wwPDB X-ray Structure Validation Summary Report (i)

Oct 23, 2021 – 09:04 AM EDT

PDB ID	:	1BI4
Title	:	CATALYTIC DOMAIN OF HIV-1 INTEGRASE
Authors	:	Maignan, S.; Guilloteau, J.P.; Zhou-Liu, Q.; Clement-Mella, C.; Mikol, V.
Deposited on	:	1998-06-22
Resolution	:	2.50 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

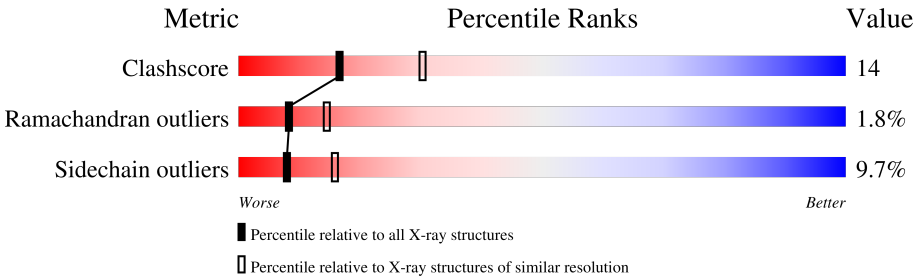
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.50 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))
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)

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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=5%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	160	64% 22% · 11%
1	C	160	65% 33% ·
2	B	160	57% 29% · 11%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3591 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	142	Total	C	N	O	S	25	0	0
			1096	700	191	201	4			
1	C	160	Total	C	N	O	S	55	0	0
			1235	784	216	229	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	185	HIS	PHE	engineered mutation	UNP P12497
C	185	HIS	PHE	engineered mutation	UNP P12497

- Molecule 2 is a protein called INTEGRASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	143	Total	C	N	O	S	64	0	0
			1110	709	193	204	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	150	LEU	VAL	conflict	UNP P12497
B	185	HIS	PHE	engineered mutation	UNP P12497

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	60	Total	O	0	0
			60	60		
3	B	33	Total	O	0	0
			33	33		
3	C	57	Total	O	0	0
			57	57		

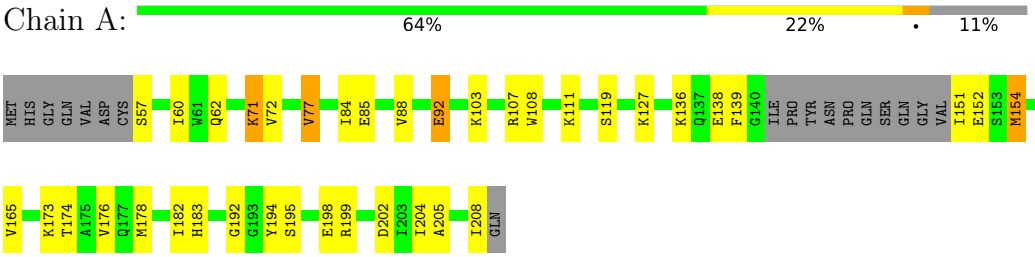
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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. 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. 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.

Note EDS was not executed.

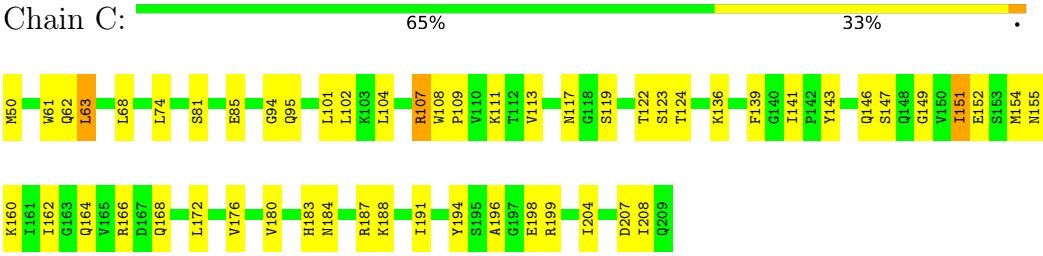
• Molecule 1: INTEGRASE

Chain A:



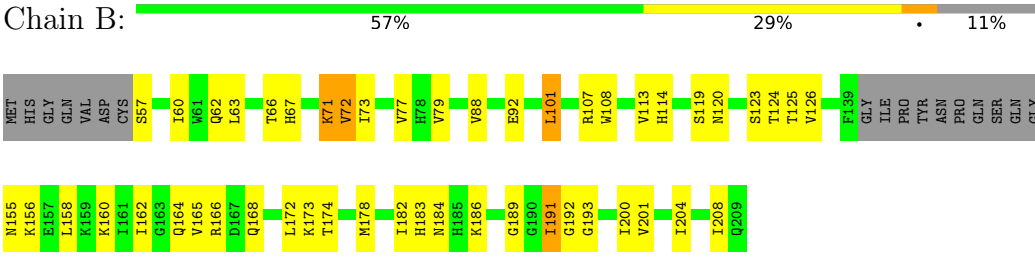
• Molecule 1: INTEGRASE

Chain C:



• Molecule 2: INTEGRASE

Chain B:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	99.10Å 72.50Å 79.40Å 90.00° 108.00° 90.00°	Depositor
Resolution (Å)	15.00 – 2.50	Depositor
% Data completeness (in resolution range)	96.8 (15.00-2.50)	Depositor
R_{merge}	0.04	Depositor
R_{sym}	0.04	Depositor
Refinement program	X-PLOR 3.851	Depositor
R, R_{free}	0.232 , 0.271	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3591	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

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	0.34	0/1118	0.58	0/1512
1	C	0.37	0/1262	0.62	0/1709
2	B	0.32	0/1132	0.55	0/1530
All	All	0.35	0/3512	0.59	0/4751

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	1096	0	1101	28	0
1	C	1235	0	1228	34	0
2	B	1110	0	1117	35	0
3	A	60	0	0	3	0
3	B	33	0	0	1	0
3	C	57	0	0	2	0
All	All	3591	0	3446	93	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 14.

The worst 5 of 93 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)
2:B:77:VAL:HG23	2:B:151:ILE:HD12	1.60	0.83
1:C:151:ILE:O	1:C:151:ILE:HD12	1.80	0.81
1:A:84:ILE:HD11	1:A:154:MET:HG3	1.67	0.77
1:A:72:VAL:HG11	1:A:92:GLU:HG3	1.68	0.74
1:A:62:GLN:HE21	1:A:151:ILE:HB	1.52	0.73

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	138/160 (86%)	134 (97%)	2 (1%)	2 (1%)	11 20
1	C	158/160 (99%)	146 (92%)	11 (7%)	1 (1%)	25 43
2	B	139/160 (87%)	127 (91%)	7 (5%)	5 (4%)	3 4
All	All	435/480 (91%)	407 (94%)	20 (5%)	8 (2%)	8 14

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	192	GLY
2	B	193	GLY
1	C	191	ILE
1	A	92	GLU
2	B	173	LYS

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	114/130 (88%)	104 (91%)	10 (9%)	10 19
1	C	130/130 (100%)	115 (88%)	15 (12%)	5 11
2	B	116/130 (89%)	106 (91%)	10 (9%)	10 20
All	All	360/390 (92%)	325 (90%)	35 (10%)	8 16

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	146	GLN
1	C	151	ILE
1	C	188	LYS
2	B	101	LEU
2	B	72	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	137	GLN
1	C	164	GLN
1	C	185	HIS
2	B	62	GLN
2	B	67	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 monosaccharides 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

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.