepanov, A supramolecular assembly mediates lentiviral DNA integration. *Science* **355**, 93–95 (2017).
8. I. K. Jóźwik, W. Li, D.-W. Zhang, D. Wong, J. Grawenhoff, A. Ballandras-Colas, S. Aiyer, P. Cherepanov, A. N. Engelman, D. Lyumkis, B-to-A transition in target DNA during retroviral integration. *Nucleic Acids Res.* **50**, 8898–8918 (2022).
9. A. Ballandras-Colas, V. Chivukula, D. T. Gruska, Z. Shan, P. K. Singh, V. E. Pye, R. K. McLean, G. J. Bedwell, W. Li, A. Nans, N. J. Cook, H. J. Fadel, E. M. Poeschla, D. J. Griffiths, J. Vargas, I. A. Taylor, D. Lyumkis, H. Yardimci, A. N. Engelman, P. Cherepanov, Multivalent interactions essential for lentiviral integrase function. *Nat. Commun.* **13**, 2416 (2022).
10. Z. Yin, K. Shi, S. Banerjee, K. K. Pandey, S. Bera, D. P. Grandgenett, H. Aihara, Crystal structure of the Rous sarcoma virus intasome. *Nature* **530**, 362–366 (2016).
11. V. Bhatt, K. Shi, D. J. Salamango, N. H. Moeller, K. K. Pandey, S. Bera, T. E. Bohl, F. Kurniawan, K. Orellana, W. Zhang, D. P. Grandgenett, R. S. Harris, A. C. Sundborger-Lunna, H. Aihara, Structural basis of host protein hijacking in human T-cell leukemia virus integration. *Nat. Commun.* **11**, 1–9 (2020).
12. K. K. Pandey, S. Bera, K. Shi, M. J. Rau, A. V. Oleru, J. A. J. Fitzpatrick, A. N. Engelman, H. Aihara, D. P. Grandgenett, Cryo-EM structure of the Rous sarcoma virus octameric cleaved synaptic complex intasome. *Commun. Biol.* **4**, 330 (2021).
13. M. S. Barski, J. J. Minnell, Z. Hodakova, V. E. Pye, A. Nans, P. Cherepanov, G. N. Maertens, Cryo-EM structure of the deltaretroviral intasome in complex with the PP2A regulatory subunit B56 γ . *Nat. Commun.* **11**, 5043 (2020).
14. D. O. Passos, M. Li, R. Yang, S. V. Rebusburg, R. Ghirlando, Y. Jeon, N. Shkriabai, M. Kvaratskhelia, R. Craigie, D. Lyumkis, Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* **355**, 89–92 (2017).
15. D. O. Passos, M. Li, I. K. Jóźwik, X. Z. Zhao, D. Santos-Martins, R. Yang, S. J. Smith, Y. Jeon, S. Forli, S. H. Hughes, T. R. Burke, R. Craigie, D. Lyumkis, Structural basis for strand-transfer inhibitor binding to HIV intasomes. *Science* **367**, 810–814 (2020).
16. D. J. Hazuda, P. Felock, M. Witmer, A. Wolfe, K. Stillmock, J. A. Grobler, A. Espeseth, L. Gabryelski, W. Schleif, C. Blau, M. D. Miller, Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **287**, 646–650 (2000).
17. A. S. Espeseth, P. Felock, A. Wolfe, M. Witmer, J. Grobler, N. Anthony, M. Egbertson, J. Y. Melamed, S. Young, T. Hamill, J. L. Cole, D. J. Hazuda, HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc. National Acad. Sci. U.S.A.* **97**, 11244–11249 (2000).
18. N. S. Pandit, D. B. Chastain, A. M. Pallotta, M. E. Badowski, E. C. Huesgen, S. M. Michienzi, Simplifying ARV therapy in the setting of resistance. *Curr. Infect. Dis. Rep.* **21**, 38 (2019).
19. J.-L. Blanco, V. Varghese, S.-Y. Rhee, J. M. Gatell, R. W. Shafer, HIV-1 integrase inhibitor resistance and its clinical implications. *J. Infect. Dis.* **203**, 1204–1214 (2011).
20. S. J. Smith, X. Z. Zhao, D. O. Passos, D. Lyumkis, T. R. Burke, S. H. Hughes, HIV-1 integrase inhibitors that are active against drug-resistant integrase mutants. *ACS Infect. Dis.* **64**, 430 (2020).
21. S. J. Smith, X. Z. Zhao, T. R. Burke Jr., S. H. Hughes, HIV-1 integrase inhibitors that are broadly effective against drug-resistant mutants. *Antimicrob. Agents Chemother.* **62**, 10.1128/aac.01035-18, (2018).

22. S. J. Smith, X. Z. Zhao, T. R. Burke, S. H. Hughes, Efficacies of Cabotegravir and Bictegravir against drug-resistant HIV-1 integrase mutants. *Retrovirology* **15**, 37 (2018).
23. S.-Y. Rhee, P. M. Grant, P. L. Tzou, G. Barrow, P. R. Harrigan, J. P. A. Ioannidis, R. W. Shafer, A systematic review of the genetic mechanisms of dolutegravir resistance. *J. Antimicrob. Chemother.* **32**, 1551 (2019).
24. J. L. Blanco, A.-G. Marcelin, C. Katlama, E. Martinez, Dolutegravir resistance mutations. *Curr. Opin. Infect. Dis.* **31**, 237–245 (2018).
25. K. Anstett, B. Brenner, T. Mesplede, M. A. Wainberg, HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology* **14**, 36 (2017).
26. J. M. Coffin, HIV population dynamics in vivo: Implications for genetic variation, pathogenesis, and therapy. *Science* **267**, 483–489 (1995).
27. S. J. Smith, X. Z. Zhao, D. O. Passos, D. Lyumkis, T. R. Burke, S. H. Hughes, Integrase strand transfer inhibitors are effective anti-HIV drugs. *Viruses* **13**, 205 (2021).
28. N. J. Cook, W. Li, D. Berta, M. Badaoui, A. Ballandras-Colas, A. Nans, A. Kotecha, E. Rosta, A. N. Engelman, P. Cherepanov, Structural basis of second-generation HIV integrase inhibitor action and viral resistance. *Science* **367**, 806–810 (2020).
29. X. Z. Zhao, S. J. Smith, D. P. Maskell, M. Métifiot, V. E. Pye, K. Fesen, C. Marchand, Y. Pommier, P. Cherepanov, S. H. Hughes, T. R. Burke, HIV-1 integrase strand transfer inhibitors with reduced susceptibility to drug resistant mutant integrases. *ACS Chem. Biol.* **11**, 1074–1081 (2016).
30. X. Z. Zhao, S. J. Smith, D. P. Maskell, M. Métifiot, V. E. Pye, K. Fesen, C. Marchand, Y. Pommier, P. Cherepanov, S. H. Hughes, T. R. Burke, Structure-guided optimization of HIV integrase strand transfer inhibitors. *J. Med. Chem.* **60**, 7315–7332 (2017).
31. S. J. Smith, X. Z. Zhao, D. O. Passos, V. E. Pye, P. Cherepanov, D. Lyumkis, T. R. Burke, S. H. Hughes, HIV-1 integrase inhibitors with modifications that affect their potencies against drug resistant integrase mutants. *ACS Infect. Dis.* **7**, 1469–1482 (2021).
32. N. A. Margot, R. R. Ram, K. L. White, M. E. Abram, C. Callebaut, Antiviral activity of HIV-1 integrase strand-transfer inhibitors against mutants with integrase resistance-associated mutations and their frequency in treatment-naïve individuals. *J. Med. Virol.* **91**, 2188–2194 (2019).
33. K. E. Hightower, R. Wang, F. DeAnda, B. A. Johns, K. Weaver, Y. Shen, G. H. Tomberlin, H. L. Carter, T. Broderick, S. Sigethy, T. Seki, M. Kobayashi, M. R. Underwood, Dolutegravir (S/GSK1349572) exhibits significantly slower dissociation than Raltegravir and Elvitegravir from wild-type and integrase inhibitor-resistant HIV-1 integrase-DNA complexes. *Antimicrob. Agents Chemother.* **55**, 4552–4559 (2011).
34. O. Delelis, I. Malet, L. Na, L. Tchertanov, V. Calvez, A.-G. Marcelin, F. Subra, E. Deprez, J.-F. Mouscadet, The G140S mutation in HIV integrases from raltegravir-resistant patients rescues catalytic defect due to the resistance Q148H mutation. *Nucleic Acids Res.* **37**, 1193–1201 (2009).
35. M. Li, K. A. Jurado, S. Lin, A. Engelman, R. Craigie, Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLOS ONE* **9**, e105078 (2014).
36. S.-Y. Rhee, M. J. Gonzales, R. Kantor, B. J. Betts, J. Ravela, R. W. Shafer, Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **31**, 298–303 (2003).
37. R. W. Shafer, Rationale and uses of a public HIV drug-resistance database. *J. Infect. Dis.* **194**, S51–S58 (2006).
38. A. L. Ferguson, J. K. Mann, S. Omarjee, T. Ndung'u, B. D. Walker, A. K. Chakraborty, Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. *Immunity* **38**, 606–617 (2013).
39. J. P. Barton, N. Goonetilleke, T. C. Butler, B. D. Walker, A. J. McMichael, A. K. Chakraborty, Relative rate and location of intra-host HIV evolution to evade cellular immunity are predictable. *Nat. Commun.* **7**, 11660 (2016).
40. M. Figliuzzi, H. Jacquier, A. Schug, O. Tenaillon, M. Weigt, Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1. *Mol. Biol. Evol.* **33**, 268–280 (2016).
41. T. A. Hopf, J. B. Ingraham, F. J. Poelwijk, C. P. I. Schärfe, M. Springer, C. Sander, D. S. Marks, Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35**, 128–135 (2017).
42. K. Shekhar, C. F. Ruberman, A. L. Ferguson, J. P. Barton, M. Kardar, A. K. Chakraborty, Spin models inferred from patient-derived viral sequence data faithfully describe HIV fitness landscapes. *Phys. Rev. E* **88**, 062705 (2013).
43. T. C. Butler, J. P. Barton, M. Kardar, A. K. Chakraborty, Identification of drug resistance mutations in HIV from constraints on natural evolution. *Phys. Rev. E* **93**, 022412 (2016).
44. A. Biswas, A. Haldane, E. Arnold, R. M. Levy, Epistasis and entrenchment of drug resistance in HIV-1 subtype B. *eLife* **8**, e05024 (2019).
45. A. Horowitz, Double-mutant cycles: A powerful tool for analyzing protein structure and function. *Fold. Des.* **1**, R121–R126 (1996).
46. T. Yoshinaga, M. Kobayashi, T. Seki, S. Miki, C. Wakasa-Morimoto, A. Suyama-Kagitani, S. Kawauchi-Miki, T. Taishi, T. Kawasaki, B. A. Johns, M. R. Underwood, E. P. Garvey, A. Sato, T. Fujiwara, Antiviral characteristics of GSK1265744, an HIV integrase inhibitor dosed orally or by long-acting injection. *Antimicrob. Agents Chemother.* **59**, 397–406 (2015).
47. J. A. Fulcher, Y. Du, T. Zhang, R. Sun, R. J. Landovitz, Emergence of integrase resistance mutations during initial therapy containing dolutegravir. *Clin. Infect. Dis.* **67**, 791–794 (2018).
48. A. M. Wensing, V. Calvez, F. Ceccherini-Silberstein, C. Charpentier, H. F. Günthard, R. Paredes, R. W. Shafer, D. D. Richman, 2022 update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* **30**, 559–574 (2022).
49. E. C. Hulme, M. A. Trevethick, Ligand binding assays at equilibrium: Validation and interpretation. *Bri. J. Pharmacol.* **161**, 1219–1237 (2010).
50. S. Hare, G. N. Maertens, P. Cherepanov, 3'-processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO J.* **31**, 3020–3028 (2012).
51. X. Z. Zhao, S. J. Smith, M. Métifiot, C. Marchand, P. L. Boyer, Y. Pommier, S. H. Hughes, T. R. Burke, 4-Amino-1-hydroxy-2-oxo-1,8-naphthyridine-containing compounds having high potency against raltegravir-resistant integrase mutants of HIV-1. *J. Med. Chem.* **57**, 5190–5202 (2014).
52. S. J. Smith, S. H. Hughes, Rapid screening of HIV reverse transcriptase and integrase inhibitors. *J. Vis. Exp.*, e51400 (2014).
53. A. Haldane, R. M. Levy, Mi3-GPU: MCMC-based inverse Ising inference on GPUs for protein covariation analysis. *Comput. Phys. Commun.* **260**, 107312 (2021).
54. R. M. Levy, A. Haldane, W. F. Flynn, Potts Hamiltonian models of protein co-variation, free energy landscapes, and evolutionary fitness. *Curr. Opin. Struct. Biol.* **43**, 55–62 (2017).
55. A. Haldane, W. F. Flynn, P. He, R. S. K. Vijayan, R. M. Levy, Structural propensities of kinase family proteins from a Potts model of residue co-variation. *Protein Sci.* **25**, 1378–1384 (2016).
56. F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, M. Weigt, Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. National. Acad. Sci. U.S.A.* **108**, E1293–E1301 (2011).
57. D. S. Marks, T. A. Hopf, C. Sander, Protein structure prediction from sequence variation. *Nat. Biotechnol.* **30**, 1072–1080 (2012).
58. A. Biswas, A. Haldane, R. M. Levy, Limits to detecting epistasis in the fitness landscape of HIV. *PLOS ONE* **17**, e0262314 (2022).
59. T. Zhang, L. Dai, J. P. Barton, Y. Du, Y. Tan, W. Pang, A. K. Chakraborty, J. O. Lloyd-Smith, R. Sun, Predominance of positive epistasis among drug resistance-associated mutations in HIV-1 protease. *PLOS Genet.* **16**, e1009009 (2020).
60. R. H. Y. Louie, K. J. Kaczorowski, J. P. Barton, A. K. Chakraborty, M. R. McKay, Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E564–E573 (2018).
61. W. F. Flynn, A. Haldane, B. E. Torbett, R. M. Levy, Inference of epistatic effects leading to entrenchment and drug resistance in HIV-1 protease. *Mol. Biol. Evol.* **34**, msx095 (2017).
62. C. Suloway, J. Pulokas, D. Fellmann, A. Cheng, F. Guerra, J. Quispe, S. Stagg, C. S. Potter, B. Carragher, Automated molecular microscopy: The new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
63. A. Cheng, C. Negro, J. F. Bruhn, W. J. Rice, S. Dallakyan, E. T. Eng, D. G. Waterman, C. S. Potter, B. Carragher, Leginon: New features and applications. *Protein Sci.* **30**, 136–150 (2020).
64. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
65. D. Kimanius, L. Dong, G. Sharov, T. Nakane, S. H. W. Scheres, New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem. J.* **478**, 4169–4185 (2021).
66. T. Grant, N. Grigorieff, Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
67. T. Grant, A. Rohou, N. Grigorieff, cisTEM, User-friendly software for single-particle image processing. *eLife* **7**, e14874 (2018).
68. D. Tegunov, P. Cramer, Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **71**, 1–7 (2019).
69. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
70. G. Harauz, M. van Heel, Exact filters for general geometry 3-dimensional reconstruction. *Optik* **73**, 146–156 (1986).
71. A. Punjani, H. Zhang, D. J. Fleet, Non-uniform refinement: Adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
72. Y. Z. Tan, P. R. Baldwin, J. H. Davis, J. R. Williamson, C. S. Potter, B. Carragher, D. Lyumkis, Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
73. P. R. Baldwin, D. Lyumkis, Non-uniformity of projection distributions attenuates resolution in Cryo-EM. *Prog. Biophys. Mol. Biol.* **150**, 160–183 (2020).
74. P. R. Baldwin, D. Lyumkis, Tools for visualizing and analyzing Fourier space sampling in Cryo-EM. *Prog. Biophys. Mol. Biol.* **160**, 53–65 (2021).

75. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
76. N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, Electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. Sect. D* **65**, 1074–1080 (2009).
77. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
78. A. Brown, F. Long, R. A. Nicholls, J. Toots, P. Emsley, G. Murshudov, Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 136–153 (2015).
79. P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, P. D. Adams, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. D* **74**, 531–544 (2018).
80. O. Carugo, D. Bordo, How many water molecules can be detected by protein crystallography? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 479–483 (1999).
81. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D* **75**, 861–877 (2019).
82. P. V. Afonine, B. P. Klaholz, N. W. Moriarty, B. K. Poon, O. V. Sobolev, T. C. Terwilliger, P. D. Adams, A. Urzhumtsev, New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 814–840 (2018).
83. C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, D. C. Richardson, MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
84. L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beumung, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, R. Abel, Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **137**, 2695–2703 (2015).
85. K. Roos, C. Wu, W. Damm, M. Reboul, J. M. Stevenson, C. Lu, M. K. Dahlgren, S. Mondal, W. Chen, L. Wang, R. Abel, R. A. Friesner, E. D. Harder, OPLS3e: Extending force field coverage for drug-like small molecules. *J. Chem. Theory Comput.* **15**, 1863–1874 (2019).
86. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration intermolecular force. *Jerus Symposia Quantum Chem. Biochem.* (1981), vol. 14.
87. C. H. Bennett, Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **22**, 245–268 (1976).
88. S. Aijer, C. Zhang, P. R. Baldwin, D. Lyumkis, Evaluating local and directional resolution of cryo-EM density maps. *Methods Mol. Biol.* **2215**, 161–187 (2021).

Acknowledgments: We would like to thank P. Afonine for help with atomic modeling and solvation. The content of this publication does not necessarily reflect the views or policies of the

Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. **Funding:** We are grateful for funding from the following sources: National Institutes of Health grant U01 AI136680 (D.L., R.M.L., T.R.B., and R.C.); National Institutes of Health grant R01 AI146017 (D.L.); National Institutes of Health grant U54 AI170855 (D.L., R.M.L., T.R.B., and R.C.); The Intramural Program of the National Institute of Diabetes and Digestive Diseases (R.C.); National Institutes of Health grant R35 GM132090 (R.M.L.); Margaret T. Morris Foundation (D.L.); Hearst Foundations (D.L.); NIH Intramural Program, Center for Cancer Research, National Cancer Institute NIH (S.J.S., X.Z.Z., T.R.B., and S.H.H.); AIDS Intramural Targeted Program, Center for Cancer Research, National Cancer Institute NIH (S.J.S., X.Z.Z., T.R.B., and S.H.H.); and National Institutes of Health F32 postdoctoral fellowship GM148049 (T.S.). The molecular graphics and analyses were performed with the USCF Chimera package (supported by NIH P41 GM103311). The cryo-EM data for this project were collected at Scripps Research (supported by S10 OD032467), University of Utah (assisted by David Timm), Stanford-SLAC (supported by the National Institutes of Health Common Fund Transformative High Resolution Cryo-Electron Microscopy program), UCLA (supported by U24GM116792), and the NIH. R.C. and M.L. acknowledge the support of the NIH Multi Institute Cryo-EM Facility (MICEF), the NIDDK Cryo-Electron Microscopy Core Facility, and the NIH Intramural Cryo-EM Consortium (NICE). A portion of this research was supported by NIH grant U24GM129547 and performed at the PNC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. **Author contributions:** Conceptualization: D.O.P., D.L., M.L., and R.C. Methodology: M.L., Z.L., D.O.P., Z.S., T.S.S., D.L., S.J.S., A.B., N.D., A.H., Q.S., I.C., X.Z., and L.B. Supervision: M.K., T.R.B., R.M.L., S.H.H., R.C., and D.L. Writing—original draft: D.L. Writing—review and editing: M.L., D.O.P., Z.S., A.B., I.C., A.H., R.M.L., S.H.H., R.C., and D.L. Methodology: M.L. developed the conditions for intasome production and purification of the mutant samples. Z.L. contributed to protein purification. D.O.P., M.L., and Z.S. performed cryo-EM data acquisition, image processing, and structure determination. T.S. and D.L. built and refined the models. S.J.S. performed virology assays and helped with the interpretation of the data. M.L. performed enzymatic assays. A.B. and A.H. performed the Potts analyses. Q.S. performed MD simulations. X.Z.Z. synthesized compound 4d for structural work. L.B. performed SPR experiments. **Competing interests:** X.Z.Z., S.S., S.H.H., and T.R.B. have been awarded a patent for the development of compound 4d, U.S. patent 10,208,035, issued 19 February 2019. The authors declare that they have no other competing interests. **Data and materials availability:** The cryo-EM maps and atomic models have been deposited into the Electron Microscopy Data Bank and Protein Data Bank under the accession codes: WT-DTG (EMD-29307, PDB 8FN7); E138K-DTG (EMD-29309, PDB 8FND); G140A-DTG (EMD-29312, PDB 8FNG); Q148K-DTG (EMD-29313, PDB 8FNH); E138K/G140A-DTG (EMD-29315, PDB 8FNJ); E138K/Q148K-DTG (EMD-29317, PDB 8FNL); G140A/Q148K-DTG (EMD-29318, PDB 8FNM); E138K/G140A/Q148K-DTG (EMD-29319, PDB 8FNN); E138K/G140A/Q148R-DTG (EMD-29320, PDB 8FNO); E138K/G140S/Q148H-DTG (EMD-29321, PDB 8FNP); E138K/G140A/Q148K-4d (EMD-29322, PDB 8FNQ). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Compound 4d can be provided by T.R.B. pending scientific review and a completed material transfer agreement. Requests for compound 4d should be submitted to burkete@mail.nih.gov.

Submitted 9 January 2023

Accepted 16 June 2023

Published 21 July 2023

10.1126/sciadv.adg5953