Structure-function analyses unravel distinct effects of allosteric inhibitors of HIV-1 integrase on viral maturation and integration

Damien Bonnard, Erwann Le Rouzic, Sylvia Eiler, Céline Amadori, Igor Orlov, Jean-Michel Bruneau, Julie Brias, Julien Barbion, Francis Chevreuil, Danièle Spehner, Sophie Chasset, Benoit Ledoussal, François Moreau, Ali Saïb, Bruno P. Klaholz, Stéphane Emiliani, Marc Ruff, Alessia Zamborlini, and Richard Benarous

Supplemental Data

Table S1. Data collection and refinement statistics. Each data set was collected from one single crystal. Values in parentheses are for the outer shell.

	IN CCD AA (50-212) (PDB entry 5OI3)	MUT-A/IN CCD TT (50-212) (PDB entry 5OI2)	MUT-A/IN CCD AA (50-212) (PDB entry 5OI8)	MUT-A03/IN CCD TT (50-212) (PDB entry 5OI5)	MUT-A03/IN CCD AA (50-212) (PDB entry 5OIA)
Data collection					
Xray wavelength (Å)	1.0000	1.0000	1.5418	1.0000	1.0000
Space group	*$P3_121$*	*$P3_121$*	*$P3_121$*	*$P3_121$*	*$P3_121$*
Unit cell parameters (Å)	a=b=72.27 ; c=65.76 α=β=90 ; γ=120	a=b=72.06 ; c=65.85 α=β=90 ; γ=120	a=b=72.66 ; c=66.17 α=β=90 ; γ=120	a=b=73.06 ; c=65.89 α=β=90 ; γ=120	a=b=72.67 ; c=66.29 α=β=90 ; γ=120
Resolution range (Å)	36.2–1.9 (2.0–1.9)	45.4–2.2 (2.3–2.2)	62.93–2.0 (2.07–2.00)	36.6–2.0 (2.13–2.0)	62.87–1.9 (2.0–1.9)
No. of measured reflections	88,879 (9774)	100,553 (16215)	66,722 (2080)	75,097 (9342)	95,910 (11864)
No. of unique reflections	26,873 (3298)	19,478 (3111)	13,944 (694)	23,646 (3300)	29,016 (4018)
Completeness (%)	88.7 (67.3)	99.6 (97.6)	99.1 (99.9)	89.2 (77.5)	93.9 (80.1)
Mean I/σ (I)	12.37 (3.17)	19.05 (2.88)	28.44 (1.13)	11.58 (1.03)	28.44 (1.13)
Multiplicity	3.3 (2.3)	5.2 (5.2)	4.78 (3.00)	3.2 (2.8)	3.3 (2.9)
R_{merge} (%)	5.6 (24.8)	5.4 (51.3)	5.9 (97.5)	5.6 (137.0)	4.6 (129.4)
Refinement statistics					
Resolution range (Å)	36.18–2.3 (2.36–2.3)	45.4–2.2 (2.6–2.2)	62.93–2.35 (2.41–2.35)	36.6–2.4 (2.46–2.4)	62.87–2.2 (2.26–2.2)
Reflections used	8663 (645)	9882 (726)	8,189 (626)	7673 (596)	9,967 (697)
R factor (%)	19.635 (29.700)	17.65 (30.0)	18.711 (18.5)	20.8 (15.7)	18.326 (42.5)
R_{free} (%)	27.197 (35.800)	21.999 (34.6)	24.367 (26.6)	26.1 (26.5)	22.514 (44.8)
R.m.s. deviation					
Bond lengths (Å)	0.017	0.018	0.017	0.016	0.020
Bond angles (°)	1.682	2.098	2.003	1.838	2.077
Ramachandran plot (%)					
Most favoured regions	99.2		99.2	100.0	99.2
Allowed regions	0.8		0.8	0.0	0.8

Table S2: Summary of structure analysis results.

Structure	Structure analysis results										Biological assay results		
	Number of residues		H-Bonds and salt bridges at the interface		Solvent accessible Area (A^2)		Solvation Energy(kcal/mol)			Ligand Binding free energy (kcal/mol)	LEDGF-IN inhibition IC_{50} (µM)	IN multimerization AC_{50} (µM) *Plateau*	MT4 multiple-round infection EC_{50} (µM)
	Interface	Surface	Number of H-Bonds	Number of salt bridges	Interface	Total	Isolated Structure	Gain on complex formation	P value				
IN CCD TT	42 32.1%	126 96.2%	12	12	1447.4 19.7%	7365.5 100%	-117.5 100%	-8.9 7.5%	0.310				
IN CCD AA	40 30.5%	123 93.9%	12	6	1446.3 19.5%	7307.4 100%	-120.7 100%	-7.8 6.4%	0.413				
IN CCD TT MUT-A	40 30.5%	122 93.1%	10	4	1312.4 18.3%	7172.4 100%	-118.8 100%	-8.6 6.7%	0.268	0.21	0.14	0.12 *532%*	0.047
IN CCD TT MUT-A03	38 29%	126 96.2%	8	10	1366.2 18.4%	7415.1 100%	-116.9 100%	-9.2 7.9%	0.222	0.21	0.19	0.37 *600%*	0.45
IN CCD AA MUT-A	39 29.8%	126 96.2%	10	4	1403.3 18.9%	7417 100%	-121.3 100%	-9.1 7.5%	0.220	-5.97	0.15	0.21 *87%*	2.3
IN CCD AA MUT-A03	39 29.8%	125 95.4%	6	2	1363.8 18.8%	7267.7 100%	-118.3 100%	-9.3 7.8%	0.223	0.21	0.17	0.34 *670%*	0.36

Table S3: Primary HIV-1 isolates used in this study.
These isolates were obtained from the NIH AIDS Research & Reference Reagent Program.

Isolate	Clade	124-125	NIH Reference	Tropism	Accession
KER2008	A	AA	11242	Dual	AF457052
33931N	B	TT	11250	R5	AY713410
NP1538	B	NA	11252	R5	AY713408
NP1525	CRF01-AE	AA	11274	Dual	AY713420

Table S4: Amino acid sequence of integrases from lab strains and primary isolates.
Amino-acid residues 124 and 125 in bold face.

Strain	Amino acid sequence of integrases from lab strains and primary isolates
NL4-3	FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLD CTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTVHTDNGSNFTS**TT**VKA ACWWAGIKQEFGIPYNPQSQGVIESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIG GYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRDPVWKGPAKLLWKGEGAVVIQDNSD IKVVPRRKAKIIRDYGKQMAGDDCVASRQDED
HxB2	FLDGIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLD CTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTIHTDNGSNFTG**AT**VRA ACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIG GYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRNPLWKGPAKLLWKGEGAVVIQDNSD IKVVPRRKAKIIRDYGKQMAGDDCVASRQDED
KER2008	FLDGIDKAQEEHERYHNNWRAMASDFNIPPIIAKEIVASCDKCQLKGEAMHGQVDCSPGMWQLD CTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKVVHTDNGSNFTS**AA**FKA ACWWANVQQEYGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIG GYSAGERIIDIIASDIQTKELQKQITKIQNFRVYYRDSRDPLWKGPAKLLWKGEGAVVIQDNSD IKVVPRRKAKIIRDYGKQMAGDDCVAGRQDED
33931N	FLDGIDKAQEEHEKYHSNWRAMASDFNLPPIVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLD CTHLEGKIILVAVHVASGYIEAEVIPAETGQDTAYFILKLAGRWPVKTIHTDNGSNFTS**TT**VKA ACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIG EYSAGERIIDIIATDIQTKELQKQITKIQNFRVYYRDNRDPLWKGPAKLLWKGEGAVVIQDNSD IKVVPRRKAKIIRDYGKQMAGDDCVASRQNED
NP1538	FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLD CTHLEEKIILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTIHTDNGRNFTS**NA**VKA ACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIG GYSAGERIVDIIATDIQTRELQKQITKIQNFRVYYRDSRDPLWKGPAKLLWKGEGAVVIQDNSD IKVVPRRKAKIIRDYGKQMAGDDCVASRQDED
NP1525	FLDGIDKAQEEHERYHSNWRTMASDFNLPPIVAKEIVANCDKCQLKGEAMHGQVDCSPGIWQLD CTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKVIHTDNGSNFTS**AA**VKA ACWWANVQQEFGIPYNPQSQGVVESMNKELKKIIGQVREQAEHLKTAVQMAVFIHNFKRKGGIG GYSAGERIIDIIATDIQTKELQKQITKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSD IKVVPRRKAKIIRDYGKQMAGDDCVAGRQDED

Figure S1: Multiple alignment of primary HIV-1 isolates used in this study.
Generated with Multalin (http://multalin.toulouse.inra.fr/multalin/) (1)

```
 Symbol comparison table: blosum62             Consensus symbols:
                                                ! is anyone of IV
                                                $ is anyone of LM
                                                % is anyone of FY
MSF:    288    Check:   0                       # is anyone of NDQEBZ
 Name: NL4-3-B        Len:   288  Check: 8559  Weight:  0.85
 Name: HXB2-B         Len:   288  Check: 8075  Weight:  0.85
 Name: 33931N-B       Len:   288  Check: 7049  Weight:  0.99
 Name: NP1538-B       Len:   288  Check: 8037  Weight:  1.07
 Name: KER2008-A      Len:   288  Check: 7556  Weight:  1.11
 Name: NP1525-AE      Len:   288  Check: 7392  Weight:  1.11
 Name: Consensus      Len:   288  Check: 2683  Weight:  0.00
//
                    1                                                      60
          NL4-3-B   FLDGIDKAQE EHEKYHSNWR AMASDFNLPP VVAKEIVASC DKCQLKGEAM HGQVDCSPGI
            HXB2-B  FLDGIDKAQD EHEKYHSNWR AMASDFNLPP VVAKEIVASC DKCQLKGEAM HGQVDCSPGI
           33931N-B FLDGIDKAQE EHEKYHSNWR AMASDFNLPP IVAKEIVASC DKCQLKGEAM HGQVDCSPGI
           NP1538-B FLDGIDKAQE EHEKYHSNWR AMASDFNLPP VVAKEIVASC DKCQLKGEAM HGQVDCSPGI
          KER2008-A FLDGIDKAQE EHERYHNNWR AMASDFNIPP IIAKEIVASC DKCQLKGEAM HGQVDCSPGM
          NP1525-AE FLDGIDKAQE EHERYHSNWR TMASDFNLPP IVAKEIVANC DKCQLKGEAM HGQVDCSPGI
          Consensus FLDGIDKAQ# EHEkYHsNWR aMASDFNlPP !!AKEIVAsC DKCQLKGEAM HGQVDCSPGi

                    61                                                     120
          NL4-3-B   WQLDCTHLEG KVILVAVHVA SGYIEAEVIP AETGQETAYF LLKLAGRWPV KTVHTDNGSN
            HXB2-B  WQLDCTHLEG KVILVAVHVA SGYIEAEVIP AETGQETAYF LLKLAGRWPV KTIHTDNGSN
           33931N-B WQLDCTHLEG KIILVAVHVA SGYIEAEVIP AETGQDTAYF ILKLAGRWPV KTIHTDNGSN
           NP1538-B WQLDCTHLEE KIILVAVHVA SGYIEAEVIP AETGQETAYF LLKLAGRWPV KTIHTDNGRN
          KER2008-A WQLDCTHLEG KVILVAVHVA SGYIEAEVIP AETGQETAYF LLKLAGRWPV KVVHTDNGSN
          NP1525-AE WQLDCTHLEG KVILVAVHVA SGYIEAEVIP AETGQETAYF LLKLAGRWPV KVIHTDNGSN
          Consensus WQLDCTHLEg K!ILVAVHVA SGYIEAEVIP AETGQ#TAYF lLKLAGRWPV Kt!HTDNGsN

                    121                                                    180
          NL4-3-B   FTSTTVKAAC WWAGIKQEFG IPYNPQSQGV IESMNKELKK IIGQVRDQAE HLKTAVQMAV
            HXB2-B  FTGATVRAAC WWAGIKQEFG IPYNPQSQGV VESMNKELKK IIGQVRDQAE HLKTAVQMAV
           33931N-B FTSTTVKAAC WWAGIKQEFG IPYNPQSQGV VESMNKELKK IIGQVRDQAE HLKTAVQMAV
           NP1538-B FTSNAVKAAC WWAGIKQEFG IPYNPQSQGV VESMNKELKK IIGQVRDQAE HLKTAVQMAV
          KER2008-A FTSAAFKAAC WWANVQQEYG IPYNPQSQGV VESMNKELKK IIGQVRDQAE HLKTAVQMAV
          NP1525-AE FTSAAVKAAC WWANVQQEFG IPYNPQSQGV VESMNKELKK IIGQVREQAE HLKTAVQMAV
          Consensus FTsaavkAAC WWAg!kQE%G IPYNPQSQGV !ESMNKELKK IIGQVR#QAE HLKTAVQMAV

                    181                                                    240
          NL4-3-B   FIHNFKRKGG IGGYSAGERI VDIIATDIQT KELQKQITKI QNFRVYYRDS RDPVWKGPAK
            HXB2-B  FIHNFKRKGG IGGYSAGERI VDIIATDIQT KELQKQITKI QNFRVYYRDS RNPLWKGPAK
           33931N-B FIHNFKRKGG IGEYSAGERI IDIIATDIQT KELQKQITKI QNFRVYYRDN RDPLWKGPAK
           NP1538-B FIHNFKRKGG IGGYSAGERI VDIIATDIQT RELQKQITKI QNFRVYYRDS RDPLWKGPAK
          KER2008-A FIHNFKRKGG IGGYSAGERI IDIIASDIQT KELQKQITKI QNFRVYYRDS RDPLWKGPAK
          NP1525-AE FIHNFKRKGG IGGYSAGERI IDIIATDIQT KELQKQITKI QNFRVYYRDS RDPIWKGPAK
          Consensus FIHNFKRKGG IGgYSAGERI !DIIAtDIQT kELQKQITKI QNFRVYYRDs R#PlWKGPAK

                    241                                             288
          NL4-3-B   LLWKGEGAVV IQDNSDIKVV PRRKAKIIRD YGKQMAGDDC VASRQDED
            HXB2-B  LLWKGEGAVV IQDNSDIKVV PRRKAKIIRD YGKQMAGDDC VASRQDED
           33931N-B LLWKGEGAVV IQDNSDIKVV PRRKAKIIRD YGKQMAGDDC VASRQNED
           NP1538-B LLWKGEGAVV IQDNSDIKVV PRRKAKIIRD YGKQMAGDDC VASRQDED
          KER2008-A LLWKGEGAVV IQDNSDIKVV PRRKAKIIRD YGKQMAGDDC VAGRQDED
          NP1525-AE LLWKGEGAVV IQDNSDIKVV PRRKAKIIRD YGKQMAGDDC VAGRQDED
          Consensus LLWKGEGAVV IQDNSDIKVV PRRKAKIIRD YGKQMAGDDC VAsRQ#ED
```

Supplementary Reference

1. Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–90