



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 6, 2026 – 10:50 AM UTC

PDB ID : 8KAI / pdb_00008kai
Title : Crystal structure of SpyCas9-crRNA-tracrRNA complex bound to 17nt target DNA
Authors : Chen, Y.; Chen, J.; Liu, L.
Deposited on : 2023-08-03
Resolution : 3.49 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

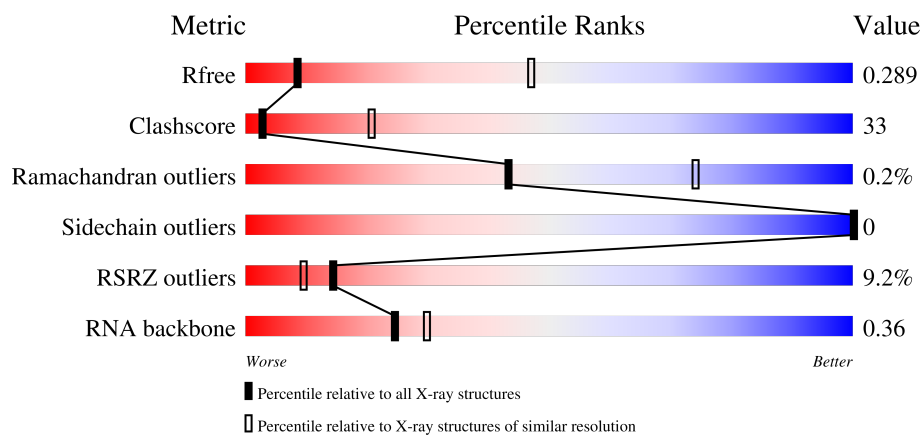
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.49 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}	180053	1085 (3.54-3.46)
Clashscore	190562	1140 (3.54-3.46)
Ramachandran outliers	187476	1113 (3.54-3.46)
Sidechain outliers	187428	1114 (3.54-3.46)
RSRZ outliers	180081	1084 (3.54-3.46)
RNA backbone	3983	1010 (3.98-3.02)

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 $\leq 5\%$. 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	A	34	3% 26% 47% 18% 9%
1	E	34	6% 26% 53% 12% 9%
2	B	1368	7% 47% 49% ..
2	F	1368	13% 45% 51% ..

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	C	25	52% 48%
3	G	25	40% 60%
4	D	11	9% 18% 82%
4	H	11	9% 45% 55%
5	I	65	14% 58% 22%
5	J	65	2% 22% 55% 20%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 27007 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 RNA chain called RNA (34-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	34	Total	C	N	O	P	0	0	0
			725	325	127	239	34			
1	E	31	Total	C	N	O	P	0	0	0
			663	297	118	217	31			

- Molecule 2 is a protein called CRISPR-associated endonuclease Cas9/Csn1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1326	Total	C	N	O	S	0	0	0
			10769	6854	1869	2024	22			
2	F	1327	Total	C	N	O	S	0	0	0
			10698	6816	1845	2014	23			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	10	ALA	ASP	engineered mutation	UNP Q99ZW2
B	840	ALA	HIS	engineered mutation	UNP Q99ZW2
F	10	ALA	ASP	engineered mutation	UNP Q99ZW2
F	840	ALA	HIS	engineered mutation	UNP Q99ZW2

- Molecule 3 is a DNA chain called DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	25	Total	C	N	O	P	0	0	0
			501	244	83	150	24			
3	G	25	Total	C	N	O	P	0	0	0
			501	244	83	150	24			

- Molecule 4 is a DNA chain called DNA (5'-D(*TP*TP*TP*AP*GP*GP*TP*AP*TP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	11	Total	C	N	O	P	0	0	0
			225	110	37	68	10			
4	H	11	Total	C	N	O	P	0	0	0
			225	110	37	68	10			

- Molecule 5 is a RNA chain called RNA (65-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	63	Total	C	N	O	P	0	0	0
			1348	603	245	437	63			
5	J	63	Total	C	N	O	P	0	0	0
			1348	603	245	437	63			

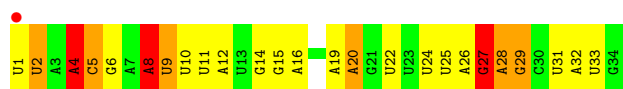
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	1	Total	O	0	0
			1	1		
6	F	3	Total	O	0	0
			3	3		

3 Residue-property plots

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.

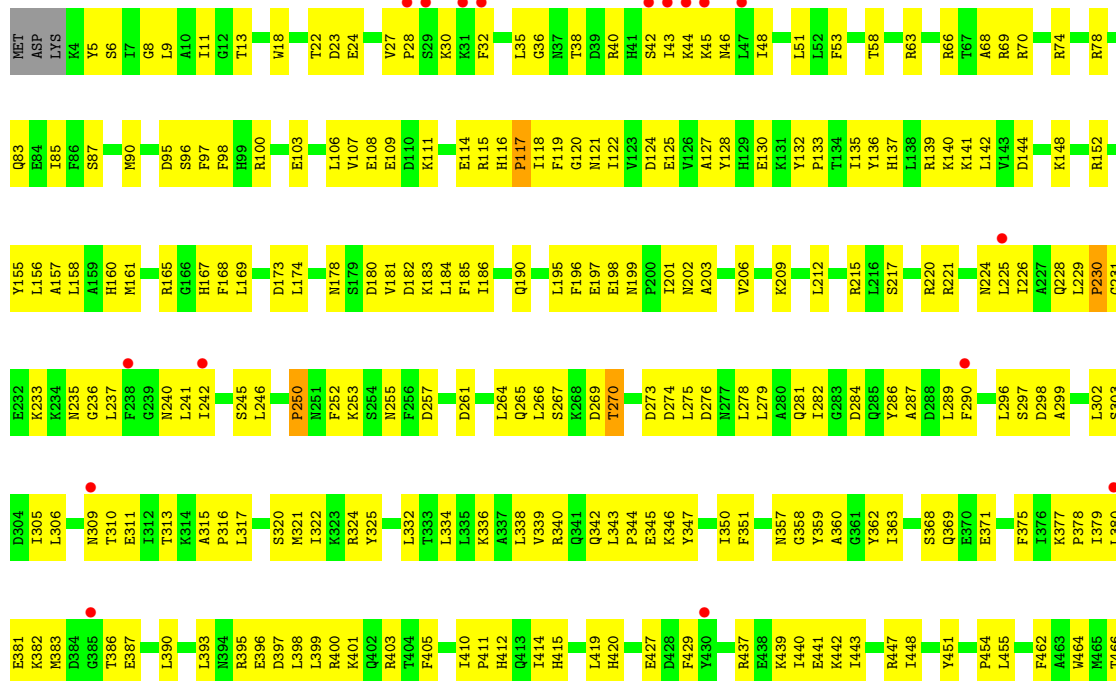
- Molecule 1: RNA (34-MER)

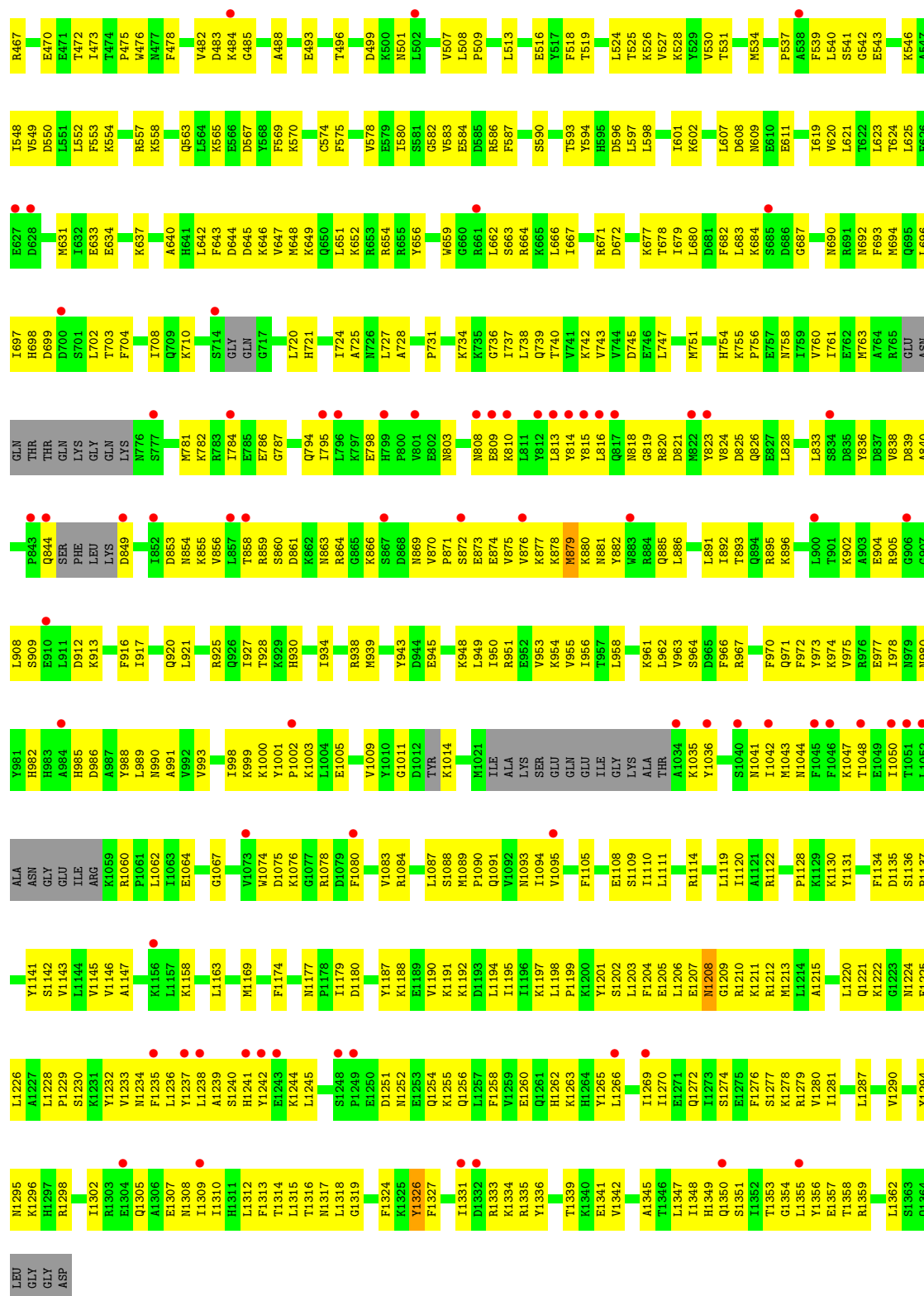


- Molecule 1: RNA (34-MER)

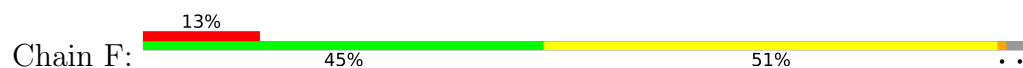


- Molecule 2: CRISPR-associated endonuclease Cas9/Csn1

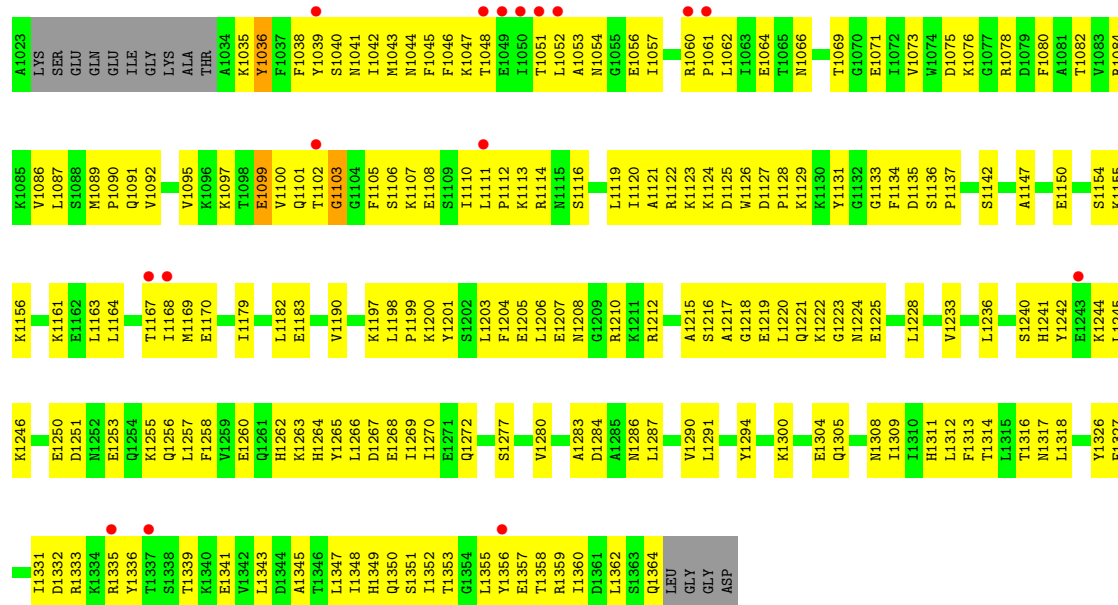




● Molecule 2: CRISPR-associated endonuclease Cas9/Csn1







• Molecule 3: DNA (25-MER)



• Molecule 3: DNA (25-MER)



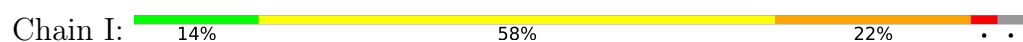
• Molecule 4: DNA (5'-D(*TP*TP*TP*AP*GP*GP*TP*AP*TP*TP*G)-3')

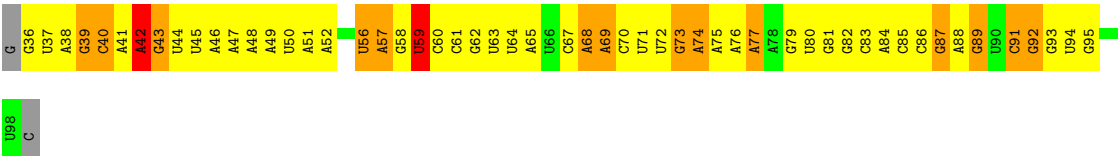


• Molecule 4: DNA (5'-D(*TP*TP*TP*AP*GP*GP*TP*AP*TP*TP*G)-3')

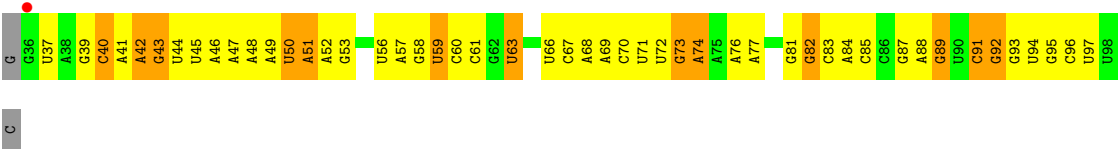


• Molecule 5: RNA (65-MER)





● Molecule 5: RNA (65-MER)



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	144.30Å 130.19Å 146.41Å 90.00° 104.02° 90.00°	Depositor
Resolution (Å)	48.57 – 3.49 48.57 – 3.49	Depositor EDS
% Data completeness (in resolution range)	79.0 (48.57-3.49) 79.0 (48.57-3.49)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.43 (at 3.48Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.225 , 0.294 0.224 , 0.289	Depositor DCC
R_{free} test set	2620 reflections (3.77%)	wwPDB-VP
Wilson B-factor (Å ²)	62.1	Xtriage
Anisotropy	0.306	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 76.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.40$, $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	0.087 for l,-k,h	Xtriage
F_o, F_c correlation	0.83	EDS
Total number of atoms	27007	wwPDB-VP
Average B, all atoms (Å ²)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.60% 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

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	0.65	0/811	1.03	4/1261 (0.3%)
1	E	0.61	0/742	0.93	0/1154
2	B	0.65	0/10954	0.97	6/14725 (0.0%)
2	F	0.67	2/10882 (0.0%)	0.96	7/14639 (0.0%)
3	C	0.81	0/559	1.03	0/859
3	G	0.77	0/559	0.99	0/859
4	D	0.86	0/251	0.87	0/387
4	H	0.78	0/251	0.93	0/387
5	I	0.66	0/1509	0.93	2/2350 (0.1%)
5	J	0.58	0/1509	0.91	0/2350
All	All	0.66	2/28027 (0.0%)	0.96	19/38971 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	1103	GLY	C-O	-17.89	1.14	1.24
2	F	425	ARG	CG-CD	-6.61	1.32	1.52

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1326	TYR	N-CA-C	-6.60	102.02	110.53
1	A	27	G	C4'-C3'-O3'	6.20	122.29	113.00
2	F	226	ILE	CA-CB-CG1	-6.15	99.95	110.40
5	I	42	A	C2'-C3'-O3'	6.11	122.86	113.70
2	F	902	LYS	CD-CE-NZ	-5.97	92.78	111.90

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	725	0	362	29	0
1	E	663	0	331	30	0
2	B	10769	0	10863	722	2
2	F	10698	0	10745	835	0
3	C	501	0	287	12	0
3	G	501	0	287	19	0
4	D	225	0	129	7	0
4	H	225	0	129	11	0
5	I	1348	0	678	70	0
5	J	1348	0	678	69	0
6	B	1	0	0	0	0
6	F	3	0	0	0	0
All	All	27007	0	24489	1682	2

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 33.

The worst 5 of 1682 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:380:LEU:O	2:B:386:THR:CG2	1.70	1.39
2:B:1222:LYS:NZ	2:B:1315:LEU:O	1.59	1.33
2:F:878:LYS:HB3	2:F:879:MET:SD	1.74	1.28
2:F:878:LYS:CD	2:F:879:MET:HE1	1.65	1.24
2:F:878:LYS:HD2	2:F:879:MET:CE	1.76	1.16

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:202:ASN:OD1	2:B:541:SER:OG[2_545]	2.07	0.13
2:B:228:GLN:NE2	2:B:543:GLU:OE2[2_545]	2.18	0.02

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	
2	B	1312/1368 (96%)	1279 (98%)	29 (2%)	4 (0%)	36	67
2	F	1313/1368 (96%)	1266 (96%)	45 (3%)	2 (0%)	43	74
All	All	2625/2736 (96%)	2545 (97%)	74 (3%)	6 (0%)	43	74

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	270	THR
2	F	1020	LYS
2	F	117	PRO
2	B	117	PRO
2	B	250	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	
2	B	1173/1225 (96%)	1173 (100%)	0	100	100
2	F	1156/1225 (94%)	1156 (100%)	0	100	100
All	All	2329/2450 (95%)	2329 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	881	ASN
2	F	1305	GLN
2	F	894	GLN
2	F	982	HIS
2	B	1252	ASN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	33/34 (97%)	9 (27%)	3 (9%)
1	E	30/34 (88%)	9 (30%)	1 (3%)
5	I	62/65 (95%)	20 (32%)	2 (3%)
5	J	62/65 (95%)	19 (30%)	1 (1%)
All	All	187/198 (94%)	57 (30%)	7 (3%)

5 of 57 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	2	U
1	A	4	A
1	A	5	C
1	A	6	G
1	A	9	U

5 of 7 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	E	27	G
5	I	42	A
5	J	42	A
5	I	68	A
1	A	28	A

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 oligosaccharides 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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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, 95th 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	A	34/34 (100%)	-0.36	1 (2%) 53 31	19, 34, 129, 166	0
1	E	31/34 (91%)	0.19	2 (6%) 25 15	40, 71, 175, 230	0
2	B	1326/1368 (96%)	0.34	89 (6%) 24 15	13, 62, 175, 215	0
2	F	1327/1368 (97%)	0.73	172 (12%) 7 6	11, 88, 145, 197	0
3	C	25/25 (100%)	-0.35	0 100 100	25, 38, 81, 88	0
3	G	25/25 (100%)	0.23	0 100 100	45, 59, 118, 139	0
4	D	11/11 (100%)	0.20	1 (9%) 15 10	30, 38, 122, 159	0
4	H	11/11 (100%)	0.20	1 (9%) 15 10	33, 60, 119, 189	0
5	I	63/65 (96%)	-0.41	0 100 100	17, 71, 122, 170	0
5	J	63/65 (96%)	-0.31	1 (1%) 70 45	26, 58, 137, 191	0
All	All	2916/3006 (97%)	0.47	267 (9%) 14 9	11, 70, 162, 230	0

The worst 5 of 267 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	305	ILE	7.8
2	B	801	VAL	7.1
2	F	73	THR	7.1
2	B	809	GLU	6.9
2	F	362	TYR	6.7

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.3 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.