



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 27, 2014 – 10:06 PM GMT

PDB ID : 1BIS  
Title : HIV-1 INTEGRASE CORE DOMAIN  
Authors : Goldgur, Y.; Dyda, F.; Hickman, A.B.; Jenkins, T.M.; Craigie, R.; Davies, D.R.  
Deposited on : 1998-06-19  
Resolution : 1.95 Å(reported)

This is a full wwPDB 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/ValidationPDFNotes.html>

---

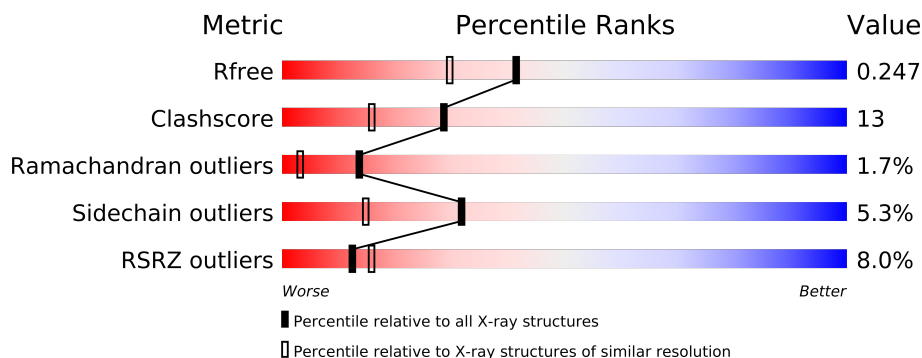
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 1.95 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))
$R_{free}$	66092	1321 (1.96-1.96)
Clashscore	79885	1488 (1.96-1.96)
Ramachandran outliers	78287	1475 (1.96-1.96)
Sidechain outliers	78261	1475 (1.96-1.96)
RSRZ outliers	66119	1321 (1.96-1.96)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	166	
1	B	166	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3274 atoms, of which 820 are hydrogens and 0 are deuterium.

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 HIV-1 INTEGRASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	146	Total	C	H	N	O	S	0	0	0
			1370	713	250	194	208	5			
1	B	154	Total	C	H	N	O	S	0	0	0
			1448	751	266	205	221	5			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	48	SER	GLU	CONFLICT	UNP P12497
A	49	HIS	ALA	CONFLICT	UNP P12497
A	131	GLU	TRP	ENGINEERED	UNP P12497
A	185	LYS	PHE	ENGINEERED	UNP P12497
B	48	SER	GLU	CONFLICT	UNP P12497
B	49	HIS	ALA	CONFLICT	UNP P12497
B	131	GLU	TRP	ENGINEERED	UNP P12497
B	185	LYS	PHE	ENGINEERED	UNP P12497

- Molecule 2 is water.

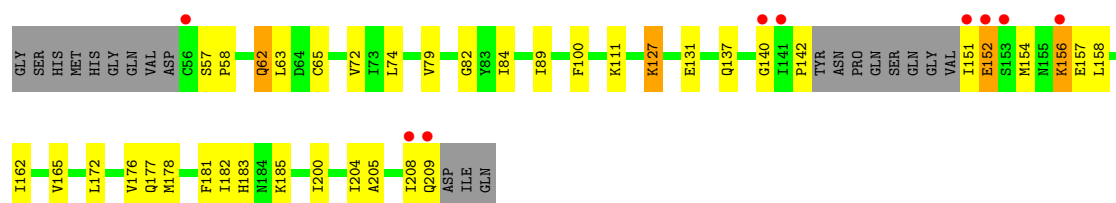
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	80	Total	H	O	0	0
			240	160	80		
2	B	72	Total	H	O	0	0
			216	144	72		

### 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 ( $RSRZ > 2$ ). 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.

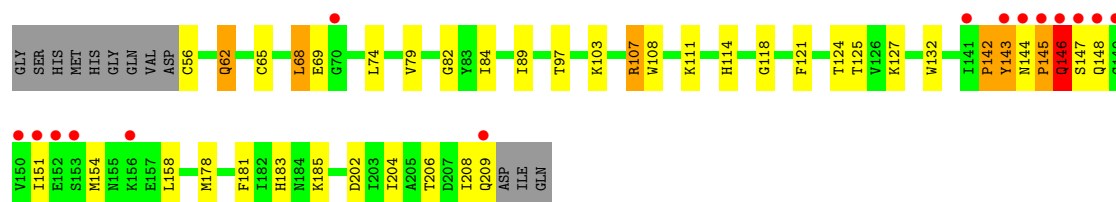
#### • Molecule 1: HIV-1 INTEGRASE

Chain A: 



#### • Molecule 1: HIV-1 INTEGRASE

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	45.09Å 45.08Å 49.43Å 68.82° 64.74° 62.63°	Depositor
Resolution (Å)	(Not available) – 1.95 21.25 – 1.95	Depositor EDS
% Data completeness (in resolution range)	93.8 ((Not available)-1.95) 94.1 (21.25-1.95)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.09 (at 1.95Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, $R_{free}$	0.204 , 0.256 0.196 , 0.247	Depositor DCC
$R_{free}$ test set	1012 reflections (5.09%)	DCC
Wilson B-factor (Å <sup>2</sup> )	24.9	Xtriage
Anisotropy	0.205	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 34.1	EDS
Estimated twinning fraction	0.017 for h,h-k,h-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 20883 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3274	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.19% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 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	0.46	0/1140	0.67	1/1540 (0.1%)
1	B	0.45	0/1205	0.68	0/1631
All	All	0.46	0/2345	0.68	1/3171 (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	A	127	LYS	CD-CE-NZ	5.40	124.11	111.70

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1120	250	884	28	0
1	B	1182	266	924	31	0
2	A	80	160	0	6	0
2	B	72	144	0	9	0
All	All	2454	820	1808	58	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 13.

All (58) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:177:GLN:HG3	2:B:669:HOH:O	1.75	0.85
1:B:124:THR:HA	2:B:738:HOH:O	1.87	0.73
1:B:84:ILE:HD11	1:B:158:LEU:HD13	1.77	0.65
1:B:56:CYS:HG	1:B:143:TYR:HH	1.47	0.63
1:A:127:LYS:HE2	2:A:856:HOH:O	1.99	0.61
1:B:79:VAL:HG22	1:B:151:ILE:HD13	1.84	0.60
1:B:118:GLY:HA3	2:B:931:HOH:O	2.04	0.57
1:B:178:MET:HE1	2:B:1005:HOH:O	2.06	0.55
1:B:144:ASN:N	1:B:145:PRO:HD3	2.23	0.54
1:B:103:LYS:HD3	2:B:669:HOH:O	2.08	0.54
1:B:181:PHE:O	1:B:185:LYS:HG2	2.10	0.52
1:A:204:ILE:O	1:A:208:ILE:HG23	2.09	0.52
1:B:143:TYR:N	1:B:143:TYR:CD1	2.77	0.51
1:B:127:LYS:HD2	2:B:738:HOH:O	2.11	0.50
1:A:84:ILE:HD11	1:A:158:LEU:HD13	1.93	0.50
1:A:82:GLY:O	1:A:183:HIS:HE1	1.94	0.50
1:B:208:ILE:O	1:B:208:ILE:HG22	2.12	0.49
1:A:62:GLN:HG3	1:A:79:VAL:HG22	1.95	0.49
1:B:79:VAL:HG11	1:B:147:SER:HB3	1.95	0.48
1:A:151:ILE:HG23	1:A:152:GLU:H	1.78	0.48
1:A:156:LYS:HD3	1:A:157:GLU:N	2.29	0.48
1:B:79:VAL:CG1	1:B:147:SER:HB3	2.44	0.47
1:B:121:PHE:HE2	2:B:934:HOH:O	1.96	0.47
1:B:107:ARG:HG3	1:B:108:TRP:CD1	2.49	0.47
1:B:82:GLY:O	1:B:183:HIS:HE1	1.99	0.46
1:A:200:ILE:HG12	1:A:204:ILE:HD12	1.97	0.45
1:B:62:GLN:HG2	1:B:114:HIS:HB2	1.99	0.45
1:A:156:LYS:HD3	1:A:157:GLU:H	1.82	0.44
1:B:125:THR:HG22	2:B:878:HOH:O	2.16	0.44
1:B:146:GLN:HA	1:B:146:GLN:NE2	2.33	0.44
1:A:140:GLY:HA2	2:A:483:HOH:O	2.18	0.44
1:B:74:LEU:HD22	1:B:89:ILE:HD13	2.00	0.44
1:B:146:GLN:NE2	1:B:148:GLN:OE1	2.51	0.44
1:A:165:VAL:HG11	1:A:182:ILE:CD1	2.48	0.43
1:B:202:ASP:O	1:B:206:THR:HG23	2.19	0.43
1:A:181:PHE:O	1:A:185:LYS:HG2	2.18	0.43
1:A:72:VAL:HG13	1:A:89:ILE:O	2.19	0.43
1:B:142:PRO:O	1:B:145:PRO:HD3	2.19	0.43
1:B:68:LEU:HD22	1:B:69:GLU:HG3	2.00	0.43
1:A:151:ILE:HG23	1:A:152:GLU:N	2.33	0.42
1:B:74:LEU:HD21	1:B:97:THR:HA	2.01	0.42
1:A:205:ALA:O	1:A:208:ILE:HG12	2.20	0.42

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:151:ILE:O	1:A:154:MET:N	2.52	0.42
1:A:178:MET:HE2	1:B:132:TRP:CE2	2.55	0.42
1:A:142:PRO:CD	2:A:537:HOH:O	2.68	0.42
1:A:142:PRO:HG3	2:A:637:HOH:O	2.20	0.42
1:A:137:GLN:NE2	2:A:741:HOH:H1	2.18	0.41
1:B:147:SER:O	1:B:151:ILE:HB	2.20	0.41
1:A:84:ILE:CD1	1:A:158:LEU:HD13	2.50	0.41
1:B:204:ILE:O	1:B:208:ILE:HG13	2.20	0.41
1:B:56:CYS:SG	1:B:143:TYR:OH	2.64	0.41
1:A:158:LEU:O	1:A:162:ILE:HG13	2.20	0.41
1:A:65:CYS:SG	1:A:74:LEU:HD13	2.61	0.41
1:A:63:LEU:HD11	1:A:100:PHE:CE2	2.56	0.41
1:A:142:PRO:HD3	2:A:537:HOH:O	2.20	0.41
1:B:103:LYS:HE3	2:B:670:HOH:H2	1.86	0.41
1:A:57:SER:HA	1:A:58:PRO:HD3	1.89	0.40
1:A:172:LEU:O	1:A:176:VAL:HG23	2.21	0.40

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	142/166 (86%)	138 (97%)	3 (2%)	1 (1%)	30	15
1	B	152/166 (92%)	147 (97%)	1 (1%)	4 (3%)	8	1
All	All	294/332 (89%)	285 (97%)	4 (1%)	5 (2%)	14	3

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	143	TYR
1	B	145	PRO
1	B	146	GLN
1	A	152	GLU

Continued on next page...



*Continued from previous page...*

Mol	Chain	Res	Type
1	B	142	PRO

### 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	118/135 (87%)	113 (96%)	5 (4%)	40	24
1	B	125/135 (93%)	117 (94%)	8 (6%)	25	10
All	All	243/270 (90%)	230 (95%)	13 (5%)	32	15

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	62	GLN
1	A	111	LYS
1	A	131	GLU
1	A	156	LYS
1	A	209	GLN
1	B	62	GLN
1	B	65	CYS
1	B	68	LEU
1	B	107	ARG
1	B	111	LYS
1	B	146	GLN
1	B	154	MET
1	B	209	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	62	GLN
1	A	67	HIS
1	A	183	HIS
1	B	146	GLN
1	B	183	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

### 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

### 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	146/166 (87%)	0.40	9 (6%)	20 25	13, 25, 54, 84	0
1	B	154/166 (92%)	0.66	15 (9%)	8 10	14, 26, 77, 93	0
All	All	300/332 (90%)	0.53	24 (8%)	12 16	13, 25, 70, 93	0

All (24) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	150	VAL	9.3
1	A	151	ILE	8.1
1	B	144	ASN	7.9
1	B	143	TYR	7.7
1	A	56	CYS	7.6
1	B	148	GLN	5.7
1	B	147	SER	5.4
1	B	146	GLN	5.3
1	B	151	ILE	4.6
1	B	141	ILE	4.5
1	B	149	GLY	4.1
1	B	156	LYS	3.4
1	A	152	GLU	3.4
1	B	153	SER	3.2
1	A	208	ILE	3.0
1	B	145	PRO	3.0
1	A	156	LYS	2.8
1	B	70	GLY	2.6
1	B	152	GLU	2.6
1	B	209	GLN	2.5
1	A	153	SER	2.3
1	A	141	ILE	2.3
1	A	140	GLY	2.2
1	A	209	GLN	2.1

## 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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

## 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 6.4 Ligands ⓘ

There are no ligands in this entry.

## 6.5 Other polymers ⓘ

There are no such residues in this entry.