PLOS PATHOGENS

Structure of a HIV-1 IN-Allosteric inhibitor complex at 2.93 Å resolution: Routes to inhibitor optimization

Grant Eilers[1,2‡], Kushol Gupta[2‡], Audrey Allen[1,2], Saira Montermoso[2], Hemma Murali[2], Robert Sharp[2], Young Hwang[1], Frederic D. Bushman[1]*, Gregory Van Duyne[2]*

1 Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **2** Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

‡ These authors share first authorship on this work.
* bushman@pennmedicine.upenn.edu (FDB); vanduyne@pennmedicine.upenn.edu (GVD)

Abstract

HIV integrase (IN) inserts viral DNA into the host genome and is the target of the strand transfer inhibitors (STIs), a class of small molecules currently in clinical use. Another potent class of antivirals is the allosteric inhibitors of integrase, or ALLINIs. ALLINIs promote IN aggregation by stabilizing an interaction between the catalytic core domain (CCD) and carboxy-terminal domain (CTD) that undermines viral particle formation in late replication. Ongoing challenges with inhibitor potency, toxicity, and viral resistance motivate research to understand their mechanism. Here, we report a 2.93 Å X-ray crystal structure of the minimal ternary complex between CCD, CTD, and the ALLINI BI-224436. This structure reveals an asymmetric ternary complex with a prominent network of π-mediated interactions that suggest specific avenues for future ALLINI development and optimization.

Author summary

The global burden of the HIV/AIDS pandemic and continued emergence of drug resistance drives the need for novel antivirals. The allosteric integrase inhibitors, or ALLINIs, are a potent class of antivirals in development that target the enzyme integrase in a surprising fashion: the small molecules act to *stabilize* an inappropriate protein-protein interaction to attain biological effect. Here, we report the first atomic resolution (2.93 Å) X-ray crystal structure of the minimal ternary complex between domains of the integrase and the first ALLINI preclinical lead BI-224436. Our structure provides a more precise view of the molecular interactions that underlie drug potency, and several aspects of our data suggest routes to improving ALLINI design and minimizing acquisition of resistance.

Introduction

Control of infection and prevention of transmission are crucial to ending the HIV epidemic [1]. Use of potent antiretrovirals with the highest genetic barriers to drug resistance is essential for minimizing the emergence and spread of resistant virus [2].

HIV integration is an essential step in the retroviral replication cycle [3], and the target of highly effective antiretrovirals [2]. Retroviral IN proteins consist of three well-defined structural domains: an N-terminal domain (NTD, residues 1–50), a catalytic core domain (CCD, residues 50–212), and C-terminal domain (CTD, residues 213–288) (Fig 1A) [4–6]. The NTD binds Zn^{2+} via a conserved HHCC motif. The CCD adopts an RNase H superfamily fold and contains a D,D-35-E motif that binds Mg^{2+} or Mn^{2+} ions, which mediate DNA cleaving and joining events. The CTD features an SH3-like fold that contributes to DNA binding and interacts with other IN domains. HIV-1 IN forms dimers and tetramers in solution [7–12] and all three domains of IN form dimers in isolation [13–21].
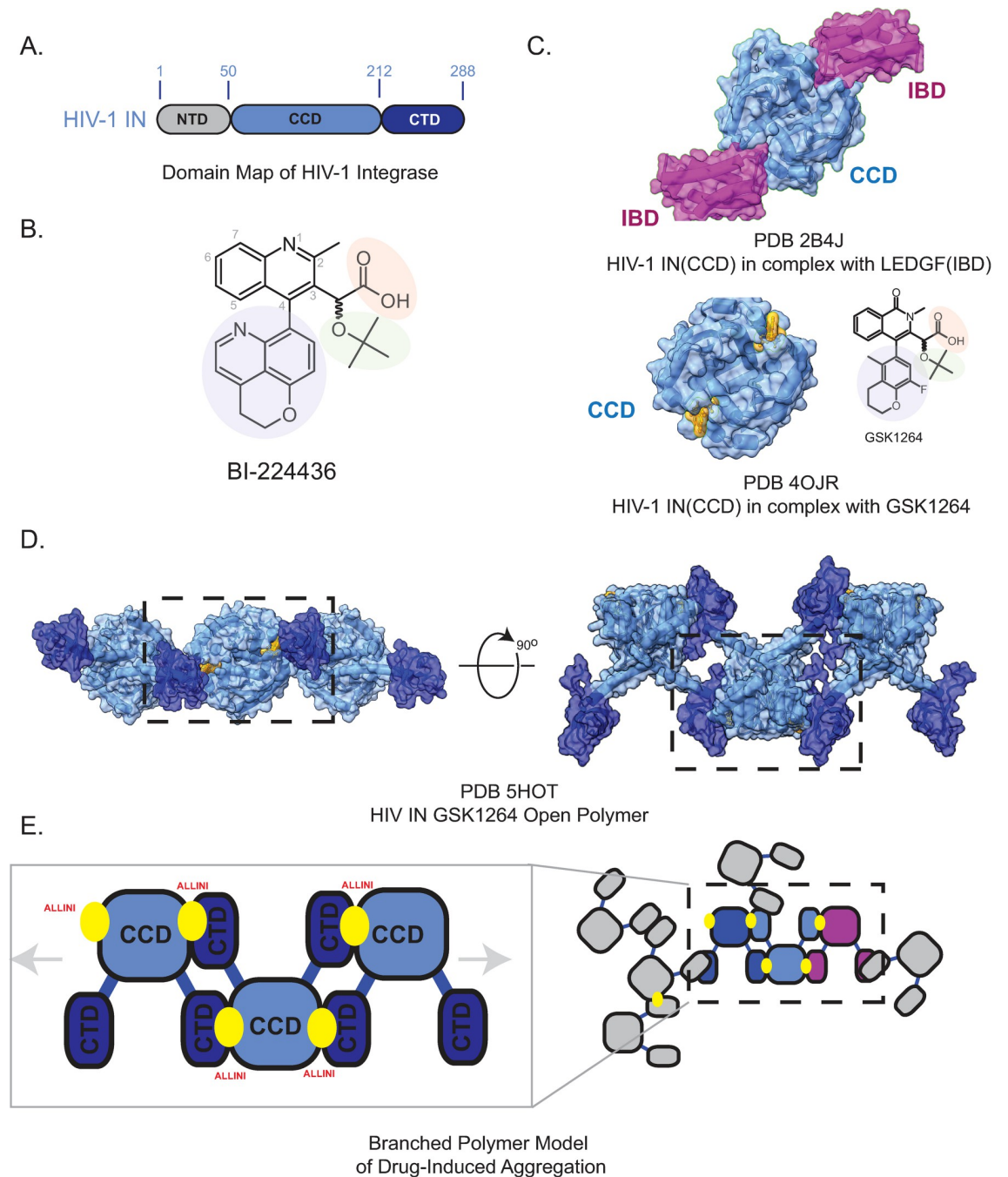
During integration, IN catalyzes two sequential chemical reactions: removal of a GT dinucleotide from the viral 3'-DNA ends and covalent insertion of the processed viral cDNA ends into host DNA [3,22]. Detailed studies of the enzymatic activity of IN led to the development of the strand transfer inhibitors (STIs), five of which have received FDA approval [23–25]. Despite long-lasting effects and favorable pharmacological profiles, resistance to STIs is well-documented [26–31] and, importantly, different STIs show overlapping resistance profiles, motivating the search for alternative approaches to targeting IN [32–35].

Another class of integrase inhibitors, the allosteric inhibitors (ALLINIs) (Fig 1B) target the pocket normally occupied by the integrase-binding domain (IBD) of the lens epithelium-derived growth factor (LEDGF/p75) [36–39] (Fig 1C). LEDGF binds to IN and directs the IN-cDNA complex toward transcriptionally active areas of the genome for integration [40–42]. Knockdown of LEDGF results in a loss of this integration site-specificity [40,43] and inhibits HIV replication [44]. The LEDGF binding pocket, formed by the dimer interface of the CCD, was thus recognized to be a potential target for drug design [11,36,45–50], leading to the subsequent development of high affinity antivirals.

Although ALLINIs were originally intended to interfere with viral integration, the major consequence of ALLINI treatment is observed late in the replication cycle, after viral DNA integration has already occurred [11,45,47,48,51]. ALLINI-treated virus producer cells yield non-infectious virions containing abnormal electron-dense aggregates and mis-localized ribonucleoprotein complexes outside of the viral core [10,11,45]. ALLINIs disrupt virion assembly by promoting abnormal polymerization of IN [10,11,52] using a mechanism analogous to that of the drug paclitaxel, which promotes the inappropriate polymerization of tubulin [53]. Studies using ALLINIs have revealed a role of IN in sequence-specific binding of the viral RNA genome, indicating that IN plays a required role in virion maturation [54,55]. A more modest effect of ALLINIs observed early in the viral replication cycle is likely due to ALLINIs directly interfering with IN binding to LEDGF.

Structures of ALLINIs bound to the isolated CCD revealed at high resolution how small molecules could inhibit LEDGF binding via high affinity interactions [11,36–39,47]. However, the structures did not indicate how drug-induced aggregation is initiated or how other domains of IN might be involved [11,56] (Fig 1C). The complete molecular structure of the ALLINI binding interface, comprised of both the CCD and CTD, was first revealed in a 4.4 Å crystal structure of IN bound with the drug GSK1264 [10] (Fig 1D). The structure revealed how CTDs of neighboring IN dimers bind to the ALLINI-CCD interface and initiate the formation of IN-ALLINI open polymers [52]. Further studies led to a model where ALLINIs induce a three-dimensional branched polymer of IN via additional homomeric CTD interactions. [52] (Fig 1E).

Here, we present a 2.93 Å resolution structure of a minimal ternary complex formed by the IN CCD, CTD, and the preclinical lead ALLINI BI-224436. The structure reveals side chain orientations and a more precise view of the molecular interactions that underlie ALLINI-induced aggregation of HIV IN. The complex has a pronounced asymmetry, with non-

A.

Domain Map of HIV-1 Integrase

B.

BI-224436

C.

PDB 2B4J
HIV-1 IN(CCD) in complex with LEDGF(IBD)

GSK1264

PDB 4OJR
HIV-1 IN(CCD) in complex with GSK1264

D.

PDB 5HOT
HIV IN GSK1264 Open Polymer

E.

Branched Polymer Model
of Drug-Induced Aggregation

Fig 1. HIV-1 Integrase and ALLINI Structure. A. Domain Map of HIV-1 Integrase. B. The preclinical lead ALLINI BI-224436. Chemical features common to the ALLINI class are highlighted, including a carboxylic acid moiety (salmon), a tert-butoxy moiety (green), and a large hydrophobic appendage at the R4 position (blue). C. Structure of HIV-1 IN CCD in complex with LEDGF/p75 Integrase Binding Domain (IBD, purple, PDB 2B4J). Shown below is the structure of HIV-1 IN CCD (blue) in complex with the ALLINI GSK1264 (PDB 5HOT). D. The IN-ALLINI structure reveals an open linear polymer configuration within the crystal lattice that underlies drug-induced aggregation. Outlined in black is the minimal ternary complex describing the ALLINI-induced protein-protein interaction between CCD and CTD. E. Cartoon Schematic of the CCD•CTD domain interactions stabilized by ALLINIs and a branched polymer model for drug-induced aggregation of HIV-1 IN.

identical ALLINI binding interfaces that depend on the nature of the proximal CTD dimers formed in the crystal lattice. We identify several new interactions including a network of cation-π and π-π interactions at the protein-protein and protein-drug interfaces. An accessible pocket adjacent to the bound ALLINI is occupied by ethylene glycol, suggesting specific directions for drug design. The minimal CCD•ALLINI•CTD assembly, like the full-length protein complex, favors the formation of drug-induced polymers in solution via two modes of CTD dimerization, providing orthogonal evidence supporting a branched polymer mechanism of aggregation. From this improved atomic model, we can generalize the mode of action for first-generation molecules and current clinical leads and suggest routes for improvement of existing ALLINI scaffolds.

Results

Reconstitution of a minimal CCD•ALLINI•CTD ternary complex

In the structure of full-length IN bound with GSK1264 (PDB 5HOT) [10] at low resolution, side chain positions only could be inferred, and atomic positions were strongly constrained by the reference models implemented in the crystallographic refinement employed [57]. This motivated the pursuit of a simpler assembly that could capture all the primary features of the drug interaction using isolated IN domains and an ALLINI (boxed region in Fig 1D). We therefore purified IN CCD (residues 50–212) and CTD (residues 220–271) and reconstituted the ternary complex with the ALLINI BI-224436 for crystallization trials. BI-224436 is a potent antiviral (EC_{50} <15 nM [58]) with favorable solubility properties and was the first ALLINI tested in preclinical trials [37,58]. Its chemical structure recapitulates the most common features among ALLINIs, including carboxylic and *tert*-butoxy groups alongside large hydrophobic moieties that optimize its fit to the hydrophobic dimer interface of the CCD (Fig 1B).

A distinguishing property of the ALLINI class is its ability to induce polymerization of HIV integrase *in vitro*, as evidenced by the formation of a cloudy precipitate composed of IN branched polymers that can be monitored by turbidity assays [10,11,52]. In a similar fashion, addition of BI-224436 to an equimolar solution of isolated CCD and the larger CTD(220–288) domains also elicits the formation of a cloudy solution (S1A Fig); turbidity is observed with the shorter CTD(220–271) truncations in complexes prepared for crystallization trials. As shown in S1B Fig, DLS analysis indicates that polymerization is both ALLINI-dependent and requires the CTD.

To further quantitate the formation of the ternary complex, sedimentation velocity analytical ultracentrifugation (SV-AUC) analysis was performed (S1C–S1E Fig). Based on our prior work [52], introduction of the L242A mutation disrupts CTD homodimerization and should therefore inhibit precipitation if such interactions underlie the observed aggregation. Fitting of data to the Lamm equation allows for the determination of a distribution of species, both in terms of sedimentation coefficient (*c(S)*), and mass (*c(M)*). SV-AUC data collected at 20°C with CCD•BI-224436•CTD(220–270)^{L242A} allowed observation of discrete peak species in *c(S)* and *c(M)* distributions, confirming the formation of a ternary complex and higher-order species (S1D–S1E Fig). Despite the predicted increase in mass upon complex formation, hydrodynamic calculations of model structures predict a decrease in the apparent S value of the ternary complex (~2.7) vs CCD (~3.1) alone due to the increase in shape anisotropy, as sedimentation is a technique sensitive to both size and shape. As predicted, we observe changes in S values consistent with the formation of a ternary complex.

Overall structure of the CCD^{F185K}•BI-224436•CTD ternary complex

The 2:2:2 ternary complex of CCD^{F185K}•BI-224436•CTD(220–271) was crystallized, and diffraction data were measured to 2.9 Å resolution (Table 1). The structure was determined by

Table 1. Data collection and refinement statistics[a].

	IN CCD[F185K]•CTD•BI-224436	IN CCD[F185K]•BI-224436	IN CCD[F185K]
PDB	8CTA	8CT7	8CT5
X-ray Source	Beamline 17-ID1 (AMX), NSLS-II	Beamline FlexX, CHESS	Beamline FlexX, CHESS
Wavelength (Å)	0.92	0.92	0.92
Temperature (K)	100	100	100
Detector	Eiger 9M	Eiger2 16M	Eiger2 16M
Resolution range (Å)	19.84–2.93 (3.04–2.93)	29.37–2.13 (2.206–2.13)	29.29–1.97 (2.044–1.97)
Space group	C 1 2 1	P 3$_1$21	P 3$_1$21
Unit cell (a,b,c,α,β,γ)	181.51,116.51,112.8 90,103.6,90	72.49,72.49,66.45 90,90,120	72.97,72.97,66.1 90,90,120
Total reflections	197,739 (20,582)	23,244 (2,261)	29,314 (2,836)
Unique reflections	48,717 (2,000)	11,625 (1,133)	14,661 (1,422)
Multiplicity	4.1 (4.3)	2.0 (2.0)	2.0 (2.0)
Completeness (%)	90.4 (41.1)	99.9 (100.00)	99.8 (98.8)
Mean I/sigma(I)	10.8 (1.8)	29.5 (3.7)	27.3 (3.6)
Wilson B-factor (Å^2)	53.5	44.8	37.8
R-merge[b]	0.078 (1.10)	0.017 (0.160)	0.018 (0.151)
R-meas[c]	0.090 (1.25)	0.024 (0.223)	0.026 (0.213)
R-pim[d]	0.043 (0.59)	0.017 (0.160)	0.018 (0.151)
CC1/2	0.997 (0.915)	1 (0.927)	0.999 (0.942)
CC*	0.999 (0.977)	1 (0.981)	1 (0.985)
Reflections used in refinement	44,409 (2,000)	11,622 (1,133)	14,661 (1,421)
Reflections used for R-free	4,439 (200)	1,176 (111)	1,460 (143)
R-work	0.209 (0.348)	0.199 (0.255)	0.200 (0.223)
R-free	0.263 (0.388)	0.228 (0.258)	0.228 (0.240)
CC(work)	0.898 (0.690)	0.953 (0.864)	0.959 (0.888)
CC(free)	0.893 (0.728)	0.948 (0.890)	0.935 (0.854)
Number of non-hydrogen atoms	12,468	1,204	1,168
macromolecules	12,017	1,120	1,084
ligands	372	62	40
solvent	79	22	44
Protein residues	1,512	146	141
RMS(bonds) (Å)	0.024	0.012	0.013
RMS(angles) (˚)	1.97	1.43	0.98
Ramachandran favored (%)	93.0	98.5	99.2
Ramachandran allowed (%)	5.1	0.75	0.8
Ramachandran outliers (%)	1.9	0.75	0.0
Rotamer outliers (%)	7.9	1.7	0.0
Clashscore	19.6	3.9	4.1
Average B-factor (Å^2)	74.7	51.1	56.6
macromolecules	75.2	51.2	54.9
ligands	65.0	49.6	108.6
solvent	38.9	49.1	50.5

(*Continued*)

Table 1. (Continued)

	IN CCD^{F185K}•CTD•BI-224436	IN CCD^{F185K}•BI-224436	IN CCD^{F185K}
Number of TLS groups	16	3	3

Statistics for the highest-resolution shell are shown in parentheses.

[a] Values in parentheses refer to the highest resolution shell.

[b] R_{merge} is calculated by the equation $R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl)—<I(hkl)>| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement.

[c] R_{meas} (or redundancy-independent R_{merge}) is calculated by the equation $R_{meas} = \Sigma_{hkl} [N/(N—1)]^{1/2} \Sigma_i |I_i(hkl)—<I(hkl)>| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement and N is the redundancy of each unique reflection hkl [59].

[d] R_{pim} is calculated by the equation $R_{pim} = \Sigma_{hkl} [1/(N—1)]^{1/2} \Sigma_i |I_i(hkl)—<I(hkl)>| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement and N is the redundancy of each unique reflection hkl [60].

molecular replacement, using high resolution CCD and CTD coordinates as search models. Four complexes are present in the crystallographic asymmetric unit, providing eight independent observations of the ALLINI-protein interface (S2A Fig). Continuous electron density is observed for backbone atoms and for most side chain atoms. The four ternary complexes superimpose with a Cα RMSD of 0.14–0.17 Å, compared to an estimated coordinate error of 0.35 Å for the refined structure.

The structure of the minimal CCD•BI-224436•CTD ternary complex is shown in Fig 2. The overall quaternary arrangement surrounding the ALLINI resembles that observed in the structure of full-length IN•GSK1264 [10], where the CTDs are connected by α-helical linkers to neighboring CCD dimers. The ALLINI binding interface is composed of residues from the α1 and α3 helices from one CCD subunit and α4 and α5 helices from the partner CCD subunit, along with the β1, β1-β2 turn, and β5 strands of the CTD. The two ALLINI binding sites are interconnected by the α4 helices and intermediate linker region (residues 138–155) of each subunit. This region spans and organizes the D-D-E active sites found within each subunit of the CCD dimer. In reported experimental structures that include the IN CCD dimer, the organization of this region is highly variable, and the linker is often disordered.

Although the CCD dimer is nearly symmetric in the ternary complex structure (Cα RMSD = 0.15 Å not including the α4 helices), the CTDs and the α4 helices adopt slightly different positions in the two halves of the complex. The same asymmetry of the CTDs and α4 is observed in each of the four independent complex copies in the asymmetric unit. Each complex contains one well-ordered ALLINI binding site (site 1) that has strong electron density and low B-factors, and one less well-ordered ALLINI binding site (site 2) that has weaker electron density for the key interacting residues and higher overall B-factors for the CTD (Fig 3A and 3B). The ALLINI binding interface buries ~512 Å^2 of solvent-accessible protein surface area between the CTD and CCD dimer, and an additional 534 Å^2 (of 598 Å^2 total) of BI-224436 at each site. However, the two binding pockets vary in surface and volume, as illustrated by CASTp analysis [61]. Site 1 has a defined volume of ~781 Å^3, whereas site 2 has a defined volume of ~971 Å^3 (Fig 3C).

An examination of crystal contacts between ternary complexes reveals that homomeric CTD interactions generate linear polymers (Fig 4A). The interactions alternate between a canonical CTD dimerization motif previously observed by crystallography (PDB 5TC2) and NMR [17,62] and an alternative dimerization interface, where the CTD subunits have been rotated ~180˚ (Fig 4B). These two modes of dimerization bury similar solvent-accessible surface (322 Å^2 vs 248 Å^2, respectively) but use key interfacial residues in distinct ways. The canonical dimer bridges site 2 from two adjacent CCDs, whereas the alternative dimer bridges site 1 from adjacent CCDs (Fig 4A). Thus, asymmetry observed in the CCD•BI-224436•CTD
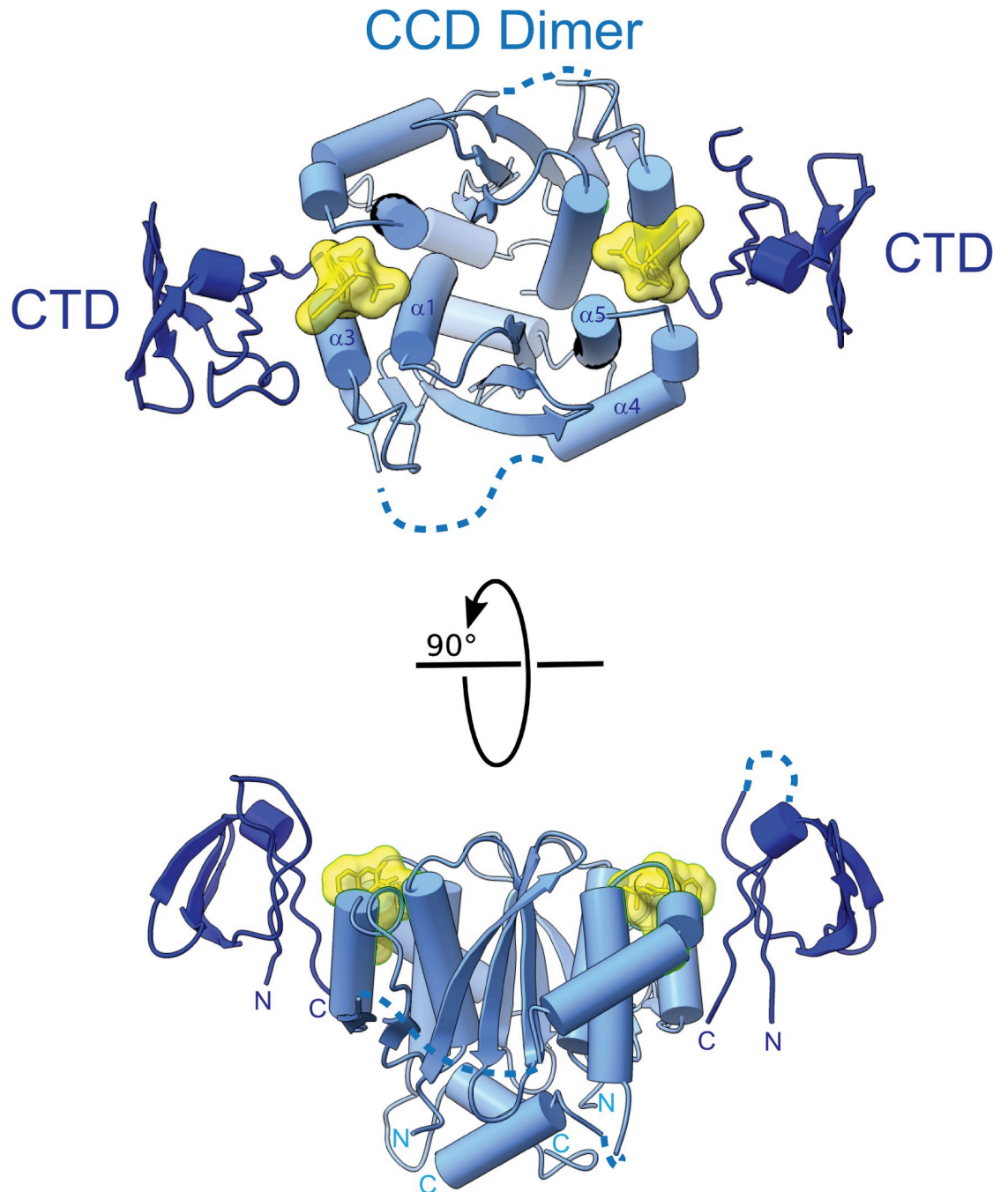
Fig 2. 2.93 Å X-ray Crystal Structure of CCD^{F185K}•BI-224436•CTD. Orthogonal views of the 2:2:2 ternary complex. The CCD is colored in dark blue and the CTD is colored in light blue. The BI-224436 molecule is shown in yellow. Disordered linkers within the CCD and CTD not observed within the electron density are shown as dotted lines.

complex is associated with the CTD dimerization modes observed in the crystal lattice. While it is known that CTD dimerization is necessary for drug-induced aggregation [52], it remains to be determined whether these observed CTD-CTD interfaces in this minimal structure exist with the full-length protein in virions.
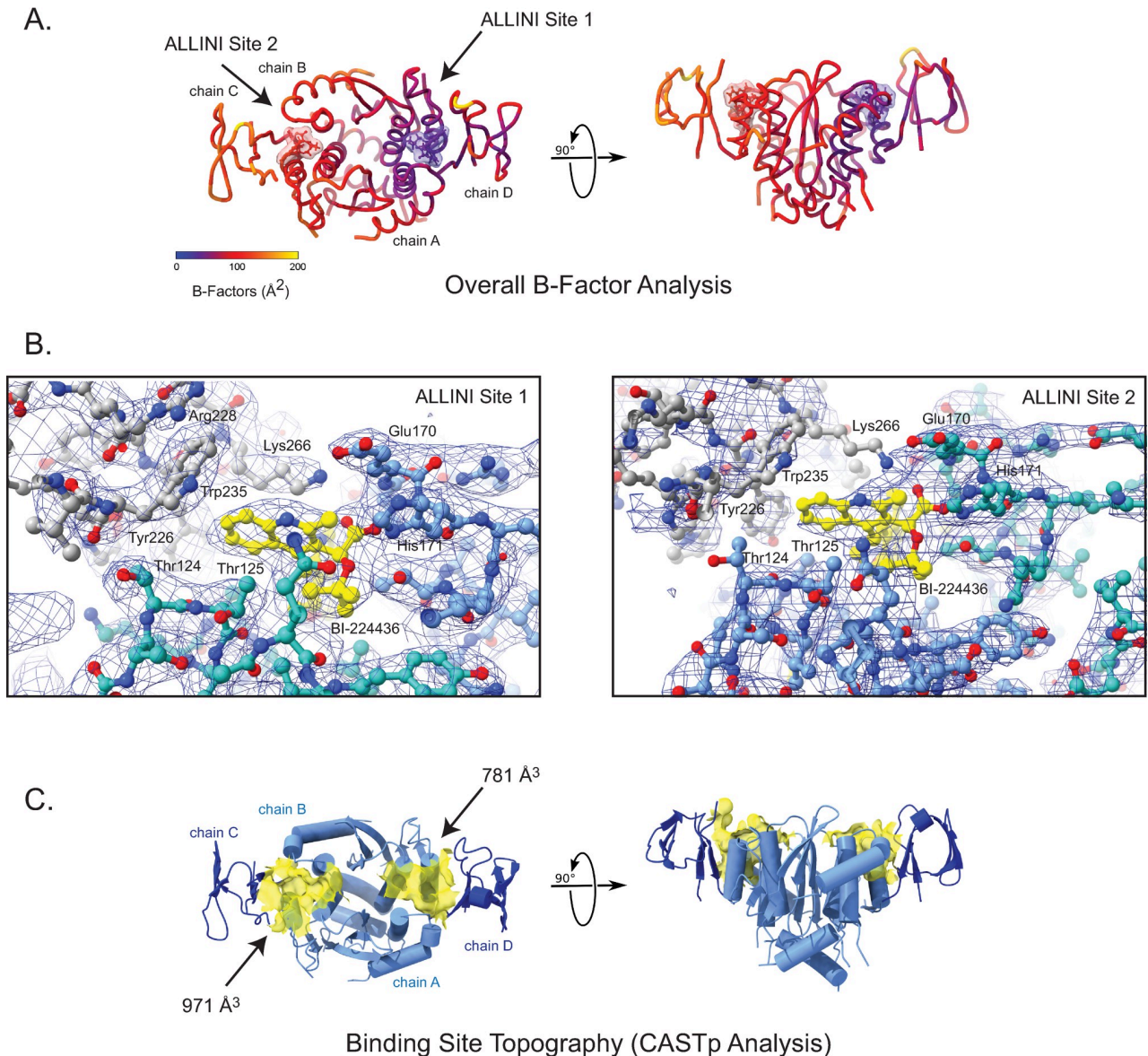
Fig 3. Evidence for asymmetry in the minimal ternary complex. A. B-factor analysis. A color key depicts the overall B-factor of the atomic structure. B. Representative sigma A-weighted $2F_o$-F_c electron density, contoured at 1σ at the first (left) and second (right) ALLINI binding sites. In the second site, side chain electron density is lost or diminished at key interfacial interactions including Lys-266, Arg-228, and Thr-124. C. Analysis of the surface of the ALLINI binding site using CASTp reveals at large binding pocket volume at ALLINI site 2 (971 A³) when compared to ALLINI site 1 (781 A³).

https://doi.org/10.1371/journal.ppat.1011097.g003

Substituents that anchor BI-224436 to the CCD

The new ternary complex structure provides the opportunity to define interactions more precisely between ALLINI and the CCD in the context of the complete ALLINI interface. The CCD dimer accounts for 440 Å² of the 534 Å² buried drug surface within the ternary complex. From one subunit, the α4 and α5 helices provide specific drug interactions with the side chains of Glu-168, Ala-169, Glu-170, His-171, Lys-173, Thr-174, and Met-178 (Fig 5A and 5B). From the second subunit, the binding interface is defined by the α1 and α3 helices and includes a series of specific drug interactions with the side chains of Gln-95, Ala-98, Tyr-99, Leu-102, Thr-124, Thr-125, Ala-128, Ala-129, and Trp-132. While Trp-131 contributes significantly to
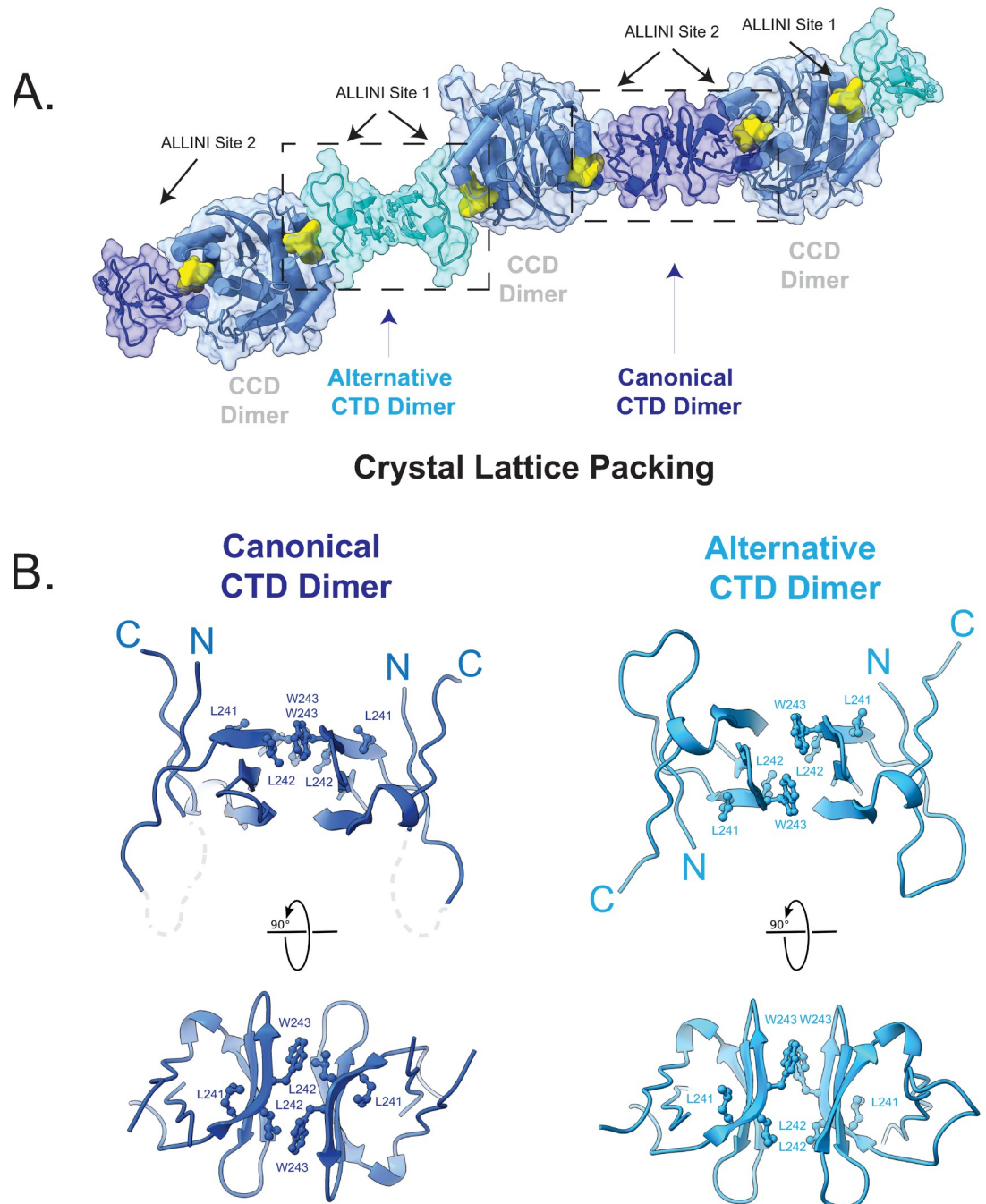
Fig 4. Crystal Lattice Packing at the C-terminal domain of HIV-1 IN. A. Crystal lattice packing. Shown is the lattice packing observed when crystallographic symmetry operations are applied. CCD dimer is shown in blue, the canonical CTD dimer in dark blue, the alternative CTD dimer in cyan, and the drug BI-224436 in yellow. Two homomeric CTD interactions are denoted (canonical and alternative), along with their position relative to ALLINI binding sites 1 and 2. B. Shown in orthogonal views are the canonical (light blue) and alternative CTD dimers (dark blue) observed. In both dimers, the interface involves the β2, β3, and β4 sheets and is predominantly hydrophobic, with the most significant contributions from residues Val-240, Leu-242, Trp-243, and Ile-257. In the canonical dimer, loop residues between the N-terminus and β2 are disordered and depicted as a dotted grey line.

https://doi.org/10.1371/journal.ppat.1011097.g004
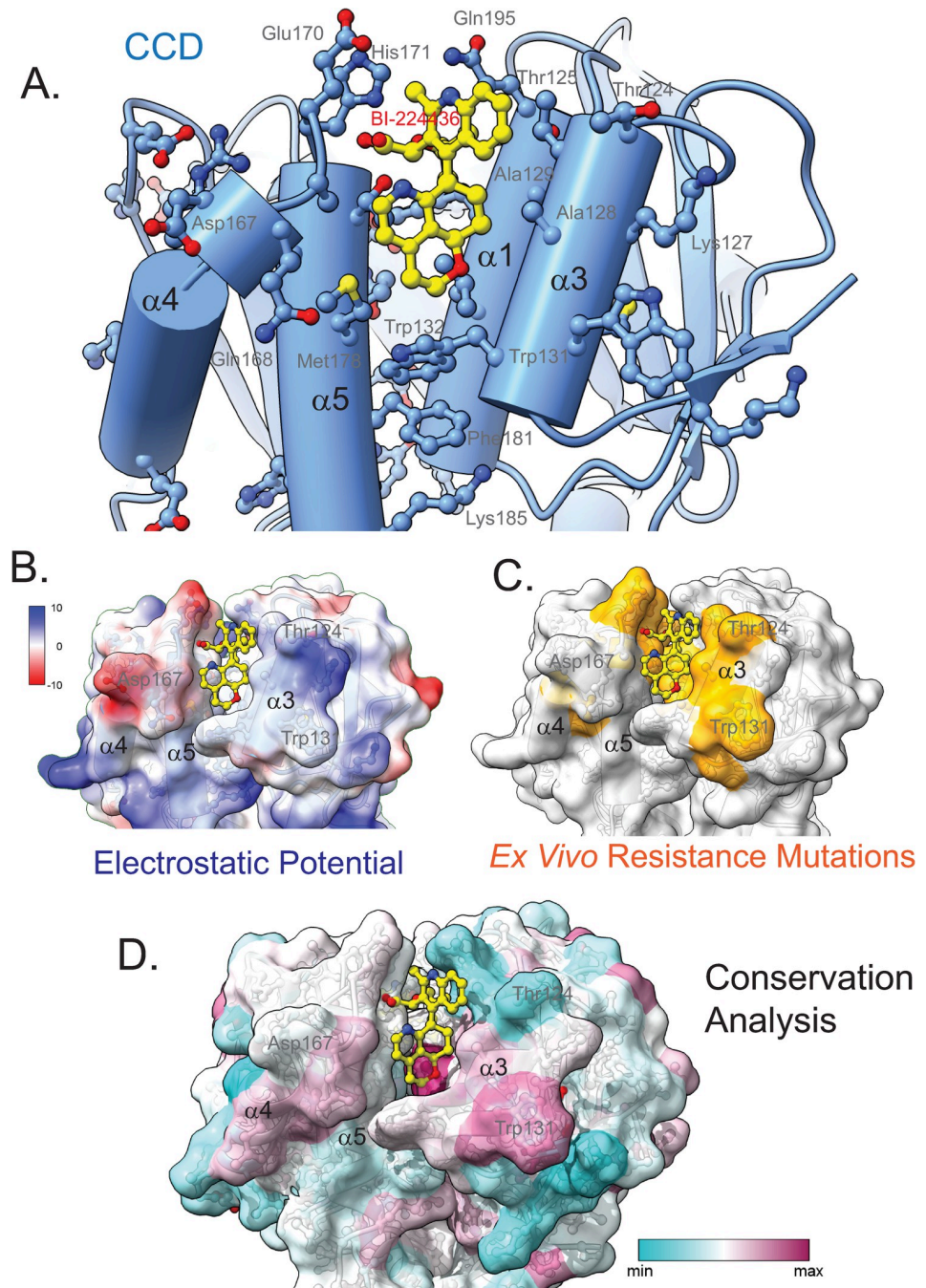
Fig 5. CCD interactions with BI-224436 within the minimal ternary complex. A. Shown is a view of the site I BI-224436 binding site on the CCD, with the CTD omitted for clarity. Helices α3 and α1 from one monomer subunit and α5 and α4 from another monomer subunit are shown cradling the drug (yellow). B. Electrostatic surface of ALLINI-bound CCD, with the CTD omitted for clarity. Red denotes electropositive surfaces and blue denotes electronegative surfaces. C. Known *ex vivo* resistance mutation sites are shown in orange on the same ALLINI-bound CCD view. D. Conservation analysis, using the 2019 HIV IN collection of sequences at the Los Alamos database that includes all subtypes (4,326 sequences) of HIV-1 IN, as well as subtype M from HIV-1/SIVcpz (2,471 sequences). Alignments were prepared and used in the calculation of sequence conservation scores using the entropy-based method from AL2CO [63] program implemented in the UCSF ChimeraX [64]. All Figs were created using the program UCSF ChimeraX. Regions of high conservation are shown in red and regions of low conservation are shown in blue.

https://doi.org/10.1371/journal.ppat.1011097.g005

the CCD-CTD protein interface stabilized by ALLINI binding, no specific drug interactions between Trp-131 and the drug occur. Most known *ex vivo* resistance mutations map to this ALLINI binding region on the CCD (Fig 5C), whereas none have been reported within the SH3-like fold of the CTD to date. The sites of variability found in resistance studies correlates strongly with conservation of the IN CCD (Fig 5D).

Among the ALLINI class, chemical moieties extending from the core pharmacophore closest to the CCD hydrophobic dimer interface (Fig 6A) anchor the molecules and underlie much of the interaction strength at the site. Like other molecules reported to-date within this class, BI-224436 features a carboxylate that interacts with backbone amide of IN residues Glu-170 and His-171, with additional hydrogen bonds to Nδ of His171 and the hydroxyl oxygen of Thr-144 (Fig 6B). Both Glu-170 and His-171 accumulate resistance mutations *ex vivo* in serial passage experiments in the presence of ALLINIs [10,50]. These interactions are observed throughout CCD-ALLINI crystal structures and in the IN•GSK1264 structure.

A second feature of BI-224436 and common to other ALLINIs is a *tert*-butoxy moiety, which emulates an isoleucine residue found at that position in the LEDGF(IBD)•CCD complex (PDBs 2B4J [65], 3F9K [20]). This ALLINI functional group is required for high-affinity interaction and binds in a pocket lined by residues Gln-95, Tyr-99, Thr-125, and Thr-174 (Fig 6B), all of which are subject to resistance mutation *ex vivo* [10,50,66] A third substituent extends from the 4-position of the central ALLINI pharmacophore (Fig 1B). This is typically a bulky hydrophobic group that is cradled by the side chains of Leu-102, Ala-128, Ala-129, Trp-132, and Met-178 (Figs 5A and 6C). Ala-128 is the location of one of the most common resistance mutants to arise against a variety of chemotypes in this drug class *ex vivo*, including BI-224436 [36,58,67] (Fig 6C). As shown in Fig 5A, Ala128 makes direct Van Der Waals contact with the R4 functional group and substitutions at this position are expected to interfere with ALLINI binding. Notably, some *ex vivo* CCD mutations do not lie anywhere in proximity to the drug binding site or CCD-CTD interface but would be predicted to instead confer oligomeric defects to the enzyme, which is prerequisite to the formation of the ALLINI binding site. For example, A205P resides at the CCD dimer interface and renders resistance to both GSK002 and GSK1264 [10], while the KF116 resistance mutation V165I confers oligomeric defects in IN [66].

CTD side chain interactions with BI-224436

Four well-conserved CTD residues interact with BI-224436: Tyr-226, Trp-235, Lys-266, and Ile-268 (Fig 6B–6D). A total of 147 Å^2 of drug surface is buried by these contacts. While the positions of these residues were inferred from the lower resolution IN-GSK1264 structure, here the side chain electron densities are clear. Trp-235 contributes a favorable π-π interaction to the aromatic BI-224436 pharmacophore, perpendicular to the plane of the CCD dimer interface and the drug's planar aromatic pharmacophore moiety. Mutagenesis of this residue to alanine abrogates drug-induced aggregation [52]. The position of the Trp-235 side chain is stabilized by a cation-π interaction with Arg-228 as determined by CaPTURE analysis [68] (Fig 6B); in the weaker ALLINI site, this interaction is not seen, as the Arg-228 side chain is disordered. The position of this Arg-Trp interaction suggests that despite its solvent accessibility, the side chain conformation of Trp-235 is constrained and may not adopt alternative rotamers in response to different chemotypes in the ALLINI class (discussed more below).

The CCD•CTD interface has a strong electrostatic component, with the acidic face of the α5 helix of the CCD packing against the basic β5 strand of the CTD. Emanating from the β5 strand is Lys-266, which when mutated to alanine abrogates ALLINI-induced aggregation of IN [56]. In our model, this lysine side chain electron density is clearly defined in ALLINI site
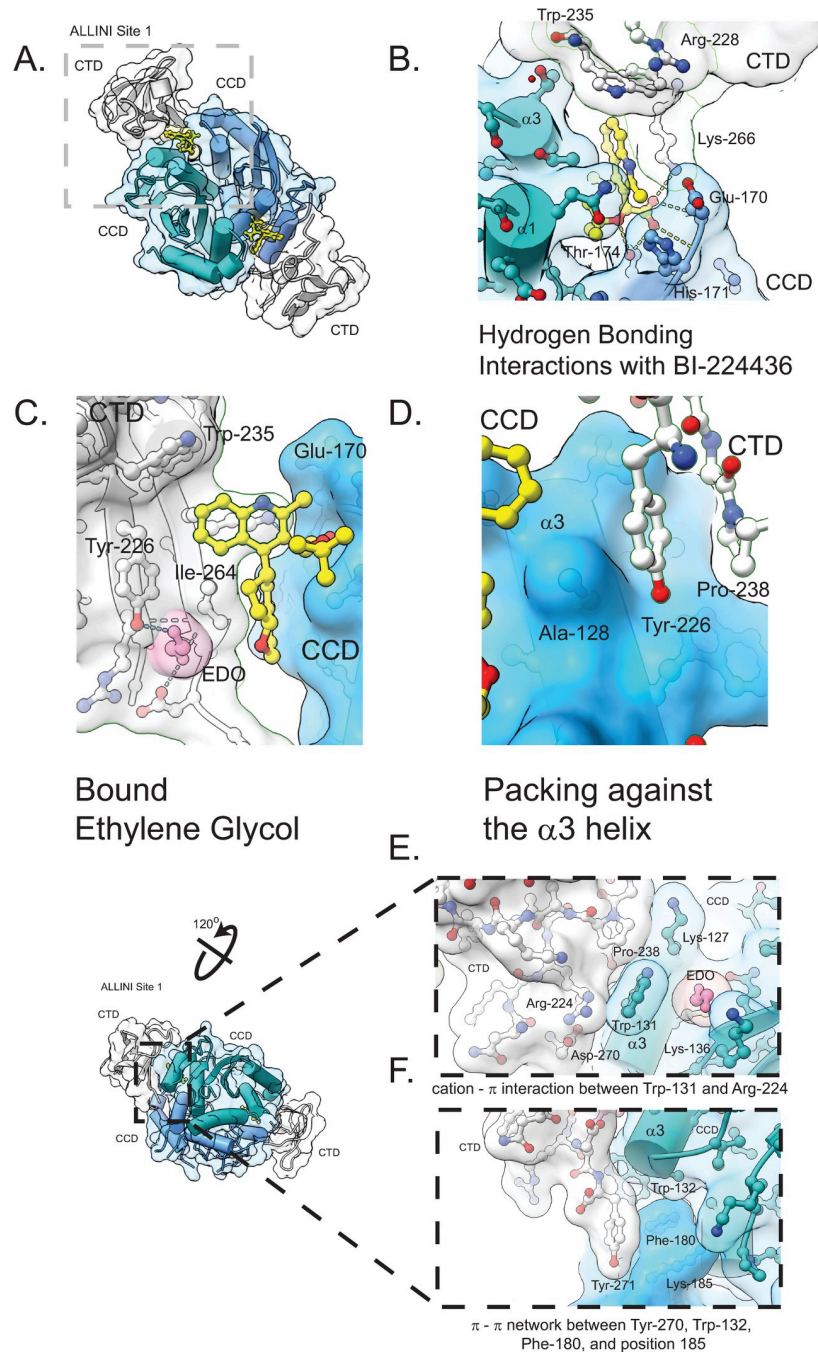
Fig 6. CTD interactions within the minimal ternary complex. A. Shown is the ternary complex, with CCD monomer chains colored in blue and cyan, BI-224436 in yellow, and the CTD colored in white. are Trp-235, Lys-266, and Ile-268 interactions with BI-224436 in ALLINI site 1. B. Hydrogen bonding interactions between Lys-266 from the CTD, the carboxylic acid moiety of BI-224436, and Glu-170, His-171, and Thr-174 from the CCD are highlighted. C. Shown is a bound molecule of ethylene glycol (EDO) within ALLINI binding site 1, created by the observed conformational change in Trp131, where bridging hydrogen bonds with Tyr-226 and Asp-270 are observed. D. Interactions between Tyr-226 and Ala-128, a common *ex vivo* resistance mutation site, on the α3 helice is shown. E. A distal interdomain cation-π interaction network between Trp-131, Arg-224, and Asp 270 is shown. Bound behind the indole plane of the Trp-131 side chain is a well-ordered molecule of EDO, precluding further stabilizing cation-π interactions by Lys-127 and Lys-136. E. An aromatic π-π network between Tyr-271 and a cluster of aromatics at the C-terminal base of the α3 helice in the CCD dimer is shown, including Trp-132, Phe-180, and Lys-185. In wild type HIV IN-1, position 185 is occupied by a phenylalanine.

https://doi.org/10.1371/journal.ppat.1011097.g006

1, adopting an extended rotamer that enables hydrogen bonding with the carboxylate of BI-224436 (Fig 6B). This side chain is also well positioned to interact with the backbone amide of Glu-170 and the ALLINI R4 substituent. Additionally, this side chain is well-positioned to provide cation-π interactions with the ALLINI ring systems. However, in ALLINI site 2, the Lys-266 side chain has weaker electron density. In HIV infected cells, the Lys-266 side chain is N6 acetylated [69–75]. Modelling suggests that this modification could be accommodated in the ALLINI binding interface with only modest adjustments to the side chain rotamer. New interactions could be formed with BI-224436, but the cation-π interaction would be lost.

The second major component of the CCD-CTD interface is a hydrophobic interaction between the α3 helix of the CCD and the β1 and β2 strands of the CTD (Fig 6D). Tyr-226 forms the core of this interface and the alanine substitution abrogates drug-induced aggregation *in vitro* [10]. However, this side chain provides minimal direct contact with BI-224436. Instead, it directly abuts Ala-128 from the CCD α3 helix (Fig 6D). In serial passage resistance studies with BI-224436, the mutations A128T and A128N arise [37,58], which would be predicted to clash with the Tyr226 side chain in addition to the R4 substituent noted earlier, resulting in a direct perturbation of the CCD-CTD interface.

Identification of a novel CTD pocket within the ALLINI binding site

The most notable difference observed between the ternary complex structure described here and the CCD^{F185H} alone-BI-224436 structure (PDB 6NUJ [76]) is the conformational change observed for Trp-131 (S2 Fig). However, the published CCD•BI-224436 structure does not have the F185K solubility substitution that we used here and that is frequently found in the PDB. To assist our comparison of these complexes, we determined structures for unliganded CCD^{F185K} and CCD^{F185K}-BI-224436 forms at 1.9 Å and 2.1 Å resolution, respectively (S2 Fig and Table 1). The CCD•BI-224436 structure is nearly identical to the CCD^{F185H}•BI-224436 structure determined by others (PDB 6N4J [76], Cα RMSD <0.2Å across matching atoms). In these reference structures, Trp-131 adopts a rotamer where the plane of the indole lies perpendicular to the long axis of the α3 helix. However, when in complex with both CTD and BI-224436, a conformational change is observed in ALLINI site 1, where the side chain of Trp-131 is rotated ~120˚ away from the CTD.

A modest number of well-ordered solvent molecules can be modelled within the asymmetric unit of the ternary complex (Table 1). Consistent with the high complementarity observed between protein and ALLINI and with previously reported CCD-ALLINI structures, ordered buried water molecules are rare within the ALLINI binding site. An exception to this is an elongated segment of electron density in a pocket defined by CTD residues Tyr-226, Ile-268, Arg-269, and Asp-270, and CCD residues Ala-128, Trp-131, and Trp-132. This pocket was created by the observed conformational change in Trp-131 (S2 Fig) that occurs upon CTD binding. The density is best modelled by a molecule of ethylene glycol and is only observed in the more ordered ALLINI site 1 (Fig 6C). The bound ethylene glycol molecule comes within Van Der Waals contact of the R4 substituent of BI-224436, suggesting that future development of the ALLINI class may be able to exploit this pocket.

Distal protein-protein interactions at the CCD•CTD interface

This structure allows for the observation of specific protein-protein interactions away from the bound drug not previously possible at lower resolution. In the configuration observed for Trp-131, the indole lies within Van Der Waals contact of Pro-238 from the bound CTD and ~2.9 Å from the Arg-224 side chain, in a configuration identified as a strong cation-π interaction by the program CaPTURE [68] (Fig 6E). Asp-270 forms a salt bridge with Arg-224, facilitating

this interaction. The potent *ex vivo* ALLINI resistance mutation W131C [10,52] would severely perturb this network of interactions.

A second distal CCD-CTD interaction involves Tyr-271 of the CTD packing against a π-π cluster of aromatic residues found at the C-terminal side of the α3 helix in the CCD (Fig 6F). The cluster includes Trp-132, Phe-181, and Phe-185, which is lysine in the current structure. Analysis using the program RING [77] identifies a face-to-edge π-π interaction between Tyr-271 and Phe-180, alongside a face-to-face π-π interaction between Trp-132 and Phe-180. No known ALLINI resistance mutants have been reported at this well-conserved distal site.

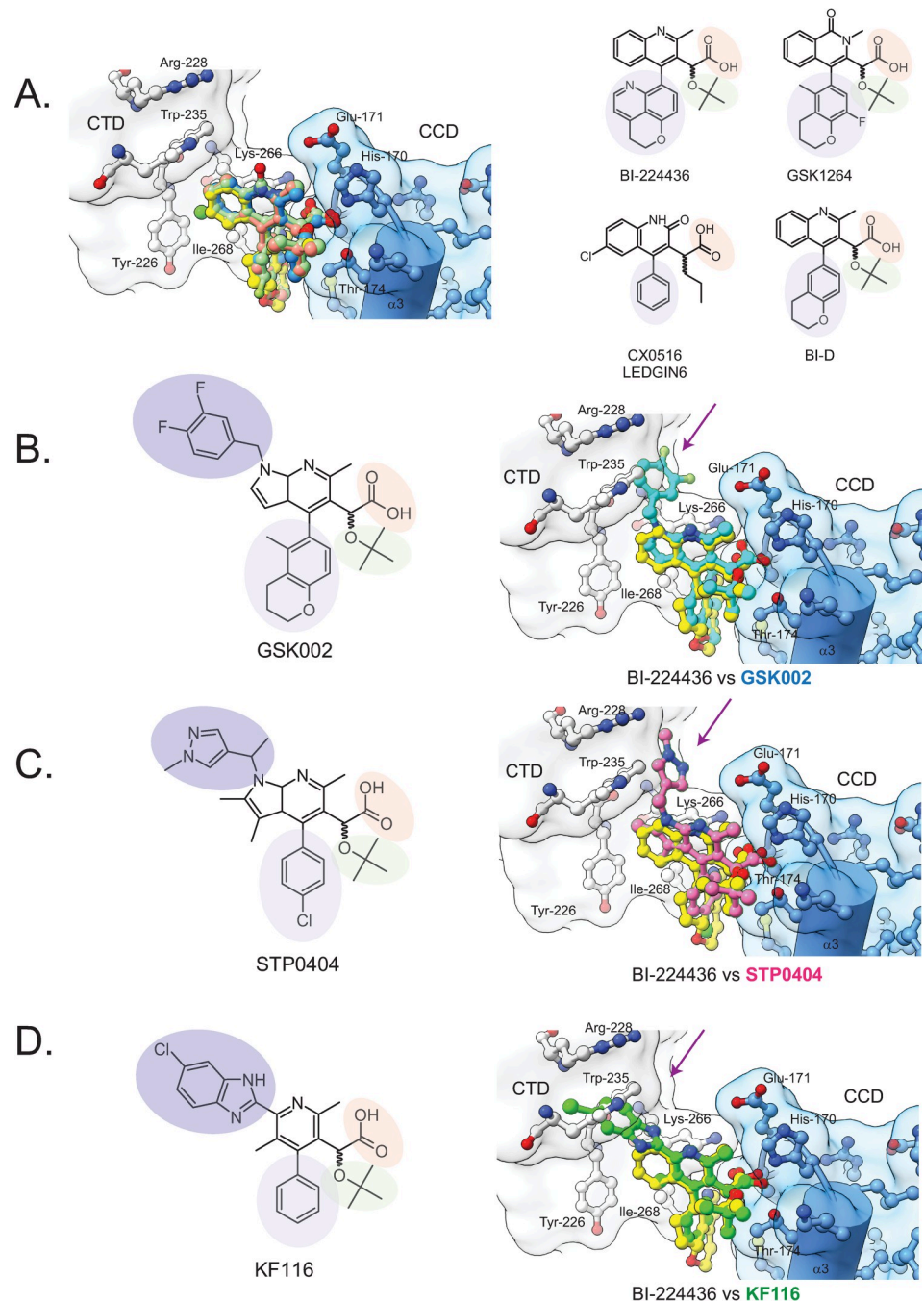## A generalized model for the ALLINI-stabilized CCD•CTD interface

To investigate the generality of the observed BI-224436 interactions, several different ALLINI•CCD crystal structures were superimposed onto the ternary complex determined here, using the better ordered ALLINI site 1 (Figs 7A, S3A, and S3B). In the cases of LEDGIN-6 (PDB 3LPU [36]), BI-D (PDB 4ID1 [47]), GSK1264 (PDB 4OJR [11]), Mut101 (PDB 4LH5 [48]), and indole acetic acid derivative (PDB 5KGW [78]), the molecules are readily accommodated within the CCD•CTD interface without the need for further minimization or manual adjustment of side chain positions or drug rotamers. Similarly, antiviral 2-thiopenecarboxylic acid derivatives [79] identified in crystallographic fragment screening that are competitive with LEDGF(IBD) binding are readily accommodated. However, they do not make any predicted interactions with the CTD (S3C Fig); their ability to induce IN aggregation has not been reported.

In recent years, more complex chemotypes have emerged with molecular weights >500 Daltons, including the ALLINIs KF116 (PDB 4O55 [49]), GSK002 (PDB 5HRN [10]), an isoquinoline derivative (PDB 6EB2 [81]), and the recently reported preclinical lead molecule STP0404 that is entering human trials (PDB 7KE0 [80]). ALLINI•CCD structures are available for all four molecules, allowing us to model their predicted interactions with the CTD.

The difluorobenzyl moiety of GSK002 would be predicted to clash with Trp-235 but could be accommodated by a rotation of the difluorobenzyl group or of Trp-235 (Fig 7B). STP0404 features a 1-methylpyrazole substituent which is readily accommodated in our structural model, with this functional group nestled between the side chains of Trp-235, Lys-266, and the α5 helix (Fig 7C). Similarly, the benzylpyrazol extension of the isoquinoline derivative reported by Wilson et al [81] would be predicted to clash with Tyr-226, which is readily accommodated in our structural model by rotamer adjustments of the side chain and the extension (S3D Fig). In contrast, the ALLINI KF116 [49] cannot be readily modeled in the interface of the ternary complex described here. The chlorinated benzimidazole ring system is predicted to clash with Trp-235 of the CTD (Fig 7D). Attempts to alleviate the steric clash by simple rotations of Trp-235 or of the benzimidazole ring were unsuccessful, implying that an alternative CTD position or side chain configuration is required to form the CCD•KF116•CTD interface. As can be seen for HIV subtype M (the major driver of the global pandemic), high levels of conservation are observed along the ALLINI-binding face of the CTD, including residues Trp-235, Lys-266, and Glu-170 that contact the 1-methylpyrazole substituent of STP0404 (S4 Fig). Thus, the barrier to development of resistance may be high in this region of the ALLINI binding pocket.

Discussion

We have determined a 2.93 Å structure of a minimal two-domain ternary complex between isolated CCD and CTD domains and the preclinical ALLINI BI-224436, revealing the highest resolution structural model to date of the full ALLINI binding site. We report eight

Fig 7. A generalized model for ALLINI Binding. A. Using available isolated CCD•ALLINI structures in the Protein Data Bank, models of ALLINI class members in the minimal ternary complex structure were created to provide a generalized model for IN-ALLINI binding. Shown on the right are the chemical structures of the ALLINI modelled, with shared features highlighted: carboxylic acid moiety (orange), tert-butoxy moiety (green), large hydrophobic moiety (lavender), and an CTD-interacting moiety (purple). The following models of the ternary complex are shown: BI-224436 (yellow), LEDGIN-6 (salmon, PDB 3LPT [36]), BI-D (green, PDB 4ID1 [47]), and GSK1264 (blue, PDB 4OJR [11]). B. Modeling of GSK002 binding by superposition of the available CCD-only structure (PDB 5HRN [11]) with the ternary complex structure. Highlighted is a clash between the diflourobenzyl moiety at R6 with the CTD residue Trp-235. C. Modelling of STP0404 binding by superposition of the available CCD-only structure (PDB 7KE0 [80]) with the ternary complex structure. Highlighted is a clash between the 1-methylpyrazole moiety at R6 with the CTD residue Trp-235. D. Modeling of KF116 binding. A predicted clash between its distal ornament and Trp 235 is shown.

independent observations of the complete drug binding site in the asymmetric unit. Four of the binding sites are well-ordered with strong electron density (site 1) and four are less well ordered with weaker density (site 2). A detailed analysis of the protein-drug interface suggests multiple avenues for improved inhibitor design.

The CTD of IN has been implicated in an array of functions in the HIV replication cycle including IN oligomerization [9,11], reverse transcription [82], binding host factors including the nucleosome [83], Gemin2•SMN [84–86] and TNPO3 [87], intasome formation and integration (reviewed in [88]); and viral RNA interactions during transcription [89] and virion maturation [54,90,91]. Given its multifaceted role, it is not surprising that the CTD is relatively intolerant of sequence substitution [92]. Thus, targeting this domain with small molecules may result in high barriers to the development of genetic resistance. Indeed, the only ALLINI resistance mutation that has been reported in this region is the N222K substitution [10,50] which is located just before the start of the SH3-like fold of the CTD.

Cation-π interactions play important roles in physiological processes such as acetylcholine binding [93], as well as drug discovery [94]. Our new structure reveals previously unobserved cation-π and π-π side chain interactions that underlie ALLINI-induced protein-protein and protein-drug interactions. Given that over 40% of small molecule drugs in development or in clinical trials between 2009 and 2018 contained heteroaromatic rings [95], these potential interactions should be considered in future drug-design efforts. Modulating the strength of these interactions is the dielectric constant of the surrounding medium. ALLINIs exert their primary effect in the crowded milieu of the nascent virion, where macromolecular crowding would be expected to significantly impact the solvent environment and effective dielectric constant [96]. This aspect of ALLINI biology may provide another tunable feature for drug optimization.

Modeling of the preclinical leads STP0404 and GSK002 suggests possible routes to improving ALLINI potency. These drugs interact with invariant residues on the CTD including Trp-235 and Lys-266. Thus, extending the functional groups on current scaffolds towards this part of the CTD may increase affinity. Such interactions would involve highly conserved CTD residues, and so potentially have a high genetic barrier to resistance. The incompatibility of the KF116 molecule with our model suggests that other modes of CCD•CTD•ALLINI interactions may exist; if so, it would be valuable to understand such complexes in detail to enhance future ALLINI designs.

We found two classes of ALLINI binding sites represented in the crystals of the ternary CCD•BI224436•CTD complex. The two site classes differ both by their average B-factors and by subtle changes in structure, where the less ordered site 2 adopts a somewhat more open, less compact configuration. It has been previously proposed that high and low affinity binding sites of LEDGF exist within the IN tetramer [12,97,98]. It has also been reported that full-length LEDGF/p75 binds the pol polyprotein cooperatively [99], and cooperative inhibition is observed with ALLINIs against IN *in vitro* and virus *ex vivo* [100]. The possibility of "structural crosstalk" between the two ALLINI binding sites warrants further investigation. In the case of the structure described here, we can identify differences in crystal packing associated with the two binding sites that could be the source of the differences observed. However, since these crystal contacts involve CTD-CTD interactions and we have shown that such interactions play a central role in the formation of ALLINI-induced aggregates *in vitro* [52], we cannot exclude the possibility that the same interactions, and thus non-equivalent ALLINI binding sites, are formed in HIV virions.

We have also identified a new binding pocket in the CCD-CTD interface that is occupied by a molecule of ethylene glycol in the new structure. The pocket is located adjacent to the ALLINI R4 substituent, suggesting that elaboration of R4 with chemical groups capable of

exploiting this pocket might be a useful direction for ALLINI development. Together, these new data provide a more detailed picture of the complete ALLINI binding site that is expected to exist in drug-induced HIV-IN aggregates and will provide much needed insights into ALLINI development.

Material and methods

Protein expression and purification

IN (NL4-3) constructs were expressed and purified as described previously [10,12,101,102]. Individual IN domains (CCD^{F185K} [11], CTD(220–271), CTD(220–270)^{L242A} [52], and CTD (220–288)) were inserted into an expression vector containing an N-terminal His7-Flag-SUMO tag [103]. Proteins were purified with nickel-nitriloacetic acid resin and subjected to a second nickel-nitriloacetic column after fusion proteins were liberated by cleavage with Ulp1 protease (Thermofisher Scientific). CCD constructs were further purified by size-exclusion chromatography using a Superdex 75 HiLoad 16/60 column (GE Healthcare) in 20 mM HEPES•NaOH pH 7.5, 300 mM NaCl, and 2 mM dithiothreitol (DTT). Purified proteins were concentrated in YM-3 or YM-10 Centricon devices (Millipore), flash-frozen with 10% glycerol in liquid N_2, and stored at -80˚C.

Inhibitor preparation

The inhibitors BI-224436 [58] was purchased from MedChemExpress (Monmouth Junction, N.J., U.S.A.). To create stock solutions for this study, drugs were resuspended in acetonitrile at 1- or 10-mM concentrations, aliquoted, and lyophilized. Dried aliquots were stored at -80˚C until use. Inhibitors were then resuspended to initial volumes using sample buffers for subsequent experimentation.

Turbidity assays and dynamic light scattering (DLS)

Purified CCD^{F185K} and CTD$^{220-288}$ in 20 mM HEPES•NaOH pH 7.5, 300 mM NaCl, and 0.1 mM TCEP were combined to a final concentration of 100 μM, and either DMSO or BI-224436 (100 μM final concentration) were added to initiate aggregation. Samples were incubated at room temperature for 30 minutes. Samples were then diluted 1:100 (1 μM final) for subsequent analysis in a Nanobrook Omni particle sizer (Brookhaven Instruments Corporation, Holtsville, NY, USA). Data were recorded at 25˚C in polystyrene 1-cm cells using a standard diode laser at 640 nm, with scattering recorded at an angle of 90˚. Three scans were recorded for each sample and hydrodynamic radii (Stokes radii) were calculated using the BIC Particle Solutions software v3.6.0.7122.

Analytical ultracentrifugation

Sedimentation velocity analytical ultracentrifugation (SV-AUC) experiments were performed at 20˚C with an XL-A analytical ultracentrifuge (Beckman-Coulter, Brea, CA) and a TiAn60 rotor with two-channel charcoal-filled Epon centerpieces and quartz windows. Experiments were performed in 20 mM HEPES•NaOH pH 7.5, 300 mM NaCl, and 1 mM DTT. Data were collected with detection at 280 nm. Complete sedimentation velocity profiles were recorded every 30 seconds at 40,000 rpm. Data were fit using the *c(S)* distribution implementations of the Lamm equation in the program SEDFIT (25) and corrected for S$_{20,w}$. Direct fitting of association models were performed using SEDPHAT [104]. The partial specific volume ($\bar{v}$), solvent density (ρ), and viscosity (η) were derived from chemical composition by SEDNTERP [105]. Figs were created using the program GUSSI [106].

Crystallization and structure determination

The CCD construct used contains the solubility-enhancing substitution F185K [9] and exists as a dimer in solution [11,52]. The CTD encompasses residues 220–288, with the latter 18 amino acids believed to be unstructured and are dispensable for viral replication [107]. A more minimal CTD truncation (residues 220–271) was chosen for this work based on the observation that Tyr-271 is the last residue of the CTD observed in the full-length IN-ALLINI structure (PDB 5HOT) to directly contact the CCD [10]. In control experiments, intact IN constructs truncated to residue 270 could be crystallized with GSK1264 in our previously published conditions, but with no improvement in resolution. CCD^{F185K} was concentrated to 8 mg/mL in 20 mM HEPES•NaOH pH 7.5, 300 mM NaCl, and fresh 1 mM DTT and crystallized using hanging drop vapor diffusion in a 1:1 drop mixture with a reservoir solution containing 7% PEG 8000, 0.2 M ammonium sulfate, 0.1 M sodium cacodylate, pH 6.5, 5 mM manganese chloride, 5 mM magnesium chloride, and 5 mM DTT as previous [11]. To prepare complexes with BI-224436, the apo crystals were harvested and soaked with 2.5 mM BI-224436 for three days at 4˚C. Apo or soaked crystals were transferred to 30% ethylene glycol in mother liquor and subsequently flash-frozen in liquid nitrogen. X-ray diffraction data were collected on a Rigaku Micromax007 rotating anode generator with CCD detector. Data collection statistics are summarized in Table 1. The structure was initially solved by molecular replacement with the CCD monomer from PDB 3L3U [108] using the program PHASER [109]. All structures were refined and rebuilt iteratively using PHENIX.refine [109], REFMAC [110], and *COOT* [111] using NCS restraints, TLS refinement, individual B-factor refinement, and energy minimization. All coordinates for the structures determined were validated using PDB-REDO [112] and deposited in the Protein Databank with accession codes 8CT5, and 8CT7.

CCD, CTD, and BI-224436 were co-crystallized using hanging drop vapor diffusion by mixing 100 μM CCD, 100 μM CTD, and 500 μM BI-224436 in 20 mM Tris•HCL pH 7.4, 250–300 mM NaCl, and 1 mM DTT with 25% ethylene glycol. Addition of 100 mM ammonium acetate improved crystal growth and yielded large single crystals >200 microns in length within one week. Crystals were harvested and frozen directly from hanging drops. Diffraction data was collected at NSLSII AMX beamline [113]. Data reduction was performed using *DIALS* [114]. Analysis of the final dataset by the UCLA diffraction anisotropy server ((http://services.mbi.ucla.edu/anisoscale/) [115] indicated that diffraction was significantly anisotropic along the a^*- and b^*-axes. Based on an $F/\sigma(F)$ cutoff of 3, reflections were subjected to an anisotropic truncation of 2.9, 3.2, and 2.9 Å along a^*, b^*, and c^*, respectively, before use in refinement. Data collection statistics are summarized in Table 1. The analyses of the Patterson function reveal a significant off-origin peak that is 41.4% of the origin peak, indicating strong translational NCS. Molecular replacement and refinement were carried out using PHASER [116] as implemented in the program PHENIX [109]. A solution was determined using the CCD dimer (PDB: 3L3U) [108] as a search model. The CTD (PDB: 1IHV[17]) could be placed with subsequent searches. The structure was refined iteratively using PHENIX.refine [109], REFMAC [110], and *COOT* [111] using NCS restraints, TLS refinement, individual B-factor refinement, and energy minimization. All coordinates for the structures determined were validated using PDB-REDO [112] and deposited in the Protein Databank with accession codes 8CT7.

Structure analysis

The program PISA [117] was used to assess protein and drug buried surface areas. The program CASTp [61] was used to analyze the surface and volume of the ALLINI binding sites. The program CaPTURE [68] (Cation–Trends Using Realistic Electrostatics) was used to

identify interactions between the cationic group of lysine or arginine and the aromatic rings of phenylalanine, tyrosine and tryptophan. The program RING [77] was used to characterize π-π interactions within the protein structure.

Alignments of representative subtypes of HIV IN were prepared from sequences extracted from curated HIV Sequence Alignments available at the Los Alamos HIV Sequence Database [118]. Using the HIV1 NL4-3 sequence as a reference (GenBank: AAA44988.2), the 2019 HIV IN collection of sequences that includes all subtypes (4,326 sequences), as well as subtype M from HIV-1/SIVcpz (2,471 sequences). The pre-aligned DNA sequences were translated in-frame one using the Los Alamos Translate Database tool (https://www.hiv.lanl.gov/content/sequence/translatev2/translate.html) and culled of ambiguous sequences. Using this data, alignments were prepared for Subtype M (6,724 sequences), Subtype B (1,282 sequences), and Subtype C (720 sequences) and used in the calculation of sequence conservation scores using the entropy-based method from AL2CO [63] program implemented in the UCSF ChimeraX [64]. All Figs were created using the program UCSF ChimeraX. Several structure coordinates available in the PDB database were used in the present studies, which can be located under accession numbers: 3LPU [36], 4ID1 [47], 4OJR [11], 4LH5 [48], 4O55 [49], 5HRN [10], 5KGW [78], 5KRS [79], 5KRT [79], 6EB2 [81], 7KE0 [80].

Supporting information

S1 Fig. ALLINI-induced polymers of HIV-1 IN CCD^{F185K}•CTD$^{220-288}$. A. Shown are 100 μM mixtures of CCD^{F185K}•CTD$^{220-288}$ in the absence or presence of 100 μM BI-224436. In the presence of ALLINI, modest turbidity is observed. B. Shown is autocorrelation data from dynamic light scattering (DLS) for 1 μM BI-224436 alone (orange), 1 μM CCD^{F185K}•CTD$^{220-288}$ with DMSO (yellow), and 1 μM CCD^{F185K}•CTD$^{220-288}$•BI-224436 (purple). Strong scattering data arises when both protein domains and drug are present, coinciding with observed turbidity. C. Particle distribution analysis of autocorrelation data shown in panel B. Data were well-described by a unimodal model with a particle diameter near ~10^4 nm. D-E. Sedimentation velocity analytical ultracentrifugation (SV-AUC) analysis on the interaction of CCD^{F185K} and CTD(220–270)^{L242A} (light blue) in the presence of BI-224436 (light blue). In the leftmost panels, fits of the experimental data (circles) to the Lamm equation are shown as lines; in middle panel the residuals from this fitting are shown. Every third boundary and are shown for clarity. Measurements were performed at 10–30 μM monomer concentrations at 20°C. D&E. c(S) (D), van Holde-Weischet (inset, D), and c(M) (E) distributions were derived from the fitting of the Lamm equation to the experimental data collected, as implemented in the program SEDFIT, with an overall RMSD of <0.005 for all fits. CTD^{L242A} data is shown in blue, CCD^{F185K} in light blue, and complex data in purple. This analysis shows evidence of ternary complex formation in the presence of BI-224436.
(TIF)

S2 Fig. Structural Comparisons. A. Packing of the crystallographic asymmetric unit, showing four ternary complexes, which in turn provides eight independent observations of the ALLINI binding site. Representative sigma-weighted 2F$_o$-F$_c$ electron density, weighted at 1σ, is shown. B. Superposition of the Cα-traces of apo CCD^{F185K} (salmon) with CCD^{F185K}•BI-224436 (blue). Highlighted is the side chain of Trp-131. The two chains with an RMSD of 0.3 Å over matching atoms, with only slight discrepancies in backbone atoms observed across the α4 region (residues 138–155). C. Superposition of the Cα-traces of apo CCD^{F185K} (salmon), CCD^{F185K}•BI-224436 (blue), and CCD^{F185K}•BI-224436•CTD ternary complex (tan). Boxed in grey dotted lines and shown is inset is the α3 helix at the ALLINI binding interface, highlight the conformational change occurring with Trp-131 upon complex formation. D. Comparison

of the Trp-131•Arg-224 cation-π interaction at ALLINI sites 1 and 2. On the left, a superposition of the Cα traces for the four observations of the first ALLINI site is shown, with preservation of this interaction observed across all four protomers. On the right, a similar rendering of the Cα traces for the four observations of the second ALLINI site. In this second site, significant variation and perturbation of the cation-π interaction is observed.
(TIF)

S3 Fig. A generalized model for ALLINI Binding. A. Shown on the right are the chemical structures of BI-224436 and GSK1264, with shared features highlighted: carboxylic acid moiety (orange), tert-butoxy moiety (green), large hydrophobic moiety (lavender), and an CTD-interacting moiety (purple). Shown on the right are models of the ternary complex: BI-224436 (yellow), GSK1264 from the CCD-only structure 4OJR [11], and GSK1264 in complex with intact IN (PDB 5HOT [10]). B. Modeling of (2S)-tert-butoxy[3-(3,4-dihydro-2H-1-benzopyran-6-yl)-1-methyl-1H-indol-2-yl]acetic acid binding by superposition of the available CCD-only structure (PDB 6EB2 [81], orange) with the ternary complex structure. C. Modeling of 3-(1H-pyrrol-1-yl)-2-thiophenecarboxylic acid derivatives (PDBs 5KRS (red) and 5KRT (green) [79]) with the ternary complex structure. D. Modeling of (2S)-[1-(1-benzyl-1H-pyrazol-4-yl)-3-(3,4-dihydro-2H-1-benzopyran-6-yl)isoquinolin-4-yl](tert-butoxy)acetic acid (PDB 5KGW [78] with the ternary complex structure. Highlighted is a clash between benzylpyrazol extension of the isoquinoline derivative with the CTD residue Tyr-226.
(TIF)

S4 Fig. Sequence conservation of the CCD and CTD domains at the ALLINI binding interface. The DNA sequence alignments of HIV Integrase were extracted from the 2019 HIV Sequence Alignments published at the Los Alamos database[1]. A non-redundant collection of 4326 sequences representing all strains, as well as 2471 sequences of Group M from HIV-1/SIVcpz organism were extracted and translated. The HIV IN sequences of subtypes B and C, on the other hand, were selected from the collection of Group M sequences. Protein sequences were aligned against HIV-1NL4-3 (GenBank: AAK08484.2) using standalone Clustal Omega (version 1.2.3) [119–121]. Conservation scores of sequence alignments were calculated using the entropy-based method from AL2CO [63] program implemented in the UCSF ChimeraX [64]. Regions of high conservations are shown in magenta while regions of low conservations are in cyan. Top: comparison of sequence conservation of the CCD and CTD domains at the ALLINI binding interface between HIV-1 all strains, Group M without circulating recombinant forms (CRFs), Subtype B, and Subtype C. Consistent among the panel, is the high sequence conservation observed along the CTD, which includes Trp-235, Lys-266, and Tyr-226. Glu-170, Ala-169, and Thr-174 at the CCD are moderately conserved. Bottom: comparison of sequence conservation of the CCD dimeric interface at the ALLINI binding site. Here, the region of low conservation, which includes Gln-95, Thr-125, and Thr-124, remained consistent.
(TIF)

Acknowledgments

School of Medicine (Philadelphia, PA) with the support of an NIH High-End Instrumentation Grant (S10-OD018483).

Author Contributions

Conceptualization: Grant Eilers, Kushol Gupta, Saira Montermoso, Frederic D. Bushman, Gregory Van Duyne.

Formal analysis: Grant Eilers, Kushol Gupta, Audrey Allen, Saira Montermoso.

Funding acquisition: Kushol Gupta, Frederic D. Bushman, Gregory Van Duyne.

Investigation: Grant Eilers, Kushol Gupta, Audrey Allen, Saira Montermoso, Hemma Murali, Robert Sharp, Young Hwang.

Writing – original draft: Grant Eilers, Kushol Gupta, Frederic D. Bushman, Gregory Van Duyne.

Writing – review & editing: Grant Eilers, Kushol Gupta, Frederic D. Bushman, Gregory Van Duyne.

References

1. Clavel F, Hance AJ. HIV drug resistance. N Engl J Med. 2004; 350(10):1023–35. Epub 2004/03/05. https://doi.org/10.1056/NEJMra025195 PMID: 14999114.

2. Brenner B, Wainberg MA. We need to use the best antiretroviral drugs worldwide to prevent HIV drug resistance. AIDS. 2016; 30(17):2725–7. Epub 2016/10/27. https://doi.org/10.1097/QAD.0000000000001234 PMID: 27782967.

3. Craigie R, Bushman FD. HIV DNA Integration. Cold Spring Harb Perspect Med. 2012; 2(7):a006890. Epub 2012/07/05. https://doi.org/10.1101/cshperspect.a006890 a006890 [pii]. PMID: 22762018; PubMed Central PMCID: PMC3385939.

4. Engelman AN, Cherepanov P. Retroviral intasomes arising. Curr Opin Struct Biol. 2017; 47:23–9. Epub 2017/05/02. https://doi.org/10.1016/j.sbi.2017.04.005 PMID: 28458055; PubMed Central PMCID: PMC5660667.

5. Engelman A, Craigie R. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. J Virol. 1992; 66(11):6361–9. Epub 1992/11/01. https://doi.org/10.1128/JVI.66.11.6361-6369.1992 PMID: 1404595; PubMed Central PMCID: PMC240128.

6. Chiu TK, Davies DR. Structure and function of HIV-1 integrase. Curr Top Med Chem. 2004; 4(9):965–77. Epub 2004/05/12. https://doi.org/10.2174/1568026043388547 PMID: 15134551.

7. Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, Hawkins ME, et al. DNA binding induces dissociation of the multimeric form of HIV-1 integrase: a time-resolved fluorescence anisotropy study. Proc Natl Acad Sci U S A. 2001; 98(18):10090–5. Epub 2001/08/16. https://doi.org/10.1073/pnas.181024498 PMID: 11504911; PubMed Central PMCID: PMC56920.

8. Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, Brochon JC. Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. Biochemistry. 2000; 39(31):9275–84. Epub 2000/08/05. https://doi.org/10.1021/bi000397j PMID: 10924120.

9. Jenkins TM, Engelman A, Ghirlando R, Craigie R. A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. J Biol Chem. 1996; 271 (13):7712–8. Epub 1996/03/29. https://doi.org/10.1074/jbc.271.13.7712 PMID: 8631811.

10. Gupta K, Turkki V, Sherrill-Mix S, Hwang Y, Eilers G, Taylor L, et al. Structural Basis for Inhibitor-Induced Aggregation of HIV Integrase. PLoS Biol. 2016; 14(12):e1002584. Epub 2016/12/10. https://doi.org/10.1371/journal.pbio.1002584 PMID: 27935939; PubMed Central PMCID: PMC5147827 contained in this manuscript to improve therapeutics for HIV infection. The other authors declare that they have no conflict of interest.

11. Gupta K, Brady T, Dyer BM, Malani N, Hwang Y, Male F, et al. Allosteric inhibition of human immunodeficiency virus integrase: late block during viral replication and abnormal multimerization involving specific protein domains. J Biol Chem. 2014; 289(30):20477–88. Epub 2014/06/07. https://doi.org/10.1074/jbc.M114.551119 PMID: 24904063; PubMed Central PMCID: PMC4110260.

12. Gupta K, Diamond T, Hwang Y, Bushman F, Van Duyne GD. Structural properties of HIV integrase. Lens epithelium-derived growth factor oligomers. J Biol Chem. 2010; 285(26):20303–15. Epub 2010/

04/22. https://doi.org/10.1074/jbc.M110.114413 PMID: 20406807; PubMed Central PMCID: PMC2888443.

13. Wang JY, Ling H, Yang W, Craigie R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. EMBO J. 2001; 20(24):7333–43. Epub 2001/12/18. https://doi.org/10.1093/emboj/20.24.7333 PMID: 11743009; PubMed Central PMCID: PMC125787.

14. Cai M, Huang Y, Caffrey M, Zheng R, Craigie R, Clore GM, et al. Solution structure of the His12 —> Cys mutant of the N-terminal zinc binding domain of HIV-1 integrase complexed to cadmium. Protein Sci. 1998; 7(12):2669–74. Epub 1998/12/29. https://doi.org/10.1002/pro.5560071221 PMID: 9865962; PubMed Central PMCID: PMC2143878.

15. Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. Proc Natl Acad Sci U S A. 1998; 95 (16):9150–4. Epub 1998/08/05. https://doi.org/10.1073/pnas.95.16.9150 PMID: 9689049; PubMed Central PMCID: PMC21307.

16. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. Nat Struct Biol. 1997; 4(7):567–77. Epub 1997/07/01. https://doi.org/10.1038/nsb0797-567 PMID: 9228950.

17. Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, et al. Solution structure of the DNA binding domain of HIV-1 integrase. Biochemistry. 1995; 34(31):9826–33. Epub 1995/08/08. https://doi.org/10.1021/bi00031a002 PMID: 7632683.

18. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science. 1994; 266 (5193):1981–6. Epub 1994/12/23. https://doi.org/10.1126/science.7801124 PMID: 7801124.

19. Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, Leavitt AD, et al. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. Proc Natl Acad Sci U S A. 2000; 97(15):8233–8. Epub 2000/07/13. https://doi.org/10.1073/pnas.150220297 PMID: 10890912; PubMed Central PMCID: PMC26930.

20. Hare S, Di Nunzio F, Labeja A, Wang J, Engelman A, Cherepanov P. Structural basis for functional tetramerization of lentiviral integrase. PLoS Pathog. 2009; 5(7):e1000515. Epub 2009/07/18. https://doi.org/10.1371/journal.ppat.1000515 PMID: 19609359; PubMed Central PMCID: PMC2705190.

21. Hare S, Shun MC, Gupta SS, Valkov E, Engelman A, Cherepanov P. A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/ LEDGF/p75. PLoS Pathog. 2009; 5(1):e1000259. Epub 2009/01/10. https://doi.org/10.1371/journal.ppat.1000259 PMID: 19132083; PubMed Central PMCID: PMC2606027.

22. Maertens GN, Engelman AN, Cherepanov P. Structure and function of retroviral integrase. Nat Rev Microbiol. 2022; 20(1):20–34. Epub 2021/07/11. https://doi.org/10.1038/s41579-021-00586-9 PMID: 34244677; PubMed Central PMCID: PMC8671357.

23. Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler JA, et al. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. Science. 2000; 287(5453):646–50. Epub 2000/01/29. https://doi.org/10.1126/science.287.5453.646 PMID: 10649997.

24. Mesplede T, Wainberg MA. Integrase Strand Transfer Inhibitors in HIV Therapy. Infect Dis Ther. 2013; 2(2):83–93. Epub 2014/08/20. https://doi.org/10.1007/s40121-013-0020-8 PMID: 25134473; PubMed Central PMCID: PMC4108112.

25. Engelman AN. Multifaceted HIV integrase functionalities and therapeutic strategies for their inhibition. J Biol Chem. 2019; 294(41):15137–57. Epub 2019/08/31. https://doi.org/10.1074/jbc.REV119.006901 PMID: 31467082; PubMed Central PMCID: PMC6791320.

26. You J, Wang H, Huang X, Qin Z, Deng Z, Luo J, et al. Therapy-Emergent Drug Resistance to Integrase Strand Transfer Inhibitors in HIV-1 Patients: A Subgroup Meta-Analysis of Clinical Trials. PLoS One. 2016; 11(8):e0160087. Epub 2016/08/18. https://doi.org/10.1371/journal.pone.0160087 PMID: 27532886; PubMed Central PMCID: PMC4988762.

27. Malet I, Delelis O, Valantin MA, Montes B, Soulie C, Wirden M, et al. Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. Antimicrob Agents Chemother. 2008; 52(4):1351–8. Epub 2008/01/30. https://doi.org/10.1128/AAC.01228-07 PMID: 18227187; PubMed Central PMCID: PMC2292515.

28. Pham HT, Labrie L, Wijting IEA, Hassounah S, Lok KY, Portna I, et al. The S230R Integrase Substitution Associated With Virus Load Rebound During Dolutegravir Monotherapy Confers Low-Level Resistance to Integrase Strand-Transfer Inhibitors. J Infect Dis. 2018; 218(5):698–706. Epub 2018/04/05. https://doi.org/10.1093/infdis/jiy175 PMID: 29617824; PubMed Central PMCID: PMC6240331.

29. Wijting I, Rokx C, Boucher C, van Kampen J, Pas S, de Vries-Sluijs T, et al. Dolutegravir as maintenance monotherapy for HIV (DOMONO): a phase 2, randomised non-inferiority trial. Lancet HIV.

2017; 4(12):e547–e54. Epub 2017/11/07. https://doi.org/10.1016/S2352-3018(17)30152-2 PMID: 29107562.

30. Cooper DA, Steigbigel RT, Gatell JM, Rockstroh JK, Katlama C, Yeni P, et al. Subgroup and resistance analyses of raltegravir for resistant HIV-1 infection. N Engl J Med. 2008; 359(4):355–65. Epub 2008/07/25. https://doi.org/10.1056/NEJMoa0708978 PMID: 18650513.

31. Engelman KD, Engelman AN. Long-Acting Cabotegravir for HIV/AIDS Prophylaxis. Biochemistry. 2021; 60(22):1731–40. Epub 2021/05/25. https://doi.org/10.1021/acs.biochem.1c00157 PMID: 34029457; PubMed Central PMCID: PMC8515172.

32. Mesplede T, Wainberg MA. Resistance against Integrase Strand Transfer Inhibitors and Relevance to HIV Persistence. Viruses. 2015; 7(7):3703–18. Epub 2015/07/23. https://doi.org/10.3390/v7072790 PMID: 26198244; PubMed Central PMCID: PMC4517119.

33. Shimura K, Kodama E, Sakagami Y, Matsuzaki Y, Watanabe W, Yamataka K, et al. Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). J Virol. 2008; 82(2):764–74. Epub 2007/11/06. https://doi.org/10.1128/JVI.01534-07 PMID: 17977962; PubMed Central PMCID: PMC2224569.

34. Molina JM, Lamarca A, Andrade-Villanueva J, Clotet B, Clumeck N, Liu YP, et al. Efficacy and safety of once daily elvitegravir versus twice daily raltegravir in treatment-experienced patients with HIV-1 receiving a ritonavir-boosted protease inhibitor: randomised, double-blind, phase 3, non-inferiority study. Lancet Infect Dis. 2012; 12(1):27–35. Epub 2011/10/22. https://doi.org/10.1016/S1473-3099(11)70249-3 PMID: 22015077.

35. Garrido C, Villacian J, Zahonero N, Pattery T, Garcia F, Gutierrez F, et al. Broad phenotypic cross-resistance to elvitegravir in HIV-infected patients failing on raltegravir-containing regimens. Antimicrob Agents Chemother. 2012; 56(6):2873–8. Epub 2012/03/28. https://doi.org/10.1128/AAC.06170-11 PMID: 22450969; PubMed Central PMCID: PMC3370736.

36. Christ F, Voet A, Marchand A, Nicolet S, Desimmie BA, Marchand D, et al. Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nature chemical biology. 2010; 6(6):442–8. Epub 2010/05/18. https://doi.org/10.1038/nchembio.370 PMID: 20473303.

37. Fader LD, Malenfant E, Parisien M, Carson R, Bilodeau F, Landry S, et al. Discovery of BI 224436, a Noncatalytic Site Integrase Inhibitor (NCINI) of HIV-1. ACS Med Chem Lett. 2014; 5(4):422–7. Epub 2014/06/06. https://doi.org/10.1021/ml500002n PMID: 24900852; PubMed Central PMCID: PMC4027581.

38. Wang H, Jurado KA, Wu X, Shun MC, Li X, Ferris AL, et al. HRP2 determines the efficiency and specificity of HIV-1 integration in LEDGF/p75 knockout cells but does not contribute to the antiviral activity of a potent LEDGF/p75-binding site integrase inhibitor. Nucleic Acids Res. 2012; 40(22):11518–30. Epub 2012/10/09. https://doi.org/10.1093/nar/gks913 PMID: 23042676; PubMed Central PMCID: PMC3526291.

39. Li G, Meanwell NA, Krystal MR, Langley DR, Naidu BN, Sivaprakasam P, et al. Discovery and Optimization of Novel Pyrazolopyrimidines as Potent and Orally Bioavailable Allosteric HIV-1 Integrase Inhibitors. J Med Chem. 2020; 63(5):2620–37. https://doi.org/10.1021/acs.jmedchem.9b01681 WOS:000526403900050. PMID: 32081010

40. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, et al. A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med. 2005; 11(12):1287–9. Epub 2005/11/29. nm1329 [pii] https://doi.org/10.1038/nm1329 PMID: 16311605.

41. Schroder ARW, Shinn P, Chen HM, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. Cell. 2002; 110(4):521–9. ISI:000177653200013. https://doi.org/10.1016/s0092-8674(02)00864-4 PMID: 12202041

42. Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, Kellam P, et al. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes & development. 2007; 21(14):1767–78. Epub 2007/07/20. https://doi.org/10.1101/gad.1565107 PMID: 17639082; PubMed Central PMCID: PMC1920171.

43. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res. 2007; 17(8):1186–94. Epub 2007/06/05. https://doi.org/10.1101/gr.6286907 PMID: 17545577; PubMed Central PMCID: PMC1933515.

44. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH, et al. An essential role for LEDGF/p75 in HIV integration. Science. 2006; 314(5798):461–4. Epub 2006/09/09. https://doi.org/10.1126/science.1132319 PMID: 16959972.

45. Balakrishnan M, Yant SR, Tsai L, O'Sullivan C, Bam RA, Tsai A, et al. Non-catalytic site HIV-1 integrase inhibitors disrupt core maturation and induce a reverse transcription block in target cells. PLoS

One. 2013; 8(9):e74163. https://doi.org/10.1371/journal.pone.0074163 PMID: 24040198; PubMed Central PMCID: PMC3767657.

46. Christ F, Shaw S, Demeulemeester J, Desimmie BA, Marchand A, Butler S, et al. Small-molecule inhibitors of the LEDGF/p75 binding site of integrase block HIV replication and modulate integrase multimerization. Antimicrob Agents Chemother. 2012; 56(8):4365–74. Epub 2012/06/06. https://doi.org/10.1128/AAC.00717-12 PMID: 22664975; PubMed Central PMCID: PMC3421592.

47. Jurado KA, Wang H, Slaughter A, Feng L, Kessl JJ, Koh Y, et al. Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation. Proc Natl Acad Sci U S A. 2013; 110 (21):8690–5. https://doi.org/10.1073/pnas.1300703110 PMID: 23610442; PubMed Central PMCID: PMC3666754.

48. Le Rouzic E, Bonnard D, Chasset S, Bruneau JM, Chevreuil F, Le Strat F, et al. Dual inhibition of HIV-1 replication by integrase-LEDGF allosteric inhibitors is predominant at the post-integration stage. Retrovirology. 2013; 10:144. Epub 2013/11/23. https://doi.org/10.1186/1742-4690-10-144 PMID: 24261564.

49. Sharma A, Slaughter A, Jena N, Feng L, Kessl JJ, Fadel HJ, et al. A new class of multimerization selective inhibitors of HIV-1 integrase. PLoS Pathog. 2014; 10(5):e1004171. Epub 2014/05/31. https://doi.org/10.1371/journal.ppat.1004171 PMID: 24874515; PubMed Central PMCID: PMC4038613.

50. Tsiang M, Jones GS, Niedziela-Majka A, Kan E, Lansdon EB, Huang W, et al. New class of HIV-1 integrase (IN) inhibitors with a dual mode of action. J Biol Chem. 2012; 287(25):21189–203. Epub 2012/04/27. https://doi.org/10.1074/jbc.M112.347534 PMID: 22535962; PubMed Central PMCID: PMC3375541.

51. Fontana J, Jurado KA, Cheng N, Ly NL, Fuchs JR, Gorelick RJ, et al. Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-Inhibited Virions: a Role for Integrase in Maturation. J Virol. 2015; 89(19):9765–80. https://doi.org/10.1128/JVI.01522-15 PMID: 26178982; PubMed Central PMCID: PMC4577894.

52. Gupta K, Allen A, Giraldo C, Eilers G, Sharp R, Hwang Y, et al. Allosteric HIV Integrase Inhibitors Promote Formation of Inactive Branched Polymers via Homomeric Carboxy-Terminal Domain Interactions. Structure. 2021; 29(3):213–25 e5. Epub 2020/12/29. https://doi.org/10.1016/j.str.2020.12.001 PMID: 33357410; PubMed Central PMCID: PMC7935764.

53. Dent EW. Dynamic microtubules at the synapse. Curr Opin Neurobiol. 2020; 63:9–14. Epub 2020/02/18. https://doi.org/10.1016/j.conb.2020.01.003 PMID: 32062144.

54. Kessl JJ, Kutluay SB, Townsend D, Rebensburg S, Slaughter A, Larue RC, et al. HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. Cell. 2016; 166(5):1257–68 e12. Epub 2016/08/28. https://doi.org/10.1016/j.cell.2016.07.044 PMID: 27565348; PubMed Central PMCID: PMC5003418.

55. Madison MK, Lawson DQ, Elliott J, Ozanturk AN, Koneru PC, Townsend D, et al. Allosteric HIV-1 Integrase Inhibitors Lead to Premature Degradation of the Viral RNA Genome and Integrase in Target Cells. J Virol. 2017; 91(17). Epub 2017/06/16. https://doi.org/10.1128/JVI.00821-17 PMID: 28615207; PubMed Central PMCID: PMC5553177.

56. Shkriabai N, Dharmarajan V, Slaughter A, Kessl JJ, Larue RC, Feng L, et al. A critical role of the C-terminal segment for allosteric inhibitor-induced aberrant multimerization of HIV-1 integrase. J Biol Chem. 2014; 289(38):26430–40. Epub 2014/08/15. https://doi.org/10.1074/jbc.M114.589572 PMID: 25118283; PubMed Central PMCID: PMC4176244.

57. Schroder GF, Levitt M, Brunger AT. Deformable elastic network refinement for low-resolution macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2014; 70(Pt 9):2241–55. Epub 2014/09/10. https://doi.org/10.1107/S1399004714016496 PMID: 25195739; PubMed Central PMCID: PMC4157441.

58. Fenwick C, Amad M, Bailey MD, Bethell R, Bos M, Bonneau P, et al. Preclinical profile of BI 224436, a novel HIV-1 non-catalytic-site integrase inhibitor. Antimicrob Agents Chemother. 2014; 58(6):3233–44. Epub 2014/03/26. https://doi.org/10.1128/AAC.02719-13 PMID: 24663024; PubMed Central PMCID: PMC4068430.

59. Diederichs K, Karplus PA. Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol. 1997; 4(4):269–75. Epub 1997/04/01. https://doi.org/10.1038/nsb0497-269 PMID: 9095194.

60. Weiss MS. Global indicators of X-ray data quality. Journal of Applied Crystallography. 2001; 34:130–5.

61. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res. 2018; 46(W1):W363–W7. Epub 2018/06/04. https://doi.org/10.1093/nar/gky473 PMID: 29860391; PubMed Central PMCID: PMC6031066.

62. Eijkelenboom AP, Lutzke RA, Boelens R, Plasterk RH, Kaptein R, Hard K. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. Nat Struct Biol. 1995; 2(9):807–10. Epub 1995/09/01. https://doi.org/10.1038/nsb0995-807 PMID: 7552753.

63. Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. Bioinformatics. 2001; 17(8):700–12. Epub 2001/08/29. https://doi.org/10.1093/bioinformatics/17.8.700 PMID: 11524371.

64. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021; 30(1):70–82. Epub 2020/09/04. https://doi.org/10.1002/pro.3943 PMID: 32881101; PubMed Central PMCID: PMC7737788.

65. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc Natl Acad Sci U S A. 2005; 102 (48):17308–13. Epub 2005/11/02. https://doi.org/10.1073/pnas.0506924102 PMID: 16260736; PubMed Central PMCID: PMC1297672.

66. Hoyte AC, Jamin AV, Koneru PC, Kobe MJ, Larue RC, Fuchs JR, et al. Resistance to pyridine-based inhibitor KF116 reveals an unexpected role of integrase in HIV-1 Gag-Pol polyprotein proteolytic processing. J Biol Chem. 2017; 292(48):19814–25. Epub 20170928. https://doi.org/10.1074/jbc.M117. 816645 PMID: 28972144; PubMed Central PMCID: PMC5712621.

67. Feng L, Sharma A, Slaughter A, Jena N, Koh Y, Shkriabai N, et al. The A128T resistance mutation reveals aberrant protein multimerization as the primary mechanism of action of allosteric HIV-1 integrase inhibitors. J Biol Chem. 2013; 288(22):15813–20. Epub 2013/04/26. https://doi.org/10.1074/jbc. M112.443390 PMID: 23615903; PubMed Central PMCID: PMC3668738.

68. Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. Proc Natl Acad Sci U S A. 1999; 96(17):9459–64. Epub 1999/08/18. https://doi.org/10.1073/pnas.96.17.9459 PMID: 10449714; PubMed Central PMCID: PMC22230.

69. Allouch A, Di Primio C, Alpi E, Lusic M, Arosio D, Giacca M, et al. The TRIM family protein KAP1 inhibits HIV-1 integration. Cell host & microbe. 2011; 9(6):484–95. Epub 2011/06/15. https://doi.org/10. 1016/j.chom.2011.05.004 PMID: 21669397.

70. Allouch A, Cereseto A. Identification of cellular factors binding to acetylated HIV-1 integrase. Amino acids. 2011; 41(5):1137–45. Epub 2009/12/18. https://doi.org/10.1007/s00726-009-0444-3 PMID: 20016921.

71. Cereseto A, Manganaro L, Gutierrez MI, Terreni M, Fittipaldi A, Lusic M, et al. Acetylation of HIV-1 integrase by p300 regulates viral integration. EMBO J. 2005; 24(17):3070–81. Epub 2005/08/13. https://doi.org/10.1038/sj.emboj.7600770 PMID: 16096645; PubMed Central PMCID: PMC1201351.

72. Kiernan RE, Vanhulle C, Schiltz L, Adam E, Xiao H, Maudoux F, et al. HIV-1 tat transcriptional activity is regulated by acetylation. EMBO J. 1999; 18(21):6106–18. Epub 1999/11/02. https://doi.org/10. 1093/emboj/18.21.6106 PMID: 10545121; PubMed Central PMCID: PMC1171675.

73. Lammers M, Neumann H, Chin JW, James LC. Acetylation regulates cyclophilin A catalysis, immunosuppression and HIV isomerization. Nature chemical biology. 2010; 6(5):331–7. Epub 2010/04/07. https://doi.org/10.1038/nchembio.342 PMID: 20364129; PubMed Central PMCID: PMC3867001.

74. Ying H, Zhang Y, Zhou X, Qu X, Wang P, Liu S, et al. Selective histonedeacetylase inhibitor M344 intervenes in HIV-1 latency through increasing histone acetylation and activation of NF-kappaB. Plos One. 2012; 7(11):e48832. Epub 2012/11/21. https://doi.org/10.1371/journal.pone.0048832 PMID: 23166597; PubMed Central PMCID: PMC3499534.

75. De Houwer S, Demeulemeester J, Thys W, Taltynov O, Zmajkovicova K, Christ F, et al. Identification of residues in the C-terminal domain of HIV-1 integrase that mediate binding to the transportin-SR2 protein. J Biol Chem. 2012; 287(41):34059–68. Epub 2012/08/09. https://doi.org/10.1074/jbc.M112. 387944 PMID: 22872638; PubMed Central PMCID: PMC3464515.

76. Koneru PC, Francis AC, Deng N, Rebensburg SV, Hoyte AC, Lindenberger J, et al. HIV-1 integrase tetramers are the antiviral target of pyridine-based allosteric integrase inhibitors. Elife. 2019; 8. Epub 2019/05/24. https://doi.org/10.7554/eLife.46344 PMID: 31120420; PubMed Central PMCID: PMC6581505.

77. Piovesan D, Minervini G, Tosatto SC. The RING 2.0 web server for high quality residue interaction networks. Nucleic Acids Res. 2016; 44(W1):W367–74. Epub 2016/05/21. https://doi.org/10.1093/nar/ gkw315 PMID: 27198219; PubMed Central PMCID: PMC4987896.

78. Patel PA, Kvaratskhelia N, Mansour Y, Antwi J, Feng L, Koneru P, et al. Indole-based allosteric inhibitors of HIV-1 integrase. Bioorg Med Chem Lett. 2016; 26(19):4748–52. Epub 20160813. https://doi. org/10.1016/j.bmcl.2016.08.037 PMID: 27568085; PubMed Central PMCID: PMC5018460.

79. Patel D, Antwi J, Koneru PC, Serrao E, Forli S, Kessl JJ, et al. A New Class of Allosteric HIV-1 Integrase Inhibitors Identified by Crystallographic Fragment Screening of the Catalytic Core Domain. J

Biol Chem. 2016; 291(45):23569–77. Epub 20160919. https://doi.org/10.1074/jbc.M116.753384 PMID: 27645997; PubMed Central PMCID: PMC5095411.

80. Maehigashi T, Ahn S, Kim UI, Lindenberger J, Oo A, Koneru PC, et al. A highly potent and safe pyrro-lopyridine-based allosteric HIV-1 integrase inhibitor targeting host LEDGF/p75-integrase interaction site. PLoS Pathog. 2021; 17(7):e1009671. Epub 2021/07/23. https://doi.org/10.1371/journal.ppat.1009671 PMID: 34293041; PubMed Central PMCID: PMC8297771 following competing interests: S. A., U.K., and K.K. are employees of ST Pharm, B.K. is consultant of ST Pharm, and A.N.E. consults for ViiV Healthcare, Co.. All other authors declare no competing interests.

81. Wilson TA, Koneru PC, Rebensburg SV, Lindenberger JJ, Kobe MJ, Cockroft NT, et al. An Isoquino-line Scaffold as a Novel Class of Allosteric HIV-1 Integrase Inhibitors. ACS Med Chem Lett. 2019; 10 (2):215–20. Epub 20190130. https://doi.org/10.1021/acsmedchemlett.8b00633 PMID: 30783506; PubMed Central PMCID: PMC6378678.

82. Tekeste SS, Wilkinson TA, Weiner EM, Xu X, Miller JT, Le Grice SF, et al. Interaction between Reverse Transcriptase and Integrase Is Required for Reverse Transcription during HIV-1 Replication. J Virol. 2015; 89(23):12058–69. Epub 2015/09/25. https://doi.org/10.1128/JVI.01471-15 PMID: 26401032; PubMed Central PMCID: PMC4645309.

83. Mauro E, Lesbats P, Lapaillerie D, Chaignepain S, Maillot B, Oladosu O, et al. Human H4 tail stimu-lates HIV-1 integration through binding to the carboxy-terminal domain of integrase. Nucleic Acids Res. 2019; 47(7):3607–18. Epub 2019/02/16. https://doi.org/10.1093/nar/gkz091 PMID: 30767014; PubMed Central PMCID: PMC6468294.

84. Sarachan KL, Valentine KG, Gupta K, Moorman VR, Gledhill JM, Bernens M Jr., et al. Solution struc-ture of the core SMN-Gemin2 complex. The Biochemical journal. 2012; 445(3):361–70. Epub 2012/05/23. https://doi.org/10.1042/BJ20120241 PMID: 22607171; PubMed Central PMCID: PMC3462613.

85. Gupta K, Martin R, Sharp R, Sarachan KL, Ninan NS, Van Duyne GD. Oligomeric Properties of Sur-vival Motor Neuron.Gemin2 Complexes. J Biol Chem. 2015; 290(33):20185–99. Epub 2015/06/21. https://doi.org/10.1074/jbc.M115.667279 PMID: 26092730; PubMed Central PMCID: PMC4536428.

86. Martin R, Gupta K, Ninan NS, Perry K, Van Duyne GD. The survival motor neuron protein forms solu-ble glycine zipper oligomers. Structure. 2012; 20(11):1929–39. Epub 2012/10/02. https://doi.org/10.1016/j.str.2012.08.024 PMID: 23022347; PubMed Central PMCID: PMC3519385.

87. Larue R, Gupta K, Wuensch C, Shkriabai N, Kessl JJ, Danhart E, et al. Interaction of the HIV-1 inta-some with transportin 3 protein (TNPO3 or TRN-SR2). J Biol Chem. 2012; 287(41):34044–58. Epub 2012/08/09. https://doi.org/10.1074/jbc.M112.384669 PMID: 22872640; PubMed Central PMCID: PMC3464514.

88. Lesbats P, Engelman AN, Cherepanov P. Retroviral DNA Integration. Chem Rev. 2016; 116 (20):12730–57. Epub 2016/05/21. https://doi.org/10.1021/acs.chemrev.6b00125 PMID: 27198982; PubMed Central PMCID: PMC5084067.

89. Winans S, Goff SP. Mutations altering acetylated residues in the CTD of HIV-1 integrase cause defects in proviral transcription at early times after integration of viral DNA. PLoS Pathog. 2020; 16 (12):e1009147. Epub 2020/12/23. https://doi.org/10.1371/journal.ppat.1009147 PMID: 33351861; PubMed Central PMCID: PMC7787678.

90. Elliott JL, Eschbach JE, Koneru PC, Li W, Puray-Chavez M, Townsend D, et al. Integrase-RNA inter-actions underscore the critical role of integrase in HIV-1 virion morphogenesis. Elife. 2020; 9. Epub 2020/09/23. https://doi.org/10.7554/eLife.54311 PMID: 32960169; PubMed Central PMCID: PMC7671690.

91. Elliott JL, Kutluay SB. Going beyond Integration: The Emerging Role of HIV-1 Integrase in Virion Mor-phogenesis. Viruses. 2020; 12(9). Epub 2020/09/13. https://doi.org/10.3390/v12091005 PMID: 32916894; PubMed Central PMCID: PMC7551943.

92. Lu R, Ghory HZ, Engelman A. Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase. J Virol. 2005; 79(16):10356–68. Epub 2005/07/30. https://doi.org/10.1128/JVI.79.16.10356-10368.2005 PMID: 16051828; PubMed Central PMCID: PMC1182625.

93. Schmitt JD, Sharples CG, Caldwell WS. Molecular recognition in nicotinic acetylcholine receptors: the importance of pi-cation interactions. J Med Chem. 1999; 42(16):3066–74. Epub 1999/08/17. https://doi.org/10.1021/jm990093z PMID: 10447950.

94. Kumar K, Woo SM, Siu T, Cortopassi WA, Duarte F, Paton RS. Cation-pi interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine. Chem Sci. 2018; 9 (10):2655–65. Epub 2018/05/03. https://doi.org/10.1039/c7sc04905f PMID: 29719674; PubMed Cen-tral PMCID: PMC5903419.

95. Pitt WR, Parry DM, Perry BG, Groom CR. Heteroaromatic rings of the future. J Med Chem. 2009; 52 (9):2952–63. Epub 2009/04/08. https://doi.org/10.1021/jm801513z PMID: 19348472.

96. Qi HW, Nakka P, Chen C, Radhakrishnan ML. The effect of macromolecular crowding on the electrostatic component of barnase-barstar binding: a computational, implicit solvent-based study. PLoS One. 2014; 9(6):e98618. Epub 2014/06/11. https://doi.org/10.1371/journal.pone.0098618 PMID: 24915485; PubMed Central PMCID: PMC4051634.

97. Tsiang M, Jones GS, Hung M, Mukund S, Han B, Liu X, et al. Affinities between the binding partners of the HIV-1 integrase dimer-lens epithelium-derived growth factor (IN dimer-LEDGF) complex. J Biol Chem. 2009; 284(48):33580–99. Epub 2009/10/06. https://doi.org/10.1074/jbc.M109.040121 PMID: 19801648; PubMed Central PMCID: PMC2785201.

98. McKee CJ, Kessl JJ, Shkriabai N, Dar MJ, Engelman A, Kvaratskhelia M. Dynamic modulation of HIV-1 integrase structure and function by cellular lens epithelium-derived growth factor (LEDGF) protein. J Biol Chem. 2008; 283(46):31802–12. Epub 2008/09/20. https://doi.org/10.1074/jbc.M805843200 PMID: 18801737; PubMed Central PMCID: PMC2581560.

99. Desimmie BA, Weydert C, Schrijvers R, Vets S, Demeulemeester J, Proost P, et al. HIV-1 IN/Pol recruits LEDGF/p75 into viral particles. Retrovirology. 2015; 12:16. Epub 2015/03/27. https://doi.org/10.1186/s12977-014-0134-4 PMID: 25809198; PubMed Central PMCID: PMC4357141.

100. Kessl JJ, Jena N, Koh Y, Taskent-Sezgin H, Slaughter A, Feng L, et al. Multimode, cooperative mechanism of action of allosteric HIV-1 integrase inhibitors. J Biol Chem. 2012; 287(20):16801–11. Epub 2012/03/23. https://doi.org/10.1074/jbc.M112.354373 PMID: 22437836; PubMed Central PMCID: PMC3351293.

101. Ciuffi A, Diamond TL, Hwang Y, Marshall HM, Bushman FD. Modulating target site selection during human immunodeficiency virus DNA integration in vitro with an engineered tethering factor. Human gene therapy. 2006; 17(9):960–7. Epub 2006/09/16. https://doi.org/10.1089/hum.2006.17.960 PMID: 16972764.

102. Diamond TL, Bushman FD. Role of metal ions in catalysis by HIV integrase analyzed using a quantitative PCR disintegration assay. Nucleic Acids Res. 2006; 34(21):6116–25. Epub 2006/11/07. https://doi.org/10.1093/nar/gkl862 PMID: 17085478; PubMed Central PMCID: PMC1693899.

103. Eilers G, Gupta K, Allen A, Zhou J, Hwang Y, Cory MB, et al. Influence of the amino-terminal sequence on the structure and function of HIV integrase. Retrovirology. 2020; 17(1):28. Epub 2020/09/02. https://doi.org/10.1186/s12977-020-00537-x PMID: 32867805; PubMed Central PMCID: PMC7457537.

104. Vistica J, Dam J, Balbo A, Yikilmaz E, Mariuzza RA, Rouault TA, et al. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal Biochem. 2004; 326(2):234–56. https://doi.org/10.1016/j.ab.2003.12.014 PMID: 15003564.

105. Laue TM, Shah BD, Ridgeway TM, Pelletier SL. Analytical ultracentrifugation in biochemistry and polymer science. In: Harding SE, Rowe AJ, Horton JC, editors. Cambridge [England]:: Royal Society of Chemistry; 1992.

106. Brautigam CA. Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. Methods Enzymol. 2015; 562:109–33. Epub 2015/09/29. https://doi.org/10.1016/bs.mie.2015.05.001 PMID: 26412649.

107. Dar MJ, Monel B, Krishnan L, Shun MC, Di Nunzio F, Helland DE, et al. Biochemical and virological analysis of the 18-residue C-terminal tail of HIV-1 integrase. Retrovirology. 2009; 6:94. Epub 2009/10/21. https://doi.org/10.1186/1742-4690-6-94 PMID: 19840380; PubMed Central PMCID: PMC2770994.

108. Wielens J, Headey SJ, Jeevarajah D, Rhodes DI, Deadman J, Chalmers DK, et al. Crystal structure of the HIV-1 integrase core domain in complex with sucrose reveals details of an allosteric inhibitory binding site. Febs Lett. 2010; 584(8):1455–62. https://doi.org/10.1016/j.febslet.2010.03.016 WOS:000276529200004. PMID: 20227411

109. Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. Journal of Applied Crystallography. 2007; 40:658–74. https://doi.org/10.1107/S0021889807021206 WOS:000248077500003. PMID: 19461840

110. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D. 2011; 67:355–67. https://doi.org/10.1107/S0907444911001314 WOS:000288532800015. PMID: 21460454

111. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D. 2004; 60:2126–32. https://doi.org/10.1107/S0907444904019158 WOS:000225360500002. PMID: 15572765

112. Joosten RP, Long F, Murshudov GN, Perrakis A. The PDB_REDO server for macromolecular structure model optimization. IUCrJ. 2014; 1(Pt 4):213–20. Epub 2014/07/31. https://doi.org/10.1107/S2052252514009324 PMID: 25075342; PubMed Central PMCID: PMC4107921.

113. Fuchs M, Sweet R, Berman L, Bhogadi D, Hendrickson W, Chubar O, et al. NSLS-II MX beamlines FMX for micro-crystallography & AMX for highly automated MX. Acta Crystallogr A. 2014; 70:C1733–C. https://doi.org/10.1107/S2053273314082667 WOS:000519606301736.

114. Winter G, Waterman DG, Parkhurst JM, Brewster AS, Gildea RJ, Gerstel M, et al. DIALS: implementation and evaluation of a new integration package. Acta Crystallogr D. 2018; 74:85–97. https://doi.org/10.1107/S2059798317017235 WOS:000426842000003. PMID: 29533234

115. Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2006; 103(21):8060–5. Epub 2006/05/13. https://doi.org/10.1073/pnas.0602606103 PMID: 16690741; PubMed Central PMCID: PMC1472429.

116. McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr D. 2007; 63:32–41. https://doi.org/10.1107/S0907444906045975 WOS:000242796000005. PMID: 17164524

117. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007; 372(3):774–97. Epub 2007/08/08. https://doi.org/10.1016/j.jmb.2007.05.022 PMID: 17681537.

118. Kuiken C, Korber B, Shafer RW. HIV sequence databases. AIDS Rev. 2003; 5(1):52–61. Epub 2003/07/24. PMID: 12875108; PubMed Central PMCID: PMC2613779.

119. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018; 27(1):135–45. Epub 2017/09/09. https://doi.org/10.1002/pro.3290 PMID: 28884485; PubMed Central PMCID: PMC5734385.

120. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7:539. Epub 2011/10/13. https://doi.org/10.1038/msb.2011.75 PMID: 21988835; PubMed Central PMCID: PMC3261699.

121. Sievers F, Barton GJ, Higgins DG. Multiple sequence alignments. Bioinformatics. 2020; 227:227–50.