



wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 07:16 am BST

PDB ID : 5EJK
Title : Crystal structure of the Rous sarcoma virus intasome
Authors : Yin, Z.; Shi, K.; Banerjee, S.; Aihara, H.
Deposited on : 2015-11-02
Resolution : 3.80 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 ⓘ 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

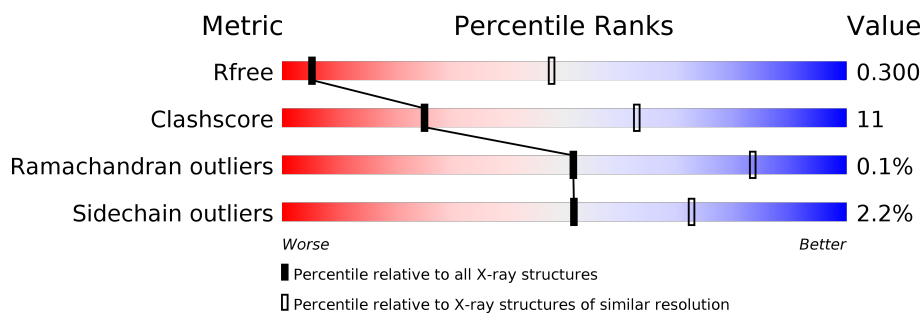
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.80 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	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)

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 ≥ 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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	270	
1	B	270	
1	C	270	
1	D	270	
1	E	270	
1	F	270	
1	G	270	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	270	
2	I	42	
2	L	42	
3	J	22	
3	M	22	
4	K	8	
4	N	8	
5	k	8	
5	n	8	

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 19335 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 Gag-Pro-Pol polyprotein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	269	Total	C	N	O	S	Se	0	0	0
			2099	1316	396	373	3	11			
1	B	253	Total	C	N	O	S	Se	0	0	0
			1993	1252	375	352	3	11			
1	C	256	Total	C	N	O	S	Se	0	0	0
			2009	1262	378	355	3	11			
1	D	250	Total	C	N	O	S	Se	0	0	0
			1967	1235	372	346	3	11			
1	E	269	Total	C	N	O	S	Se	0	0	0
			2099	1316	396	373	3	11			
1	F	254	Total	C	N	O	S	Se	0	0	0
			1997	1254	376	353	3	11			
1	G	250	Total	C	N	O	S	Se	0	0	0
			1972	1241	372	345	3	11			
1	H	263	Total	C	N	O	S	Se	0	0	0
			2058	1290	389	365	3	11			

There are 64 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	23	SER	CYS	engineered mutation	UNP P03354
A	112	MSE	LEU	engineered mutation	UNP P03354
A	135	MSE	LEU	engineered mutation	UNP P03354
A	162	MSE	LEU	engineered mutation	UNP P03354
A	163	MSE	LEU	engineered mutation	UNP P03354
A	166	LYS	ARG	conflict	UNP P03354
A	188	MSE	LEU	engineered mutation	UNP P03354
A	189	MSE	LEU	engineered mutation	UNP P03354
B	23	SER	CYS	engineered mutation	UNP P03354
B	112	MSE	LEU	engineered mutation	UNP P03354
B	135	MSE	LEU	engineered mutation	UNP P03354
B	162	MSE	LEU	engineered mutation	UNP P03354
B	163	MSE	LEU	engineered mutation	UNP P03354

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	166	LYS	ARG	conflict	UNP P03354
B	188	MSE	LEU	engineered mutation	UNP P03354
B	189	MSE	LEU	engineered mutation	UNP P03354
C	23	SER	CYS	engineered mutation	UNP P03354
C	112	MSE	LEU	engineered mutation	UNP P03354
C	135	MSE	LEU	engineered mutation	UNP P03354
C	162	MSE	LEU	engineered mutation	UNP P03354
C	163	MSE	LEU	engineered mutation	UNP P03354
C	166	LYS	ARG	conflict	UNP P03354
C	188	MSE	LEU	engineered mutation	UNP P03354
C	189	MSE	LEU	engineered mutation	UNP P03354
D	23	SER	CYS	engineered mutation	UNP P03354
D	112	MSE	LEU	engineered mutation	UNP P03354
D	135	MSE	LEU	engineered mutation	UNP P03354
D	162	MSE	LEU	engineered mutation	UNP P03354
D	163	MSE	LEU	engineered mutation	UNP P03354
D	166	LYS	ARG	conflict	UNP P03354
D	188	MSE	LEU	engineered mutation	UNP P03354
D	189	MSE	LEU	engineered mutation	UNP P03354
E	23	SER	CYS	engineered mutation	UNP P03354
E	112	MSE	LEU	engineered mutation	UNP P03354
E	135	MSE	LEU	engineered mutation	UNP P03354
E	162	MSE	LEU	engineered mutation	UNP P03354
E	163	MSE	LEU	engineered mutation	UNP P03354
E	166	LYS	ARG	conflict	UNP P03354
E	188	MSE	LEU	engineered mutation	UNP P03354
E	189	MSE	LEU	engineered mutation	UNP P03354
F	23	SER	CYS	engineered mutation	UNP P03354
F	112	MSE	LEU	engineered mutation	UNP P03354
F	135	MSE	LEU	engineered mutation	UNP P03354
F	162	MSE	LEU	engineered mutation	UNP P03354
F	163	MSE	LEU	engineered mutation	UNP P03354
F	166	LYS	ARG	conflict	UNP P03354
F	188	MSE	LEU	engineered mutation	UNP P03354
F	189	MSE	LEU	engineered mutation	UNP P03354
G	23	SER	CYS	engineered mutation	UNP P03354
G	112	MSE	LEU	engineered mutation	UNP P03354
G	135	MSE	LEU	engineered mutation	UNP P03354
G	162	MSE	LEU	engineered mutation	UNP P03354
G	163	MSE	LEU	engineered mutation	UNP P03354
G	166	LYS	ARG	conflict	UNP P03354
G	188	MSE	LEU	engineered mutation	UNP P03354

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	189	MSE	LEU	engineered mutation	UNP P03354
H	23	SER	CYS	engineered mutation	UNP P03354
H	112	MSE	LEU	engineered mutation	UNP P03354
H	135	MSE	LEU	engineered mutation	UNP P03354
H	162	MSE	LEU	engineered mutation	UNP P03354
H	163	MSE	LEU	engineered mutation	UNP P03354
H	166	LYS	ARG	conflict	UNP P03354
H	188	MSE	LEU	engineered mutation	UNP P03354
H	189	MSE	LEU	engineered mutation	UNP P03354

- Molecule 2 is a DNA chain called RSV Integrase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	42	Total	C	N	O	P	0	0	0
			869	413	178	237	41			
2	L	38	Total	C	N	O	P	0	0	0
			787	373	161	215	38			

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*AP*TP*GP*TP*TP*GP*TP*CP*TP*TP*AP*TP*GP*CP*AP*AP*TP*AP*CP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	J	22	Total	C	N	O	P	0	0	0
			445	216	75	133	21			
3	M	18	Total	C	N	O	P	0	0	0
			366	178	62	109	17			

- Molecule 4 is a DNA chain called DNA (5'-D(*CP*TP*TP*CP*TP*CP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	K	8	Total	C	N	O	P	0	0	0
			154	77	19	51	7			
4	N	8	Total	C	N	O	P	0	0	0
			154	77	19	51	7			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	k	8	Total	C	N	O	P	0	0	0
			161	79	26	49	7			
5	n	8	Total	C	N	O	P	0	0	0
			161	79	26	49	7			

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	G	1	Total 1	Zn 1	0	0
6	D	1	Total 1	Zn 1	0	0
6	E	1	Total 1	Zn 1	0	0
6	H	1	Total 1	Zn 1	0	0
6	B	1	Total 1	Zn 1	0	0
6	C	1	Total 1	Zn 1	0	0
6	A	1	Total 1	Zn 1	0	0
6	F	1	Total 1	Zn 1	0	0

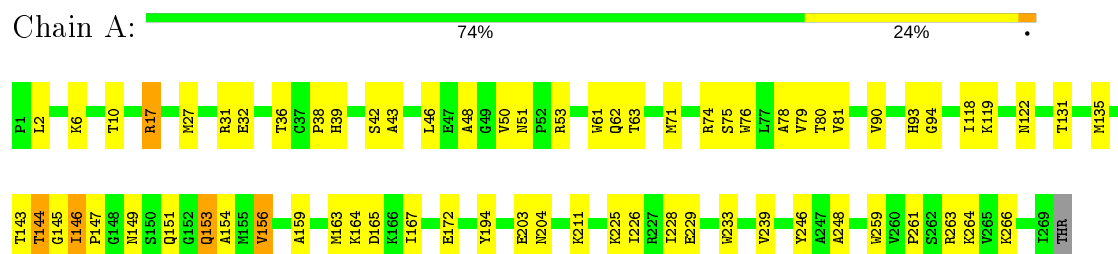
- Molecule 7 is TUNGSTEN ION (three-letter code: W) (formula: W).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	G	1	Total 1	W 1	0	0
7	B	20	Total 20	W 20	0	0
7	A	4	Total 4	W 4	1	0
7	F	11	Total 11	W 11	0	0

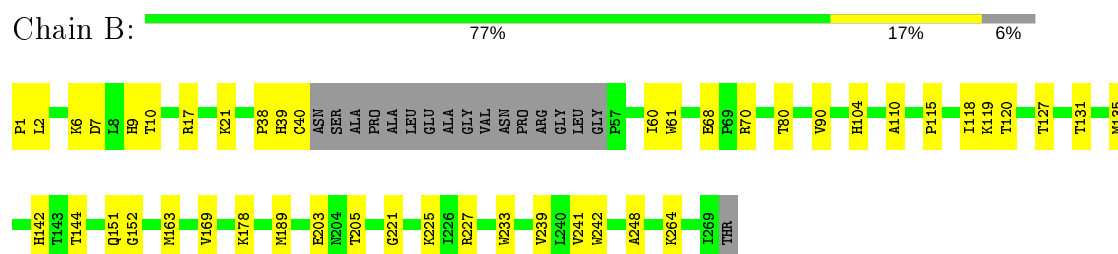
3 Residue-property plots

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

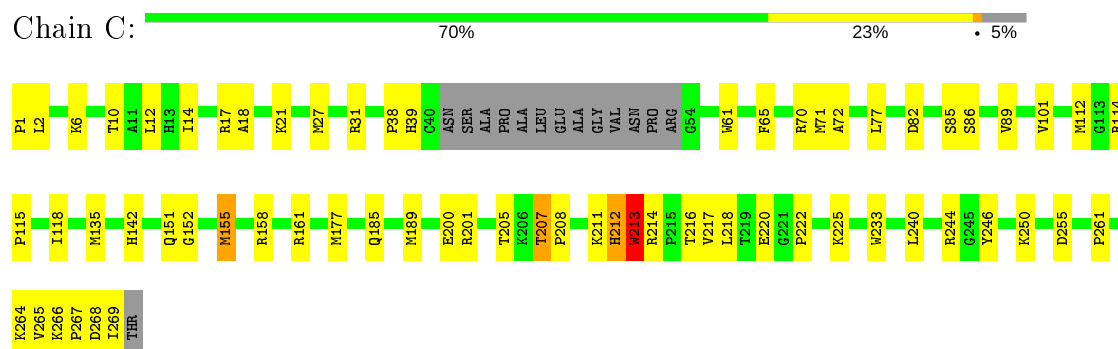
• Molecule 1: Gag-Pro-Pol polyprotein



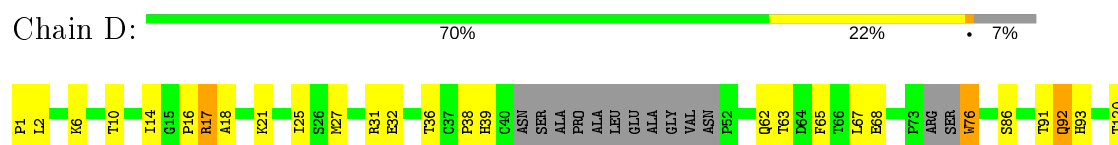
• Molecule 1: Gag-Pro-Pol polyprotein

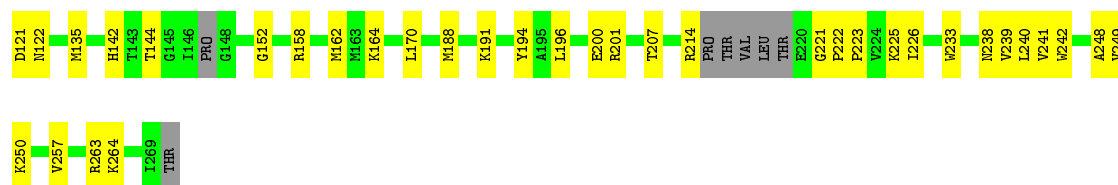


• Molecule 1: Gag-Pro-Pol polyprotein



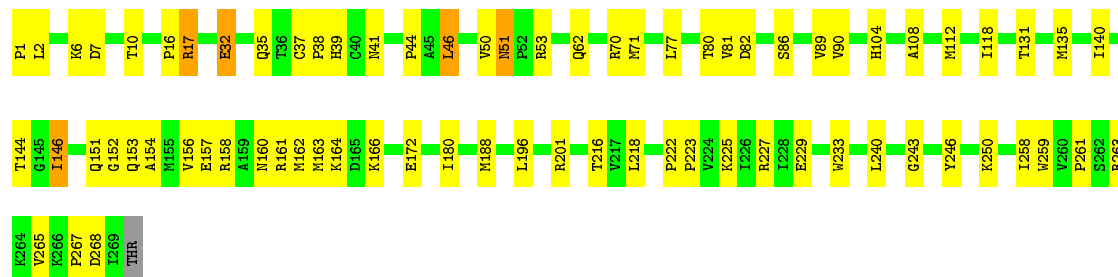
• Molecule 1: Gag-Pro-Pol polyprotein





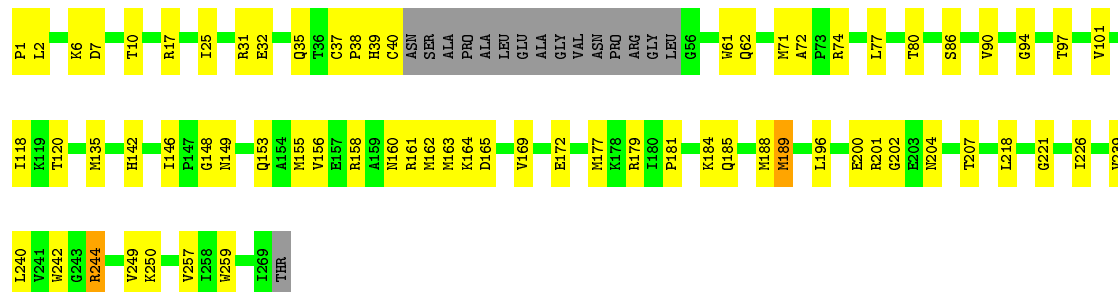
• Molecule 1: Gag-Pro-Pol polyprotein

Chain E: 72% 26%



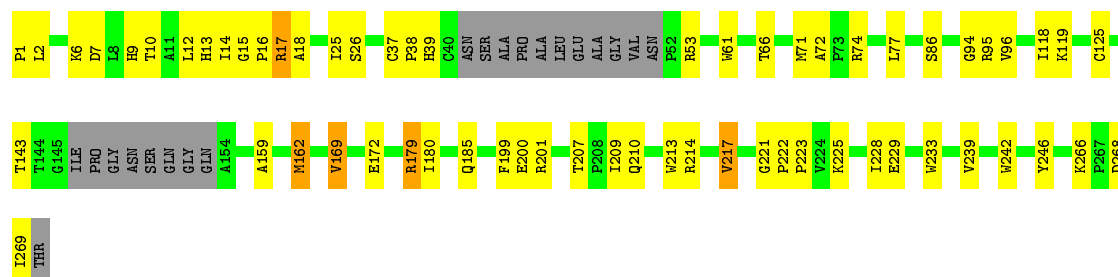
• Molecule 1: Gag-Pro-Pol polyprotein

Chain F: 69% 25% 6%



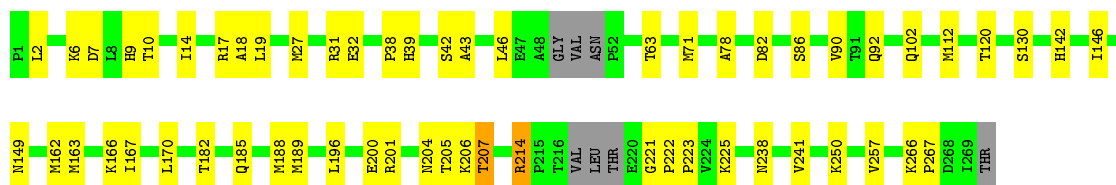
• Molecule 1: Gag-Pro-Pol polyprotein

Chain G: 70% 21% 7%

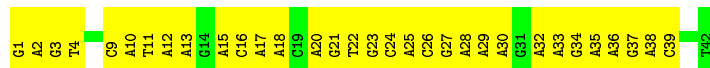


• Molecule 1: Gag-Pro-Pol polyprotein

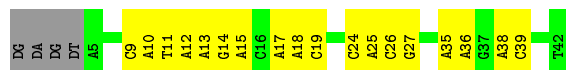
Chain H: 76% 21%



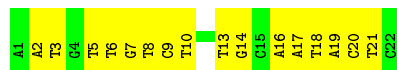
● Molecule 2: RSV Integrase



● Molecule 2: RSV Integrase



● Molecule 3: DNA (5'-D(*AP*AP*TP*GP*TP*TP*GP*TP*CP*TP*TP*AP*TP*GP*CP*AP*AP*TP*AP*CP*TP*C)-3')



● Molecule 3: DNA (5'-D(*AP*AP*TP*GP*TP*TP*GP*TP*CP*TP*TP*AP*TP*GP*CP*AP*AP*TP*AP*CP*TP*C)-3')



● Molecule 4: DNA (5'-D(*CP*TP*TP*CP*TP*CP*TP*C)-3')



● Molecule 4: DNA (5'-D(*CP*TP*TP*CP*TP*CP*TP*C)-3')



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

Chain k:  100%

There are no outlier residues recorded for this chain.

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

Chain n:  100%

There are no outlier residues recorded for this chain.

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	124.94Å 157.85Å 126.58Å 90.00° 110.94° 90.00°	Depositor
Resolution (Å)	49.22 – 3.80 49.22 – 3.61	Depositor EDS
% Data completeness (in resolution range)	97.0 (49.22-3.80) 77.5 (49.22-3.61)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.18 (at 3.57Å)	Xtriage
Refinement program	PHENIX (dev_2210: ???)	Depositor
R, R_{free}	0.254 , 0.294 0.266 , 0.300	Depositor DCC
R_{free} test set	2548 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å ²)	137.5	Xtriage
Anisotropy	0.201	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 150.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.049 for l,-k,h	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	19335	wwPDB-VP
Average B, all atoms (Å ²)	206.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.17% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

5.1 Standard geometry

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

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	0.25	0/2139	0.47	0/2882
1	B	0.24	0/2030	0.46	0/2730
1	C	0.26	0/2046	0.51	0/2752
1	D	0.25	0/2000	0.50	0/2681
1	E	0.25	0/2139	0.48	0/2882
1	F	0.24	0/2034	0.46	0/2736
1	G	0.25	0/2008	0.47	0/2698
1	H	0.25	0/2096	0.47	0/2818
2	I	0.56	0/981	0.84	0/1513
2	L	0.53	0/888	0.84	0/1368
3	J	0.57	0/497	1.02	0/765
3	M	0.55	0/409	1.08	0/630
4	K	0.59	0/169	1.14	0/258
4	N	0.49	0/169	1.11	0/258
5	k	0.64	0/179	1.02	0/275
5	n	0.57	0/179	1.02	0/275
All	All	0.32	0/19963	0.60	0/27521

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	G	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	G	214	ARG	Peptide

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	2099	0	2121	49	0
1	B	1993	0	2016	30	0
1	C	2009	0	2032	55	0
1	D	1967	0	1984	46	0
1	E	2099	0	2121	64	0
1	F	1997	0	2018	45	0
1	G	1972	0	2001	43	0
1	H	2058	0	2075	38	0
2	I	869	0	470	23	0
2	L	787	0	424	16	0
3	J	445	0	253	16	0
3	M	366	0	208	12	0
4	K	154	0	95	3	0
4	N	154	0	95	2	0
5	k	161	0	94	0	0
5	n	161	0	94	0	0
6	A	1	0	0	0	0
6	B	1	0	0	0	0
6	C	1	0	0	0	0
6	D	1	0	0	0	0
6	E	1	0	0	0	0
6	F	1	0	0	0	0
6	G	1	0	0	0	0
6	H	1	0	0	0	0
7	A	4	0	0	0	0
7	B	20	0	0	0	0
7	F	11	0	0	0	0
7	G	1	0	0	0	0
All	All	19335	0	18101	392	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 11.

The worst 5 of 392 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:H:120:THR:HG1	1:H:142:HIS:HE2	1.15	0.94
1:C:135:MSE:HE2	1:C:142:HIS:HB2	1.64	0.79
1:E:80:THR:HG1	1:E:104:HIS:HE2	1.30	0.77
2:L:12:DA:H2'	2:L:13:DA:C8	2.21	0.76
1:C:244:ARG:HH22	1:E:258:ILE:HB	1.52	0.74

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	267/270 (99%)	253 (95%)	14 (5%)	0	100	100
1	B	249/270 (92%)	243 (98%)	6 (2%)	0	100	100
1	C	252/270 (93%)	237 (94%)	13 (5%)	2 (1%)	19	57
1	D	240/270 (89%)	230 (96%)	10 (4%)	0	100	100
1	E	267/270 (99%)	254 (95%)	13 (5%)	0	100	100
1	F	250/270 (93%)	244 (98%)	6 (2%)	0	100	100
1	G	244/270 (90%)	236 (97%)	8 (3%)	0	100	100
1	H	257/270 (95%)	248 (96%)	9 (4%)	0	100	100
All	All	2026/2160 (94%)	1945 (96%)	79 (4%)	2 (0%)	51	83

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	213	TRP
1	C	212	HIS

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	222/212 (105%)	213 (96%)	9 (4%)	30	59
1	B	212/212 (100%)	212 (100%)	0	100	100
1	C	213/212 (100%)	208 (98%)	5 (2%)	50	72
1	D	207/212 (98%)	203 (98%)	4 (2%)	57	76
1	E	222/212 (105%)	217 (98%)	5 (2%)	50	72
1	F	212/212 (100%)	207 (98%)	5 (2%)	49	71
1	G	209/212 (99%)	203 (97%)	6 (3%)	42	67
1	H	217/212 (102%)	214 (99%)	3 (1%)	67	81
All	All	1714/1696 (101%)	1677 (98%)	37 (2%)	52	72

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	91	THR
1	E	46	LEU
1	H	32	GLU
1	D	92	GLN
1	E	17	ARG

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

Mol	Chain	Res	Type
1	D	142	HIS
1	E	51	ASN
1	F	102	GLN
1	D	92	GLN
1	E	151	GLN

5.3.3 RNA ⓘ

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 44 ligands modelled in this entry, 44 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.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.