



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 08:55 pm BST

PDB ID : 2X74  
Title : Human foamy virus integrase - catalytic core.  
Authors : Rety, S.; Delelis, O.; Rezaczkova, L.; Dubanchet, B.; Legrand, P.; Silhan, J.;  
Lewit-Bentley, A.  
Deposited on : 2010-02-23  
Resolution : 2.34 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

<https://www.wwpdb.org/validation/2017/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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

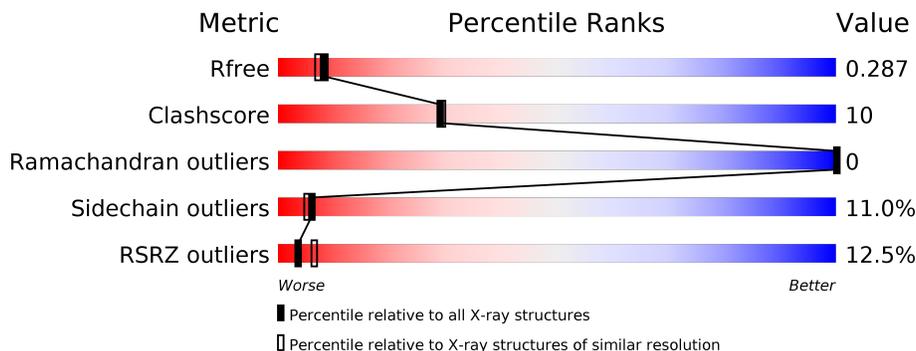
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.34 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))
$R_{free}$	130704	2096 (2.36-2.32)
Clashscore	141614	2193 (2.36-2.32)
Ramachandran outliers	138981	2159 (2.36-2.32)
Sidechain outliers	138945	2160 (2.36-2.32)
RSRZ outliers	127900	2067 (2.36-2.32)

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 respectively. 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	200	 5% 66% 24% • 7%
1	B	200	 7% 66% 23% 5% 7%
1	C	200	 13% 68% 19% • 11%
1	D	200	 18% 66% 20% • • 12%
1	E	200	 13% 72% 19% • 7%
1	F	200	 12% 71% 19% • 10%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 9069 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
			Total	C	N	O	Se			
1	A	187	1514	982	249	279	4	0	2	0
1	B	186	1512	978	250	280	4	0	4	0
1	C	178	1443	933	242	264	4	0	3	0
1	D	177	1429	925	237	263	4	0	2	0
1	E	186	1524	986	253	281	4	0	6	0
1	F	180	1459	946	238	271	4	0	3	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	127	MSE	ILE	engineered mutation	UNP P14350
A	180	ARG	LYS	conflict	UNP P14350
A	227	MSE	ILE	engineered mutation	UNP P14350
A	253	MSE	LEU	engineered mutation	UNP P14350
B	127	MSE	ILE	engineered mutation	UNP P14350
B	180	ARG	LYS	conflict	UNP P14350
B	227	MSE	ILE	engineered mutation	UNP P14350
B	253	MSE	LEU	engineered mutation	UNP P14350
C	127	MSE	ILE	engineered mutation	UNP P14350
C	180	ARG	LYS	conflict	UNP P14350
C	227	MSE	ILE	engineered mutation	UNP P14350
C	253	MSE	LEU	engineered mutation	UNP P14350
D	127	MSE	ILE	engineered mutation	UNP P14350
D	180	ARG	LYS	conflict	UNP P14350
D	227	MSE	ILE	engineered mutation	UNP P14350
D	253	MSE	LEU	engineered mutation	UNP P14350
E	127	MSE	ILE	engineered mutation	UNP P14350

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	180	ARG	LYS	conflict	UNP P14350
E	227	MSE	ILE	engineered mutation	UNP P14350
E	253	MSE	LEU	engineered mutation	UNP P14350
F	127	MSE	ILE	engineered mutation	UNP P14350
F	180	ARG	LYS	conflict	UNP P14350
F	227	MSE	ILE	engineered mutation	UNP P14350
F	253	MSE	LEU	engineered mutation	UNP P14350

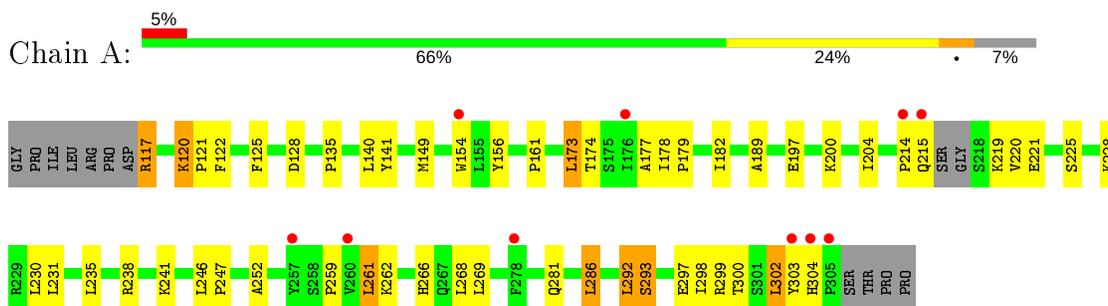
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	55	Total O 55 55	0	0
2	B	54	Total O 54 54	0	0
2	C	19	Total O 19 19	0	0
2	D	6	Total O 6 6	0	0
2	E	35	Total O 35 35	0	0
2	F	19	Total O 19 19	0	0

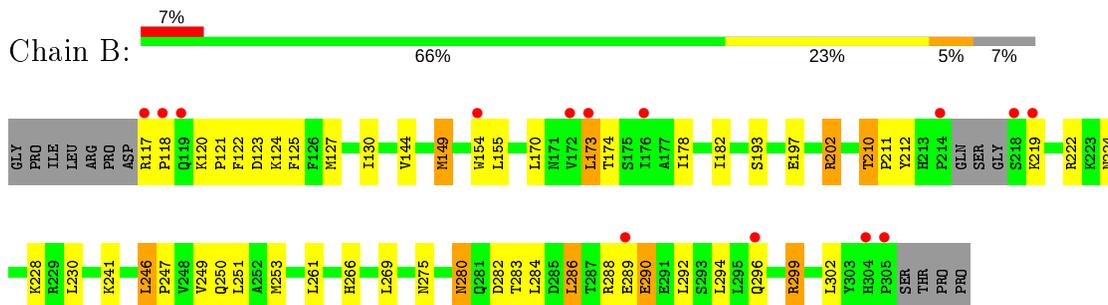
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

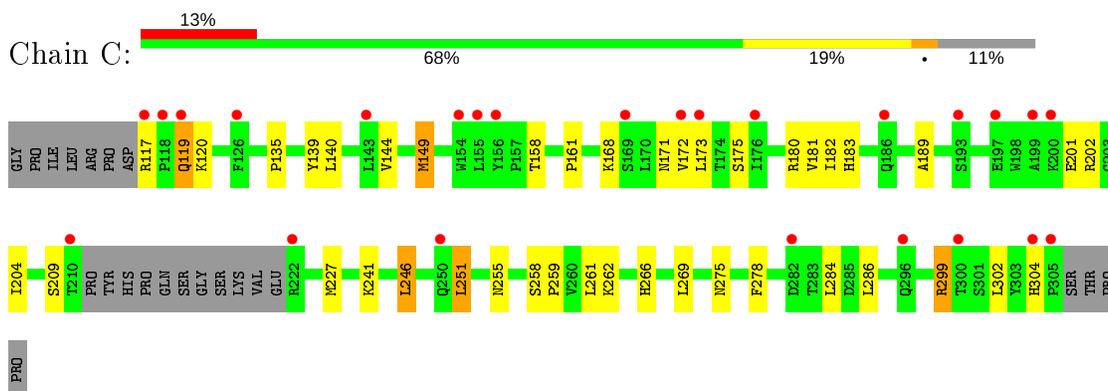
- Molecule 1: INTEGRASE



- Molecule 1: INTEGRASE

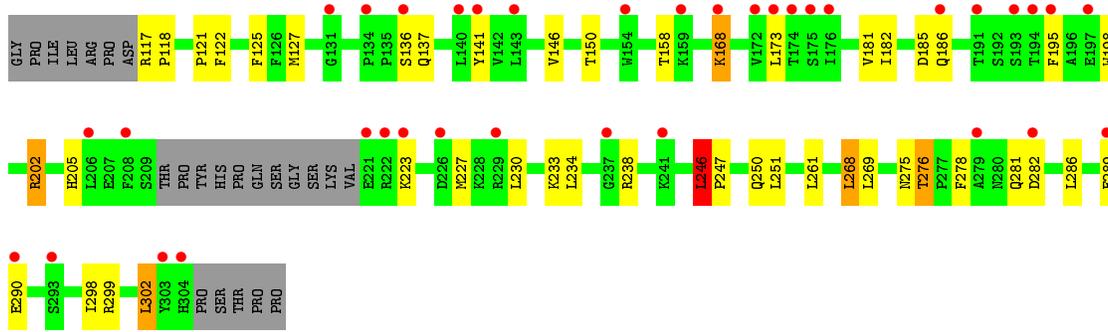


- Molecule 1: INTEGRASE

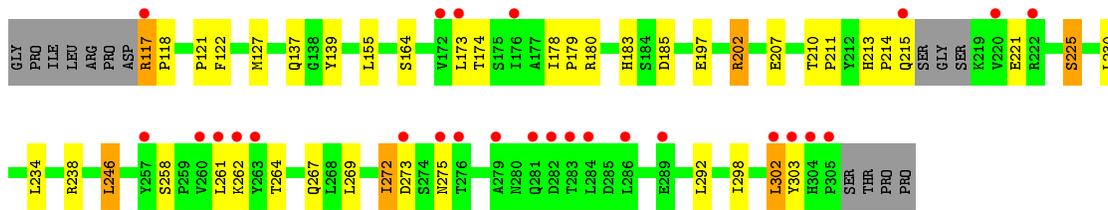


- Molecule 1: INTEGRASE

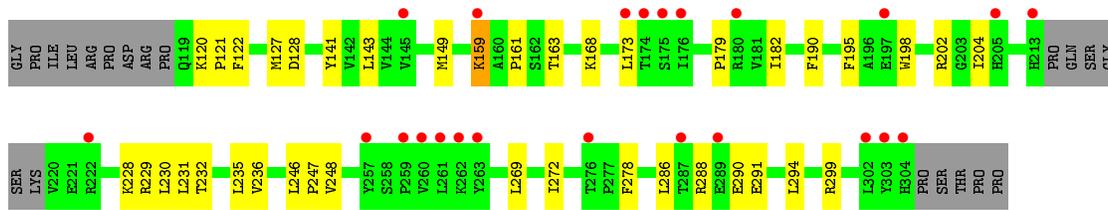




• Molecule 1: INTEGRASE



• Molecule 1: INTEGRASE



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	85.18Å 89.11Å 177.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.58 – 2.34 34.58 – 2.34	Depositor EDS
% Data completeness (in resolution range)	98.9 (34.58-2.34) 98.9 (34.58-2.34)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.93 (at 2.34Å)	Xtrriage
Refinement program	REFMAC 5.5.0088	Depositor
R, $R_{free}$	0.233 , 0.294 0.230 , 0.287	Depositor DCC
$R_{free}$ test set	2843 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	49.6	Xtrriage
Anisotropy	0.119	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 47.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.018 for k,h,-l	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9069	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	57.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.15% 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 [i](#)

### 5.1 Standard geometry [i](#)

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	0.54	0/1564	0.65	0/2126
1	B	0.58	0/1571	0.68	0/2134
1	C	0.45	0/1493	0.61	0/2027
1	D	0.42	0/1474	0.59	1/1999 (0.1%)
1	E	0.51	0/1594	0.60	0/2167
1	F	0.49	0/1510	0.61	0/2050
All	All	0.50	0/9206	0.63	1/12503 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	246	LEU	CA-CB-CG	5.10	127.02	115.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

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	1514	0	1510	36	0
1	B	1512	0	1510	47	0
1	C	1443	0	1452	25	0
1	D	1429	0	1433	23	0
1	E	1524	0	1516	24	0
1	F	1459	0	1456	21	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	55	0	0	2	0
2	B	54	0	0	3	0
2	C	19	0	0	1	0
2	D	6	0	0	2	0
2	E	35	0	0	1	0
2	F	19	0	0	0	0
All	All	9069	0	8877	171	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.

The worst 5 of 171 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:202[A]:ARG:HH11	1:B:202[A]:ARG:HG2	1.18	1.05
1:B:299:ARG:HH11	1:B:299:ARG:HG2	1.23	1.04
1:D:127:MSE:HE2	1:D:182:ILE:HG21	1.42	0.99
1:B:202[A]:ARG:HH11	1:B:202[A]:ARG:CG	1.76	0.98
1:C:299:ARG:HH11	1:C:299:ARG:HG2	1.28	0.96

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	185/200 (92%)	182 (98%)	3 (2%)	0	100	100
1	B	186/200 (93%)	180 (97%)	6 (3%)	0	100	100
1	C	177/200 (88%)	172 (97%)	5 (3%)	0	100	100
1	D	175/200 (88%)	169 (97%)	6 (3%)	0	100	100
1	E	188/200 (94%)	181 (96%)	7 (4%)	0	100	100

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	F	179/200 (90%)	171 (96%)	8 (4%)	0	100	100
All	All	1090/1200 (91%)	1055 (97%)	35 (3%)	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	172/177 (97%)	154 (90%)	18 (10%)	7	6
1	B	173/177 (98%)	153 (88%)	20 (12%)	5	4
1	C	164/177 (93%)	144 (88%)	20 (12%)	5	4
1	D	162/177 (92%)	139 (86%)	23 (14%)	3	3
1	E	174/177 (98%)	154 (88%)	20 (12%)	5	4
1	F	166/177 (94%)	152 (92%)	14 (8%)	11	10
All	All	1011/1062 (95%)	896 (89%)	115 (11%)	6	4

5 of 115 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	275	ASN
1	D	202	ARG
1	F	229	ARG
1	C	278	PHE
1	D	136	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	280	ASN
1	C	224	ASN
1	E	183	HIS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	275	ASN
1	E	213	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](#)

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	183/200 (91%)	0.37	10 (5%) 25 34	24, 42, 73, 86	0
1	B	182/200 (91%)	0.37	14 (7%) 13 20	22, 44, 75, 90	0
1	C	174/200 (87%)	0.84	25 (14%) 2 4	39, 60, 88, 98	0
1	D	173/200 (86%)	1.08	36 (20%) 1 1	47, 79, 100, 100	0
1	E	182/200 (91%)	0.62	26 (14%) 2 4	27, 51, 100, 100	3 (1%)
1	F	176/200 (88%)	0.72	23 (13%) 3 5	35, 57, 90, 100	1 (0%)
All	All	1070/1200 (89%)	0.66	134 (12%) 3 6	22, 54, 94, 100	4 (0%)

The worst 5 of 134 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	220	VAL	8.1
1	C	210	THR	7.9
1	D	304	HIS	6.2
1	A	305	PRO	6.0
1	A	260	VAL	5.0

### 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

There are no such residues in this entry.