



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 04:59 PM GMT

PDB ID : 4GVM  
Title : HIV-1 Integrase Catalytic Core Domain A128T Mutant Complexed with Allosteric Inhibitor  
Authors : Feng, L.; Kvaratskhelia, M.  
Deposited on : 2012-08-30  
Resolution : 2.16 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

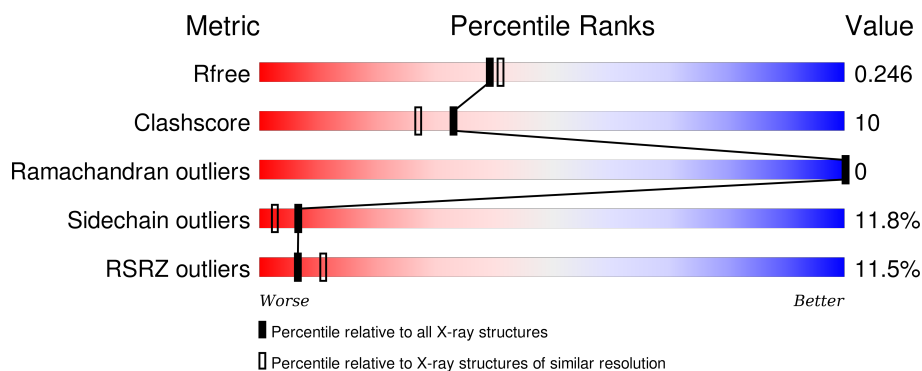
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.16 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	1045 (2.16-2.16)
Clashscore	102246	1152 (2.16-2.16)
Ramachandran outliers	100387	1131 (2.16-2.16)
Sidechain outliers	100360	1131 (2.16-2.16)
RSRZ outliers	91569	1050 (2.16-2.16)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	163	



In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Gag-Pol polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	148	Total	C	N	O	S	0	1	0
			1147	731	198	213	5			

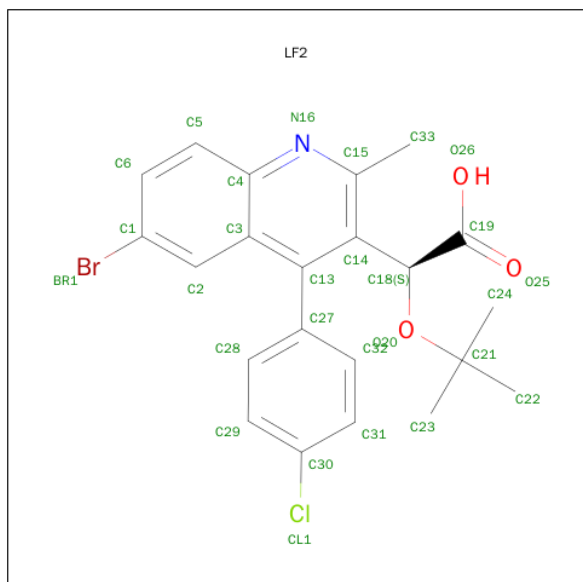
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	128	THR	ALA	ENGINEERED MUTATION	UNP P12497
A	185	LYS	PHE	CONFLICT	UNP P12497

- Molecule 2 is ARSENIC (three-letter code: ARS) (formula: As).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total As 2 2	0	0

- Molecule 3 is (2S)-[6-BROMO-4-(4-CHLOROPHENYL)-2-METHYLQUINOLIN-3-YL](TE RT-BUTOXY)ETHANOIC ACID (three-letter code: LF2) (formula:  $C_{22}H_{21}BrClNO_3$ ).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	Br	C	Cl	N	O	0	0
			28	1	22	1	1	3		

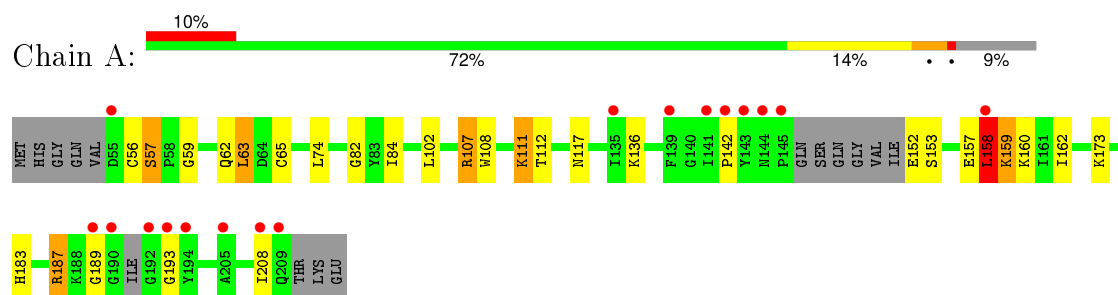
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	42	Total	O	0	0
			42	42		

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $\text{RSRZ} > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

#### • Molecule 1: Gag-Pol polypeptide



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.39 Å 72.39 Å 66.17 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	45.51 – 2.16 45.51 – 2.16	Depositor EDS
% Data completeness (in resolution range)	99.1 (45.51-2.16) 99.1 (45.51-2.16)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.30 (at 2.16 Å)	Xtriage
Refinement program	REFMAC 5.5.0110	Depositor
R, $R_{free}$	0.199 , 0.249 0.202 , 0.246	Depositor DCC
$R_{free}$ test set	1088 reflections (10.99%)	DCC
Wilson B-factor (Å <sup>2</sup> )	46.0	Xtriage
Anisotropy	0.172	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 51.9	EDS
Estimated twinning fraction	0.046 for -h,-k,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 10982 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	1219	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.77% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ARS, LF2

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	1.02	3/1170 (0.3%)	0.85	3/1583 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	158	LEU	C-N	11.70	1.60	1.34
1	A	159	LYS	C-O	9.48	1.41	1.23
1	A	56	CYS	CB-SG	5.58	1.91	1.82

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	63	LEU	CB-CG-CD2	-6.92	99.23	111.00
1	A	158	LEU	O-C-N	-5.73	113.53	122.70
1	A	158	LEU	CA-CB-CG	5.44	127.81	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	158	LEU	Mainchain

## 5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1147	0	1136	24	0
2	A	2	0	0	0	0
3	A	28	0	20	0	0
4	A	42	0	0	2	0
All	All	1219	0	1156	24	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 10.

All (24) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:187:ARG:HH11	1:A:187:ARG:HG3	1.11	1.13
1:A:111[A]:LYS:HE3	4:A:423:HOH:O	1.60	1.00
1:A:107:ARG:HG2	1:A:108:TRP:CD1	2.17	0.78
1:A:187:ARG:HG3	1:A:187:ARG:NH1	1.91	0.76
1:A:111[A]:LYS:HG2	1:A:112:THR:HG23	1.75	0.69
1:A:63:LEU:C	1:A:63:LEU:HD23	2.17	0.65
1:A:59:GLY:HA3	1:A:111[A]:LYS:HD2	1.81	0.61
1:A:65:CYS:SG	1:A:74:LEU:HD13	2.41	0.60
1:A:82:GLY:O	1:A:183:HIS:HE1	1.84	0.60
1:A:152:GLU:OE2	1:A:153:SER:HA	2.03	0.58
1:A:187:ARG:HD2	1:A:187:ARG:O	2.02	0.57
1:A:112:THR:HG22	1:A:136:LYS:HB2	1.88	0.56
1:A:107:ARG:HG2	1:A:108:TRP:NE1	2.23	0.53
1:A:63:LEU:HD21	1:A:74:LEU:HD11	1.94	0.49
1:A:187:ARG:HH11	1:A:187:ARG:CG	2.00	0.48
1:A:142:PRO:HD3	4:A:432:HOH:O	2.14	0.47
1:A:189:GLY:HA2	1:A:193:GLY:HA2	1.97	0.46
1:A:158:LEU:HD13	1:A:158:LEU:C	2.36	0.46
1:A:57:SER:OG	1:A:111[A]:LYS:NZ	2.45	0.45
1:A:158:LEU:HD13	1:A:162:ILE:CD1	2.47	0.45
1:A:65:CYS:SG	1:A:74:LEU:CD1	3.06	0.43
1:A:112:THR:HA	1:A:136:LYS:O	2.18	0.43
1:A:59:GLY:HA3	1:A:111[B]:LYS:HD3	2.00	0.42

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:158:LEU:HD13	1:A:162:ILE:HD12	2.02	0.42

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	143/163 (88%)	141 (99%)	2 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	120/134 (90%)	105 (88%)	15 (12%)	6	2

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	57	SER
1	A	62	GLN
1	A	84	ILE
1	A	102	LEU
1	A	107	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	111[A]	LYS
1	A	111[B]	LYS
1	A	117	ASN
1	A	157	GLU
1	A	158	LEU
1	A	159	LYS
1	A	160	LYS
1	A	173	LYS
1	A	187	ARG
1	A	208	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	114	HIS
1	A	117	ASN
1	A	183	HIS

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected

value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	LF2	A	303	-	26,30,30	1.49	4 (15%)	39,45,45	1.94	13 (33%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	LF2	A	303	-	-	0/13/17/17	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	303	LF2	C14-C18	-4.23	1.47	1.53
3	A	303	LF2	C2-C1	2.44	1.40	1.36
3	A	303	LF2	C13-C14	2.60	1.41	1.38
3	A	303	LF2	C15-N16	3.38	1.35	1.32

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	303	LF2	C3-C4-N16	-3.33	119.65	122.90
3	A	303	LF2	C14-C15-N16	-3.12	119.89	122.54
3	A	303	LF2	C6-C5-C4	-3.01	117.61	120.88
3	A	303	LF2	C33-C15-N16	-2.60	113.33	116.94
3	A	303	LF2	C31-C32-C27	-2.38	117.66	121.14
3	A	303	LF2	C32-C27-C13	-2.24	117.27	120.76
3	A	303	LF2	C13-C14-C18	-2.11	118.76	121.46
3	A	303	LF2	C27-C13-C14	-2.09	119.47	121.58
3	A	303	LF2	C13-C3-C4	2.25	119.82	118.00
3	A	303	LF2	C28-C27-C32	2.53	122.34	117.55
3	A	303	LF2	C33-C15-C14	2.74	126.78	122.33
3	A	303	LF2	C5-C4-C3	3.53	123.06	119.07
3	A	303	LF2	C15-N16-C4	6.06	122.79	118.40

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	148/163 (90%)	0.60	17 (11%) <b>6</b> <b>11</b>	32, 47, 111, 121	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	190	GLY	14.2
1	A	143	TYR	10.1
1	A	144	ASN	8.2
1	A	192	GLY	6.7
1	A	145	PRO	6.2
1	A	55	ASP	5.7
1	A	189	GLY	4.7
1	A	139	PHE	4.4
1	A	158	LEU	3.7
1	A	193	GLY	3.3
1	A	194	TYR	3.1
1	A	208	ILE	2.8
1	A	142	PRO	2.6
1	A	135	ILE	2.6
1	A	209	GLN	2.3
1	A	205	ALA	2.3
1	A	141	ILE	2.2

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	ARS	A	301	1/1	0.98	0.14	1.21	44,44,44,44	0
2	ARS	A	302	1/1	0.97	0.18	-0.09	50,50,50,50	0
3	LF2	A	303	28/28	0.97	0.09	-1.26	32,39,46,58	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.