



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 3, 2016 – 02:35 PM EDT

PDB ID : 5EU7  
Title : Crystal structure of HIV-1 integrase catalytic core in complex with Fab  
Authors : Galilee, M.; Griner, S.L.; Stroud, R.M.; Alian, A.  
Deposited on : 2015-11-18  
Resolution : 2.64 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20028320  
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) : rb-20028320

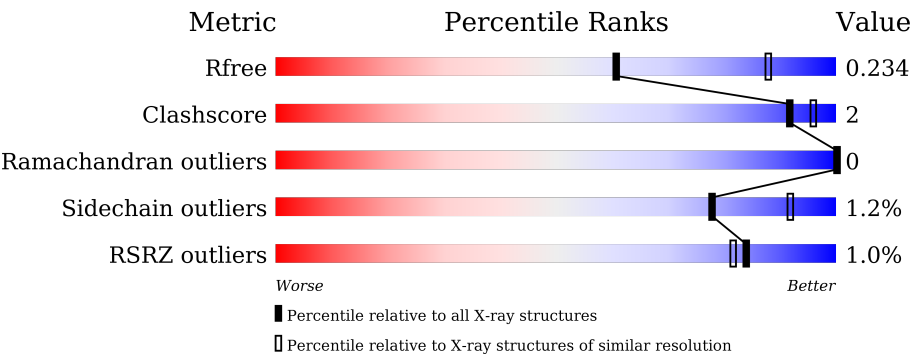


# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:  
*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.64 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	3377 (2.68-2.60)
Clashscore	102246	3781 (2.68-2.60)
Ramachandran outliers	100387	3722 (2.68-2.60)
Sidechain outliers	100360	3722 (2.68-2.60)
RSRZ outliers	91569	3388 (2.68-2.60)

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	153	<div><div>%</div><div>84%9%7%</div></div>
1	B	153	<div><div>%</div><div>87%7%6%</div></div>
2	E	233	<div><div>%</div><div>92%6%</div></div>
2	F	233	<div><div>%</div><div>94%.</div></div>
3	C	214	<div><div>2%</div><div>91%7%</div></div>
3	D	214	<div><div></div><div>93%7%</div></div>



## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 9118 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Integrase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	143	Total	C	N	O	S	0	0	0
			1098	698	191	205	4			
1	B	144	Total	C	N	O	S	0	0	0
			1103	701	192	206	4			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	123	SER	GLY	conflict	UNP P04585
A	124	THR	ALA	conflict	UNP P04585
A	127	LYS	ARG	conflict	UNP P04585
A	131	ASP	TRP	engineered mutation	UNP P04585
A	185	LYS	PHE	engineered mutation	UNP P04585
B	123	SER	GLY	conflict	UNP P04585
B	124	THR	ALA	conflict	UNP P04585
B	127	LYS	ARG	conflict	UNP P04585
B	131	ASP	TRP	engineered mutation	UNP P04585
B	185	LYS	PHE	engineered mutation	UNP P04585

- Molecule 2 is a protein called FAB Heavy Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	228	Total	C	N	O	S	0	0	0
			1723	1092	277	346	8			
2	F	227	Total	C	N	O	S	0	0	0
			1716	1089	276	344	7			

- Molecule 3 is a protein called FAB light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	213	Total	C	N	O	S	0	0	0
			1628	1016	273	334	5			

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	211	Total	C	N	O	S	0	0	0
			1615	1009	271	330	5			

- Molecule 4 is water.

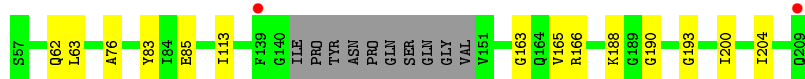
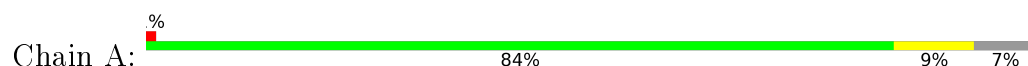
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	21	Total	O	0	0
			21	21		
4	B	24	Total	O	0	0
			24	24		
4	E	59	Total	O	0	0
			59	59		
4	F	36	Total	O	0	0
			36	36		
4	D	53	Total	O	0	0
			53	53		
4	C	42	Total	O	0	0
			42	42		



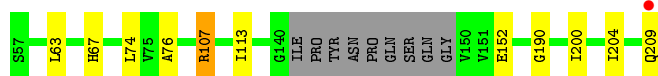
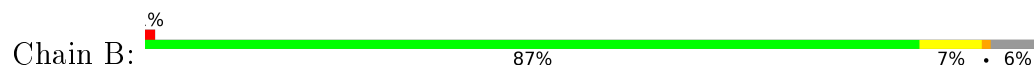
### 3 Residue-property plots [i](#)

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 ( $RSRZ > 2$ ). 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: Integrase



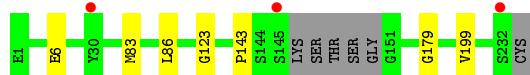
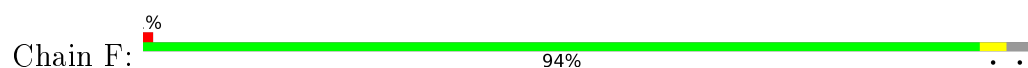
- Molecule 1: Integrase



- Molecule 2: FAB Heavy Chain



- Molecule 2: FAB Heavy Chain



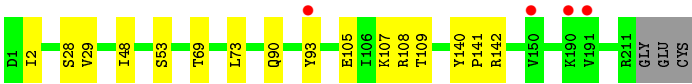
- Molecule 3: FAB light chain



- Molecule 3: FAB light chain









## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	82.49Å 87.65Å 96.34Å 90.00° 95.83° 90.00°	Depositor
Resolution (Å)	95.84 – 2.64 43.34 – 2.64	Depositor EDS
% Data completeness (in resolution range)	98.7 (95.84-2.64) 98.7 (43.34-2.64)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.79 (at 2.65Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, $R_{free}$	0.181 , 0.233 0.186 , 0.234	Depositor DCC
$R_{free}$ test set	2013 reflections (5.30%)	DCC
Wilson B-factor (Å <sup>2</sup> )	40.1	Xtriage
Anisotropy	0.645	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 30.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9118	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.48% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



## 5 Model quality

### 5.1 Standard geometry

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	0.73	0/1117	0.84	2/1508 (0.1%)
1	B	0.80	2/1122 (0.2%)	0.85	2/1515 (0.1%)
2	E	0.78	2/1769 (0.1%)	0.81	1/2410 (0.0%)
2	F	0.73	0/1762	0.78	0/2402
3	C	0.71	0/1649	0.83	2/2239 (0.1%)
3	D	0.73	0/1662	0.83	1/2256 (0.0%)
All	All	0.75	4/9081 (0.0%)	0.82	8/12330 (0.1%)

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
2	F	0	1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	107	ARG	CZ-NH1	5.21	1.39	1.33
2	E	198	VAL	CB-CG1	-5.18	1.42	1.52
2	E	107	SER	CB-OG	-5.13	1.35	1.42
1	B	209	GLN	CD-NE2	-5.08	1.20	1.32

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	142	ARG	NE-CZ-NH2	-8.59	116.01	120.30
3	C	142	ARG	NE-CZ-NH1	-8.42	116.09	120.30
2	E	73	ASP	CB-CG-OD1	6.52	124.17	118.30
3	C	142	ARG	NH1-CZ-NH2	6.07	126.07	119.40
1	A	165	VAL	CB-CA-C	-6.05	99.90	111.40



There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	F	143	PRO	Peptide

## 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	1098	0	1109	8	0
1	B	1103	0	1111	9	0
2	E	1723	0	1641	6	0
2	F	1716	0	1636	3	0
3	C	1615	0	1578	8	0
3	D	1628	0	1587	7	0
4	A	21	0	0	0	0
4	B	24	0	0	1	0
4	C	42	0	0	0	0
4	D	53	0	0	0	0
4	E	59	0	0	1	0
4	F	36	0	0	0	0
All	All	9118	0	8662	36	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 2.

The worst 5 of 36 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:B:152:GLU:HA	4:B:312:HOH:O	1.84	0.77
1:A:85:GLU:OE2	1:B:107:ARG:HD3	1.98	0.62
2:E:210:THR:HG22	4:E:355:HOH:O	2.06	0.54
3:C:48:ILE:HD12	3:C:73:LEU:CD1	2.38	0.53
3:D:2:ILE:CD1	3:D:93:TYR:CD2	2.92	0.53

There are no symmetry-related clashes.



## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	139/153 (91%)	135 (97%)	4 (3%)	0	100	100
1	B	140/153 (92%)	137 (98%)	3 (2%)	0	100	100
2	E	224/233 (96%)	217 (97%)	7 (3%)	0	100	100
2	F	223/233 (96%)	217 (97%)	6 (3%)	0	100	100
3	C	209/214 (98%)	202 (97%)	7 (3%)	0	100	100
3	D	211/214 (99%)	203 (96%)	8 (4%)	0	100	100
All	All	1146/1200 (96%)	1111 (97%)	35 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains

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	115/124 (93%)	114 (99%)	1 (1%)	84	94
1	B	115/124 (93%)	114 (99%)	1 (1%)	84	94
2	E	190/194 (98%)	187 (98%)	3 (2%)	70	88
2	F	189/194 (97%)	189 (100%)	0	100	100
3	C	187/189 (99%)	185 (99%)	2 (1%)	80	92
3	D	188/189 (100%)	183 (97%)	5 (3%)	52	78
All	All	984/1014 (97%)	972 (99%)	12 (1%)	78	92

5 of 12 residues with a non-rotameric sidechain are listed below:



Mol	Chain	Res	Type
3	D	27	GLN
3	D	105	GLU
3	D	213	GLU
2	E	198	VAL
3	D	199	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

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	143/153 (93%)	-0.19	2 (1%) 78 74	24, 45, 79, 102	0
1	B	144/153 (94%)	-0.29	1 (0%) 89 87	24, 41, 75, 92	0
2	E	228/233 (97%)	-0.49	2 (0%) 85 83	24, 37, 63, 120	0
2	F	227/233 (97%)	-0.36	3 (1%) 79 76	24, 39, 63, 98	0
3	C	211/214 (98%)	-0.25	4 (1%) 70 65	24, 43, 71, 88	0
3	D	213/214 (99%)	-0.52	0 100 100	23, 36, 59, 79	0
All	All	1166/1200 (97%)	-0.36	12 (1%) 84 81	23, 39, 71, 120	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	232	SER	3.5
2	F	232	SER	3.3
1	A	139	PHE	3.1
1	A	209	GLN	3.0
3	C	191	VAL	3.0

### 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](#)

There are no ligands in this entry.



## 6.5 Other polymers [i](#)

There are no such residues in this entry.