



wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Jan 12, 2017 – 04:30 PM EST

PDB ID : 5U1C
EMDB ID: : EMD-8481
Title : Structure of tetrameric HIV-1 Strand Transfer Complex Intasome
Authors : Lyumkis, D.; Passos, D.
Deposited on : 2016-11-28
Resolution : 3.90 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report
for a publicly released PDB/EMDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Mogul : unknown
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20028442

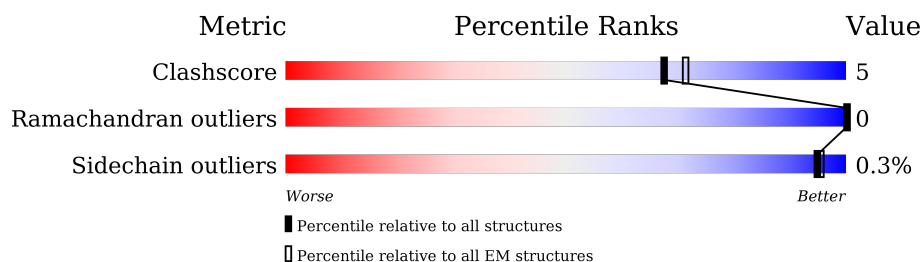
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.90 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)	EM structures (#Entries)
Clashscore	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for ≥ 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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	383	59% 6% • 34%
1	B	383	40% 5% 55%
1	C	383	58% 7% • 34%
1	D	383	40% 5% 55%
2	G	11	64% 9% 27%
2	H	11	64% 9% 27%
3	E	23	43% 17% 39%
3	I	23	48% 13% 39%
4	F	37	57% 8% • 32%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	J	37	<div><div></div><div>57%</div><div>8%</div><div>•</div><div>32%</div></div>

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 8596 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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, Sso7d chimera.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	251	Total	C	N	O	S	0	0
			1975	1260	350	356	9		
1	B	172	Total	C	N	O	S	0	0
			1362	875	243	239	5		
1	C	251	Total	C	N	O	S	0	0
			1975	1260	350	356	9		
1	D	172	Total	C	N	O	S	0	0
			1362	875	243	239	5		

There are 128 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-94	MET	-	expression tag	UNP A0A157T5S7
A	-93	GLY	-	expression tag	UNP A0A157T5S7
A	-92	SER	-	expression tag	UNP A0A157T5S7
A	-91	SER	-	expression tag	UNP A0A157T5S7
A	-90	HIS	-	expression tag	UNP A0A157T5S7
A	-89	HIS	-	expression tag	UNP A0A157T5S7
A	-88	HIS	-	expression tag	UNP A0A157T5S7
A	-87	HIS	-	expression tag	UNP A0A157T5S7
A	-86	HIS	-	expression tag	UNP A0A157T5S7
A	-85	HIS	-	expression tag	UNP A0A157T5S7
A	-84	SER	-	expression tag	UNP A0A157T5S7
A	-83	SER	-	expression tag	UNP A0A157T5S7
A	-82	GLY	-	expression tag	UNP A0A157T5S7
A	-81	LEU	-	expression tag	UNP A0A157T5S7
A	-80	VAL	-	expression tag	UNP A0A157T5S7
A	-79	PRO	-	expression tag	UNP A0A157T5S7
A	-78	ARG	-	expression tag	UNP A0A157T5S7
A	-77	GLY	-	expression tag	UNP A0A157T5S7
A	-76	SER	-	expression tag	UNP A0A157T5S7
A	-75	HIS	-	expression tag	UNP A0A157T5S7
A	-10	GLY	-	linker	UNP A0A157T5S7
A	-9	GLY	-	linker	UNP A0A157T5S7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-8	GLY	-	linker	UNP A0A157T5S7
A	-7	GLY	-	linker	UNP A0A157T5S7
A	-6	GLY	-	linker	UNP A0A157T5S7
A	-5	GLY	-	linker	UNP A0A157T5S7
A	-4	GLY	-	linker	UNP A0A157T5S7
A	-3	GLY	-	linker	UNP A0A157T5S7
A	-2	GLY	-	linker	UNP A0A157T5S7
A	-1	GLY	-	linker	UNP A0A157T5S7
A	0	GLY	-	linker	UNP A0A157T5S7
A	152	GLN	GLU	engineered mutation	UNP F2WR39
B	-94	MET	-	expression tag	UNP A0A157T5S7
B	-93	GLY	-	expression tag	UNP A0A157T5S7
B	-92	SER	-	expression tag	UNP A0A157T5S7
B	-91	SER	-	expression tag	UNP A0A157T5S7
B	-90	HIS	-	expression tag	UNP A0A157T5S7
B	-89	HIS	-	expression tag	UNP A0A157T5S7
B	-88	HIS	-	expression tag	UNP A0A157T5S7
B	-87	HIS	-	expression tag	UNP A0A157T5S7
B	-86	HIS	-	expression tag	UNP A0A157T5S7
B	-85	HIS	-	expression tag	UNP A0A157T5S7
B	-84	SER	-	expression tag	UNP A0A157T5S7
B	-83	SER	-	expression tag	UNP A0A157T5S7
B	-82	GLY	-	expression tag	UNP A0A157T5S7
B	-81	LEU	-	expression tag	UNP A0A157T5S7
B	-80	VAL	-	expression tag	UNP A0A157T5S7
B	-79	PRO	-	expression tag	UNP A0A157T5S7
B	-78	ARG	-	expression tag	UNP A0A157T5S7
B	-77	GLY	-	expression tag	UNP A0A157T5S7
B	-76	SER	-	expression tag	UNP A0A157T5S7
B	-75	HIS	-	expression tag	UNP A0A157T5S7
B	-10	GLY	-	linker	UNP A0A157T5S7
B	-9	GLY	-	linker	UNP A0A157T5S7
B	-8	GLY	-	linker	UNP A0A157T5S7
B	-7	GLY	-	linker	UNP A0A157T5S7
B	-6	GLY	-	linker	UNP A0A157T5S7
B	-5	GLY	-	linker	UNP A0A157T5S7
B	-4	GLY	-	linker	UNP A0A157T5S7
B	-3	GLY	-	linker	UNP A0A157T5S7
B	-2	GLY	-	linker	UNP A0A157T5S7
B	-1	GLY	-	linker	UNP A0A157T5S7
B	0	GLY	-	linker	UNP A0A157T5S7
B	152	GLN	GLU	engineered mutation	UNP F2WR39

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-94	MET	-	expression tag	UNP A0A157T5S7
C	-93	GLY	-	expression tag	UNP A0A157T5S7
C	-92	SER	-	expression tag	UNP A0A157T5S7
C	-91	SER	-	expression tag	UNP A0A157T5S7
C	-90	HIS	-	expression tag	UNP A0A157T5S7
C	-89	HIS	-	expression tag	UNP A0A157T5S7
C	-88	HIS	-	expression tag	UNP A0A157T5S7
C	-87	HIS	-	expression tag	UNP A0A157T5S7
C	-86	HIS	-	expression tag	UNP A0A157T5S7
C	-85	HIS	-	expression tag	UNP A0A157T5S7
C	-84	SER	-	expression tag	UNP A0A157T5S7
C	-83	SER	-	expression tag	UNP A0A157T5S7
C	-82	GLY	-	expression tag	UNP A0A157T5S7
C	-81	LEU	-	expression tag	UNP A0A157T5S7
C	-80	VAL	-	expression tag	UNP A0A157T5S7
C	-79	PRO	-	expression tag	UNP A0A157T5S7
C	-78	ARG	-	expression tag	UNP A0A157T5S7
C	-77	GLY	-	expression tag	UNP A0A157T5S7
C	-76	SER	-	expression tag	UNP A0A157T5S7
C	-75	HIS	-	expression tag	UNP A0A157T5S7
C	-10	GLY	-	linker	UNP A0A157T5S7
C	-9	GLY	-	linker	UNP A0A157T5S7
C	-8	GLY	-	linker	UNP A0A157T5S7
C	-7	GLY	-	linker	UNP A0A157T5S7
C	-6	GLY	-	linker	UNP A0A157T5S7
C	-5	GLY	-	linker	UNP A0A157T5S7
C	-4	GLY	-	linker	UNP A0A157T5S7
C	-3	GLY	-	linker	UNP A0A157T5S7
C	-2	GLY	-	linker	UNP A0A157T5S7
C	-1	GLY	-	linker	UNP A0A157T5S7
C	0	GLY	-	linker	UNP A0A157T5S7
C	152	GLN	GLU	engineered mutation	UNP F2WR39
D	-94	MET	-	expression tag	UNP A0A157T5S7
D	-93	GLY	-	expression tag	UNP A0A157T5S7
D	-92	SER	-	expression tag	UNP A0A157T5S7
D	-91	SER	-	expression tag	UNP A0A157T5S7
D	-90	HIS	-	expression tag	UNP A0A157T5S7
D	-89	HIS	-	expression tag	UNP A0A157T5S7
D	-88	HIS	-	expression tag	UNP A0A157T5S7
D	-87	HIS	-	expression tag	UNP A0A157T5S7
D	-86	HIS	-	expression tag	UNP A0A157T5S7
D	-85	HIS	-	expression tag	UNP A0A157T5S7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-84	SER	-	expression tag	UNP A0A157T5S7
D	-83	SER	-	expression tag	UNP A0A157T5S7
D	-82	GLY	-	expression tag	UNP A0A157T5S7
D	-81	LEU	-	expression tag	UNP A0A157T5S7
D	-80	VAL	-	expression tag	UNP A0A157T5S7
D	-79	PRO	-	expression tag	UNP A0A157T5S7
D	-78	ARG	-	expression tag	UNP A0A157T5S7
D	-77	GLY	-	expression tag	UNP A0A157T5S7
D	-76	SER	-	expression tag	UNP A0A157T5S7
D	-75	HIS	-	expression tag	UNP A0A157T5S7
D	-10	GLY	-	linker	UNP A0A157T5S7
D	-9	GLY	-	linker	UNP A0A157T5S7
D	-8	GLY	-	linker	UNP A0A157T5S7
D	-7	GLY	-	linker	UNP A0A157T5S7
D	-6	GLY	-	linker	UNP A0A157T5S7
D	-5	GLY	-	linker	UNP A0A157T5S7
D	-4	GLY	-	linker	UNP A0A157T5S7
D	-3	GLY	-	linker	UNP A0A157T5S7
D	-2	GLY	-	linker	UNP A0A157T5S7
D	-1	GLY	-	linker	UNP A0A157T5S7
D	0	GLY	-	linker	UNP A0A157T5S7
D	152	GLN	GLU	engineered mutation	UNP F2WR39

- Molecule 2 is a DNA chain called DNA (11-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
2	G	8	Total	C	N	O	P	0	0
			162	77	28	49	8		
2	H	8	Total	C	N	O	P	0	0
			162	77	28	49	8		

- Molecule 3 is a DNA chain called DNA (23-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
3	E	14	Total	C	N	O	P	0	0
			285	138	51	83	13		
3	I	14	Total	C	N	O	P	0	0
			285	138	51	83	13		

- Molecule 4 is a DNA chain called DNA (37-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
4	F	25	Total	C	N	O	P	0	0
			512	244	95	148	25		
4	J	25	Total	C	N	O	P	0	0
			512	244	95	148	25		

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
5	A	1	Total	Zn	0
			1	1	
5	C	1	Total	Zn	0
			1	1	

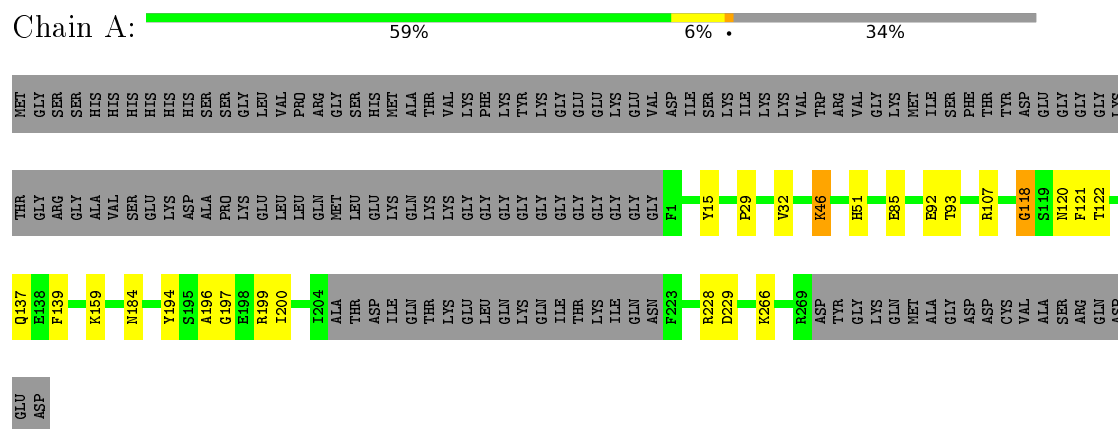
- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
6	A	1	Total	Mg	0
			1	1	
6	C	1	Total	Mg	0
			1	1	

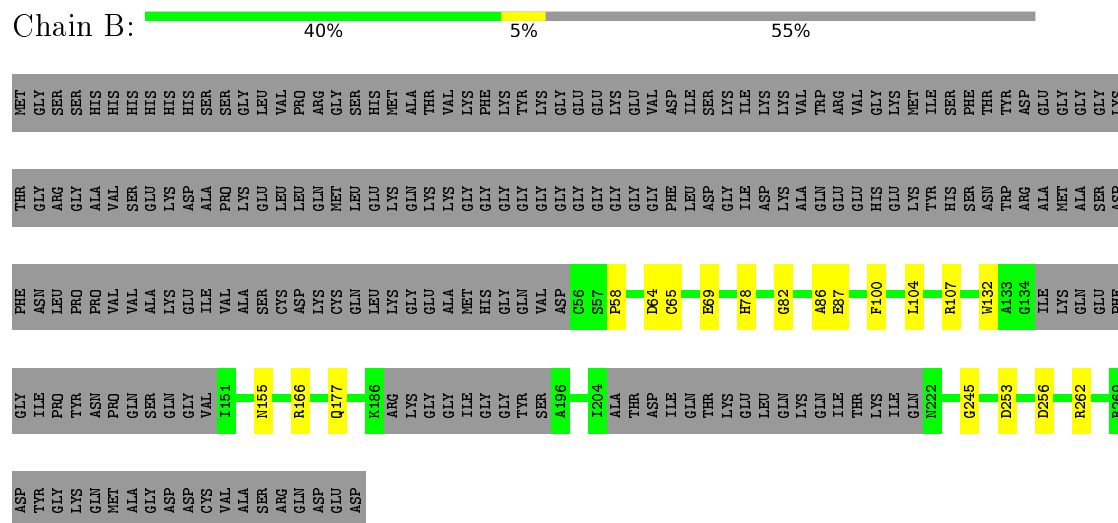
3 Residue-property plots

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. 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.

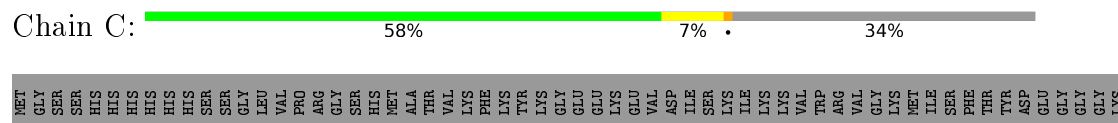
- Molecule 1: HIV-1 Integrase, Sso7d chimera

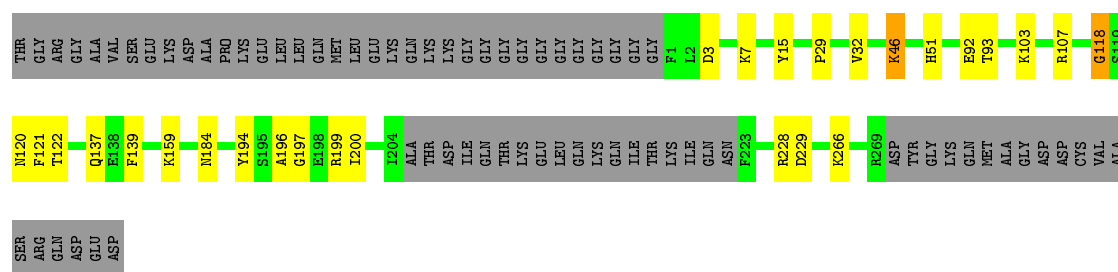


- Molecule 1: HIV-1 Integrase, Sso7d chimera



- Molecule 1: HIV-1 Integrase, Sso7d chimera





- Molecule 1: HIV-1 Integrase, Sso7d chimera

Chain D: 40% 5% 55%



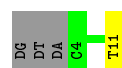
- Molecule 2: DNA (11-MER)

Chain G: 64% 9% 27%



- Molecule 2: DNA (11-MER)

Chain H: 64% 9% 27%



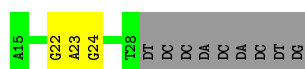
- Molecule 3: DNA (23-MER)

Chain E: 43% 17% 39%

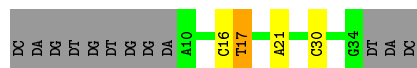


- Molecule 3: DNA (23-MER)

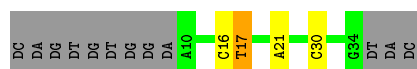
Chain I: 48% 13% 39%



- Molecule 4: DNA (37-MER)



- Molecule 4: DNA (37-MER)



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	83766	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	Not provided	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	Not provided	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	38167	Depositor
Image detector	Not provided	Depositor

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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

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 >2	RMSZ	# Z >2
1	A	0.40	0/2020	0.61	1/2730 (0.0%)
1	B	0.35	0/1390	0.56	0/1879
1	C	0.40	0/2020	0.61	1/2730 (0.0%)
1	D	0.35	0/1390	0.56	0/1879
2	G	0.96	0/180	1.13	0/275
2	H	0.97	0/180	1.13	0/275
3	E	0.82	0/319	1.07	0/491
3	I	0.81	0/319	1.07	0/491
4	F	0.87	0/574	1.01	1/883 (0.1%)
4	J	0.87	0/574	1.01	1/883 (0.1%)
All	All	0.54	0/8966	0.74	4/12516 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	118	GLY	O-C-N	-7.13	111.29	122.70
1	A	118	GLY	O-C-N	-6.92	111.62	122.70
4	J	17	DT	N3-C4-O4	5.18	123.01	119.90
4	F	17	DT	N3-C4-O4	5.17	123.00	119.90

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	1975	0	1981	28	0
1	B	1362	0	1388	12	0
1	C	1975	0	1981	28	0
1	D	1362	0	1388	11	0
2	G	162	0	91	1	0
2	H	162	0	91	1	0
3	E	285	0	161	10	0
3	I	285	0	161	8	0
4	F	512	0	282	6	0
4	J	512	0	282	6	0
5	A	1	0	0	0	0
5	C	1	0	0	0	0
6	A	1	0	0	0	0
6	C	1	0	0	0	0
All	All	8596	0	7806	78	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 5.

The worst 5 of 78 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:C:228:ARG:O	1:C:229:ASP:OD1	1.97	0.83
1:A:228:ARG:O	1:A:229:ASP:OD1	1.96	0.82
1:A:266:LYS:HE2	3:E:24:DG:P	2.24	0.77
1:C:266:LYS:HE2	3:I:24:DG:P	2.26	0.76
1:A:92:GLU:HG2	1:A:120:ASN:HD21	1.51	0.76

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 PDB entries followed by that with respect to all EM entries.

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	247/383 (64%)	228 (92%)	19 (8%)	0	100	100
1	B	164/383 (43%)	157 (96%)	7 (4%)	0	100	100
1	C	247/383 (64%)	228 (92%)	19 (8%)	0	100	100
1	D	164/383 (43%)	157 (96%)	7 (4%)	0	100	100
All	All	822/1532 (54%)	770 (94%)	52 (6%)	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 PDB entries followed by that with respect to all EM entries.

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	208/311 (67%)	207 (100%)	1 (0%)	92	96
1	B	144/311 (46%)	144 (100%)	0	100	100
1	C	208/311 (67%)	207 (100%)	1 (0%)	92	96
1	D	144/311 (46%)	144 (100%)	0	100	100
All	All	704/1244 (57%)	702 (100%)	2 (0%)	95	97

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	46	LYS
1	C	46	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	120	ASN
1	C	120	ASN

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](#)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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.