wwPDB X-ray Structure Validation Summary Report (i)

Oct 12, 2021 – 03:27 PM EDT

PDB ID	:	1Z19
Title	:	Crystal structure of a lambda integrase(75-356) dimer bound to a COC' core site
Authors	:	Biswas, T.; Aihara, H.; Radman-Livaja, M.; Filman, D.; Landy, A.; Ellenberger, T.
Deposited on	:	2005-03-03
Resolution	:	2.80 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 (1)) were used in the production of this report:

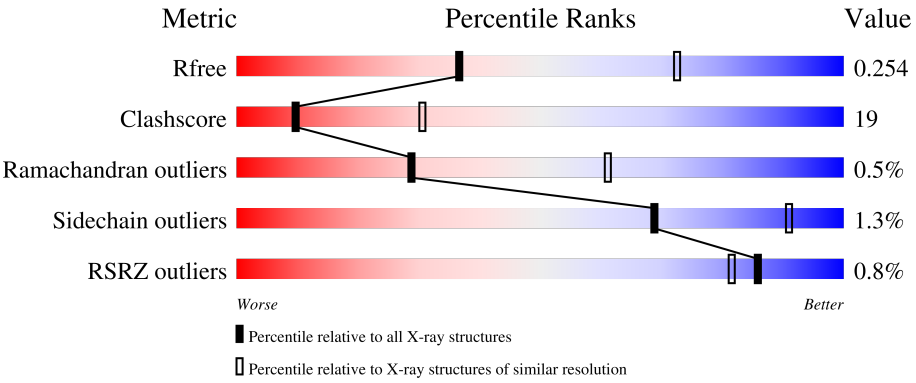
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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% 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	C	16	6% 56% 19% 19%
2	D	20	10% 5% 95%
3	E	33	6% 15% 79% 6%
4	A	283	66% 29%
4	B	283	76% 24%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5677 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 DNA chain called 5'-D(*CP*TP*CP*GP*TP*TP*CP*AP*GP*CP*TP*TP *TP*TP*TP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	13	243	117	36	78	12	0	0	1

- Molecule 2 is a DNA chain called 5'-D(P*TP*TP*TP*AP*TP*AP*CP*TP*AP*AP*GP*T P*TP*GP*GP*CP*AP*TP*TP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	20	394	188	64	122	20	0	0	0

- Molecule 3 is a DNA chain called 33-MER.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	E	33	679	325	134	188	32	0	0	0

- Molecule 4 is a protein called Integrase.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se				
4	A	273	2111	1333	374	393	3	8		0	0	0
			Total	C	N	O	P	S	Se			
4	B	283	2219	1391	397	419	1	3	8	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	74	MSE	VAL	SEE REMARK 999	UNP P03700
A	101	MSE	MET	modified residue	UNP P03700
A	127	MSE	MET	modified residue	UNP P03700
A	174	LYS	GLU	engineered mutation	UNP P03700

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	203	MSE	MET	modified residue	UNP P03700
A	219	MSE	MET	modified residue	UNP P03700
A	255	MSE	MET	modified residue	UNP P03700
A	290	MSE	MET	modified residue	UNP P03700
A	338	MSE	MET	modified residue	UNP P03700
A	342	PTR	TYR	modified residue	UNP P03700
B	74	MSE	VAL	SEE REMARK 999	UNP P03700
B	101	MSE	MET	modified residue	UNP P03700
B	127	MSE	MET	modified residue	UNP P03700
B	174	LYS	GLU	engineered mutation	UNP P03700
B	203	MSE	MET	modified residue	UNP P03700
B	219	MSE	MET	modified residue	UNP P03700
B	255	MSE	MET	modified residue	UNP P03700
B	290	MSE	MET	modified residue	UNP P03700
B	338	MSE	MET	modified residue	UNP P03700
B	342	PTR	TYR	modified residue	UNP P03700

- Molecule 5 is water.

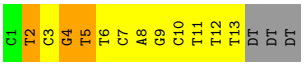
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	4	Total O 4 4	0	0
5	A	12	Total O 12 12	0	0
5	B	15	Total O 15 15	0	0

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 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.

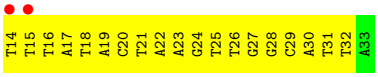
• Molecule 1: 5'-D(*CP*TP*CP*GP*TP*TP*CP*AP*GP*CP*TP*TP*TP*TP*TP*T)-3'

Chain C:



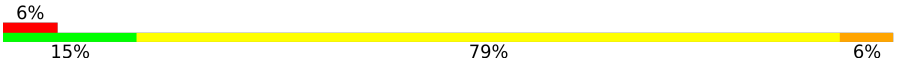
• Molecule 2: 5'-D(P*TP*TP*TP*AP*TP*AP*CP*TP*AP*AP*GP*TP*TP*GP*GP*CP*AP*TP*TP*A)-3'

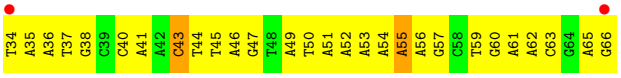
Chain D:



• Molecule 3: 33-MER

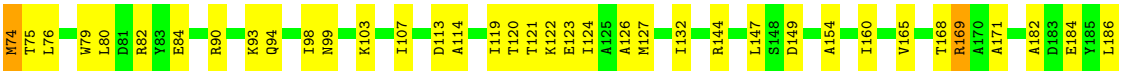
Chain E:

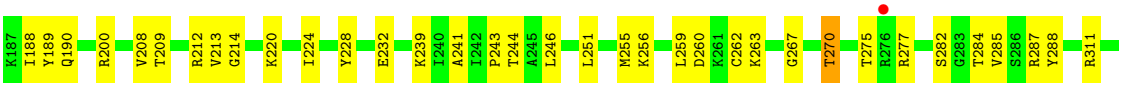


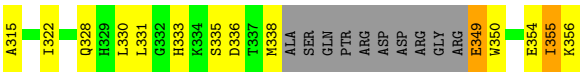


• Molecule 4: Integrase

Chain A:

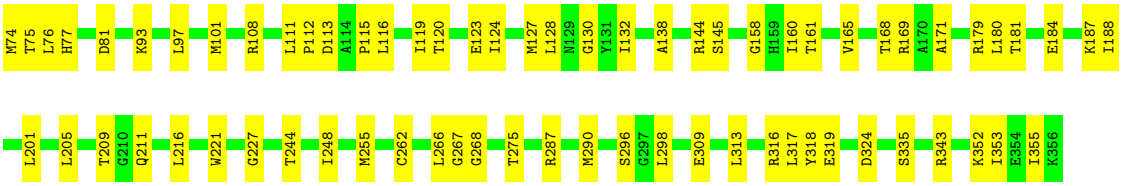






• Molecule 4: Integrase

Chain B:


M74
T75
L76
H77
D81
K93
L97
M101
R108
L111
P112
D113
A114
P115
L116
I119
T120
E123
I124
M127
L128
N129
G130
Y131
I132
A138
R144
S145
G158
H159
I160
T161
V165
T168
R169
A170
A171
R179
L180
T181
E184
K187
I188
L201
L205
T209
G210
Q211
L216
W221
G227
T244
I248
M255
C262
L266
G267
G268
T275
R287
M290
S296
G297
L298
E309
L313
R316
L317
Y318
E319
D324
S335
R343
K352
I353
E354
I355
K356

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	109.30Å 121.48Å 65.39Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80 49.84 – 2.80	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-2.80) 98.7 (49.84-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.47 (at 2.81Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.218 , 0.262 0.205 , 0.254	Depositor DCC
R_{free} test set	2226 reflections (10.07%)	wwPDB-VP
Wilson B-factor (Å^2)	62.3	Xtriage
Anisotropy	0.401	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 39.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	5677	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.91% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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

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	C	0.54	0/269	0.92	1/414 (0.2%)
2	D	0.53	0/439	0.87	0/675
3	E	0.51	0/765	0.85	1/1179 (0.1%)
4	A	0.42	0/2138	0.61	0/2868
4	B	0.42	0/2231	0.61	0/2988
All	All	0.45	0/5842	0.69	2/8124 (0.0%)

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	C	0	3
3	E	0	2
All	All	0	5

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	13	DT	N1-C1'-C2'	5.27	122.61	112.60
3	E	59	DT	N1-C1'-C2'	5.21	122.51	112.60

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	2	DT	Sidechain

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	C	4	DG	Sidechain
1	C	5	DT	Sidechain
3	E	43	DC	Sidechain
3	E	55	DA	Sidechain

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	C	243	0	138	19	0
2	D	394	0	218	28	0
3	E	679	0	372	48	0
4	A	2111	0	2137	67	0
4	B	2219	0	2247	60	0
5	A	12	0	0	0	0
5	B	15	0	0	1	0
5	C	4	0	0	0	0
All	All	5677	0	5112	208	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 19.

The worst 5 of 208 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)
4:A:120:THR:HG22	4:A:122:LYS:H	1.20	1.06
4:A:165:VAL:O	4:A:168:THR:HG22	1.62	0.99
4:B:244:THR:HA	4:B:255:MSE:HE2	1.49	0.94
4:A:120:THR:HG22	4:A:122:LYS:N	1.83	0.93
4:B:120:THR:HG22	4:B:123:GLU:OE1	1.70	0.89

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
4	A	269/283 (95%)	253 (94%)	14 (5%)	2 (1%)	22 53
4	B	280/283 (99%)	263 (94%)	16 (6%)	1 (0%)	34 66
All	All	549/566 (97%)	516 (94%)	30 (6%)	3 (0%)	29 61

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	267	GLY
4	B	267	GLY
4	A	355	ILE

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
4	A	219/228 (96%)	215 (98%)	4 (2%)	59 86
4	B	232/228 (102%)	230 (99%)	2 (1%)	78 94
All	All	451/456 (99%)	445 (99%)	6 (1%)	69 91

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	A	349	GLU
4	B	93	LYS
4	B	145	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	A	169	ARG
4	A	74	MSE

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

Mol	Chain	Res	Type
4	A	159	HIS
4	A	329	HIS
4	A	333	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)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

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	$\#\|Z\| > 2$
4	PTR	B	342	1,4	11,15,17	0.63	0	12,19,24	0.66	0

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
4	PTR	B	342	1,4	-	2/7/10/13	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	342	PTR	N-CA-CB-CG
4	B	342	PTR	C-CA-CB-CG

There are no ring outliers.

No monomer is involved in short contacts.

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

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^{th} 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^2)	Q<0.9
1	C	13/16 (81%)	-0.11	0 100 100	36, 58, 72, 95	0
2	D	20/20 (100%)	0.09	2 (10%) 7 4	34, 64, 146, 150	0
3	E	33/33 (100%)	-0.08	2 (6%) 21 13	34, 51, 84, 94	2 (6%)
4	A	265/283 (93%)	-0.08	1 (0%) 92 91	31, 54, 73, 108	0
4	B	274/283 (96%)	-0.18	0 100 100	33, 49, 71, 89	0
All	All	605/635 (95%)	-0.12	5 (0%) 86 81	31, 52, 75, 150	2 (0%)

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	E	34	DT	3.1
3	E	66	DG	3.0
4	A	276	ARG	2.6
2	D	14	DT	2.4
2	D	15	DT	2.4

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

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. The B-factors column lists the minimum, median, 95^{th} 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	B-factors(Å^2)	Q<0.9
4	PTR	B	342	15/17	0.98	0.19	39,43,48,50	0

6.3 Carbohydrates

There are no monosaccharides 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.