



wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 02:35 am BST

PDB ID : 6IFO
Title : Crystal structure of AcrIIA2-SpyCas9-sgRNA ternary complex
Authors : Liu, L.; Wang, Y.
Deposited on : 2018-09-20
Resolution : 3.31 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
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

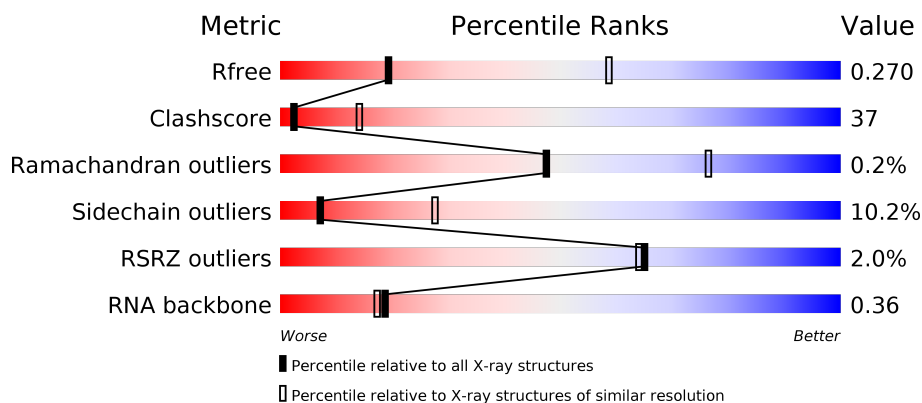
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.31 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	1089 (3.36-3.28)
Clashscore	141614	1137 (3.36-3.28)
Ramachandran outliers	138981	1115 (3.36-3.28)
Sidechain outliers	138945	1114 (3.36-3.28)
RSRZ outliers	127900	1059 (3.36-3.28)
RNA backbone	3102	1125 (3.74-2.90)

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 $\leq 5\%$. 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	1369	<div> <div>2%</div> <div>65% 30% . .</div> </div>
1	B	1369	<div> <div>2%</div> <div>63% 31% 6% .</div> </div>
2	C	99	<div> <div>37% 29% 22% 11%</div> </div>
2	D	99	<div> <div>% 27% 40% 23% 9%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	123	<div><div><div></div><div></div><div></div><div></div><div></div></div><div><div>2%</div><div>28%</div><div>46%</div><div>19%</div><div>• 7%</div></div></div>
3	F	123	<div><div><div></div><div></div><div></div><div></div><div></div></div><div><div>2%</div><div>17%</div><div>53%</div><div>25%</div><div>• •</div></div></div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 27432 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 CRISPR-associated endonuclease Cas9/Csn1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1360	Total	C	N	O	S	0	0	0
			10904	6939	1886	2055	24			
1	B	1360	Total	C	N	O	S	0	0	0
			10900	6937	1886	2053	24			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP Q99ZW2
B	0	SER	-	expression tag	UNP Q99ZW2

- Molecule 2 is a RNA chain called RNA (99-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	88	Total	C	N	O	P	0	0	0
			1881	841	338	614	88			
2	D	90	Total	C	N	O	P	0	0	0
			1927	861	348	628	90			

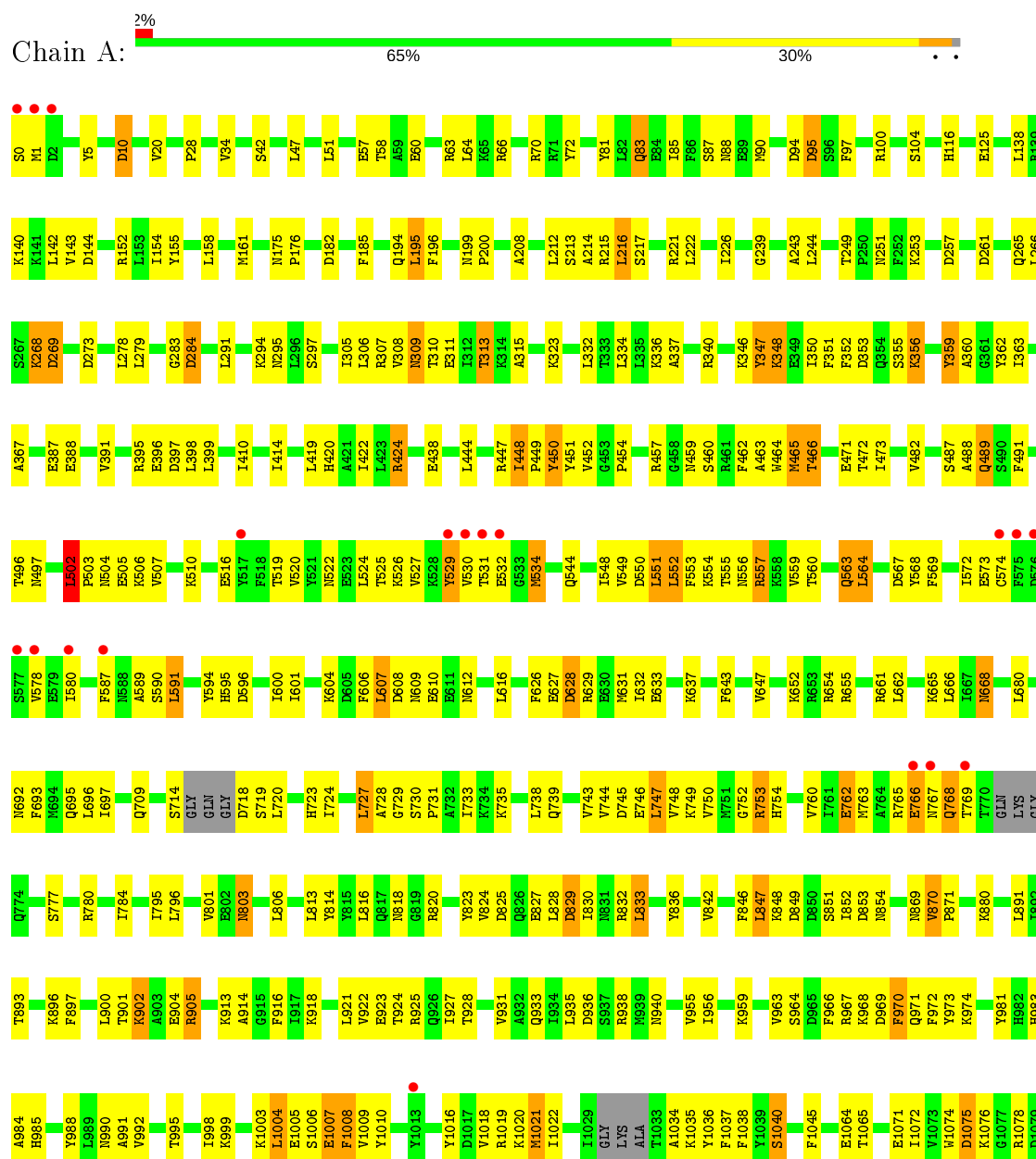
- Molecule 3 is a protein called AcrIIA2.

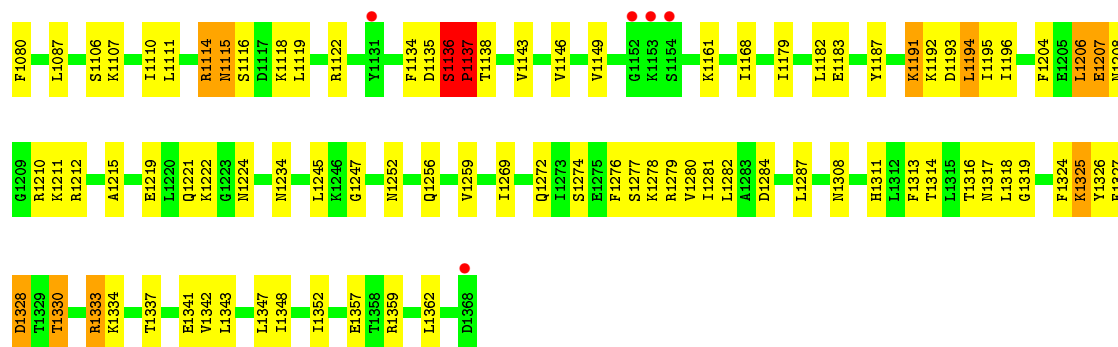
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	118	Total	C	N	O	S	0	0	0
			920	575	136	205	4			
3	E	115	Total	C	N	O	S	0	0	0
			900	565	136	195	4			

3 Residue-property plots

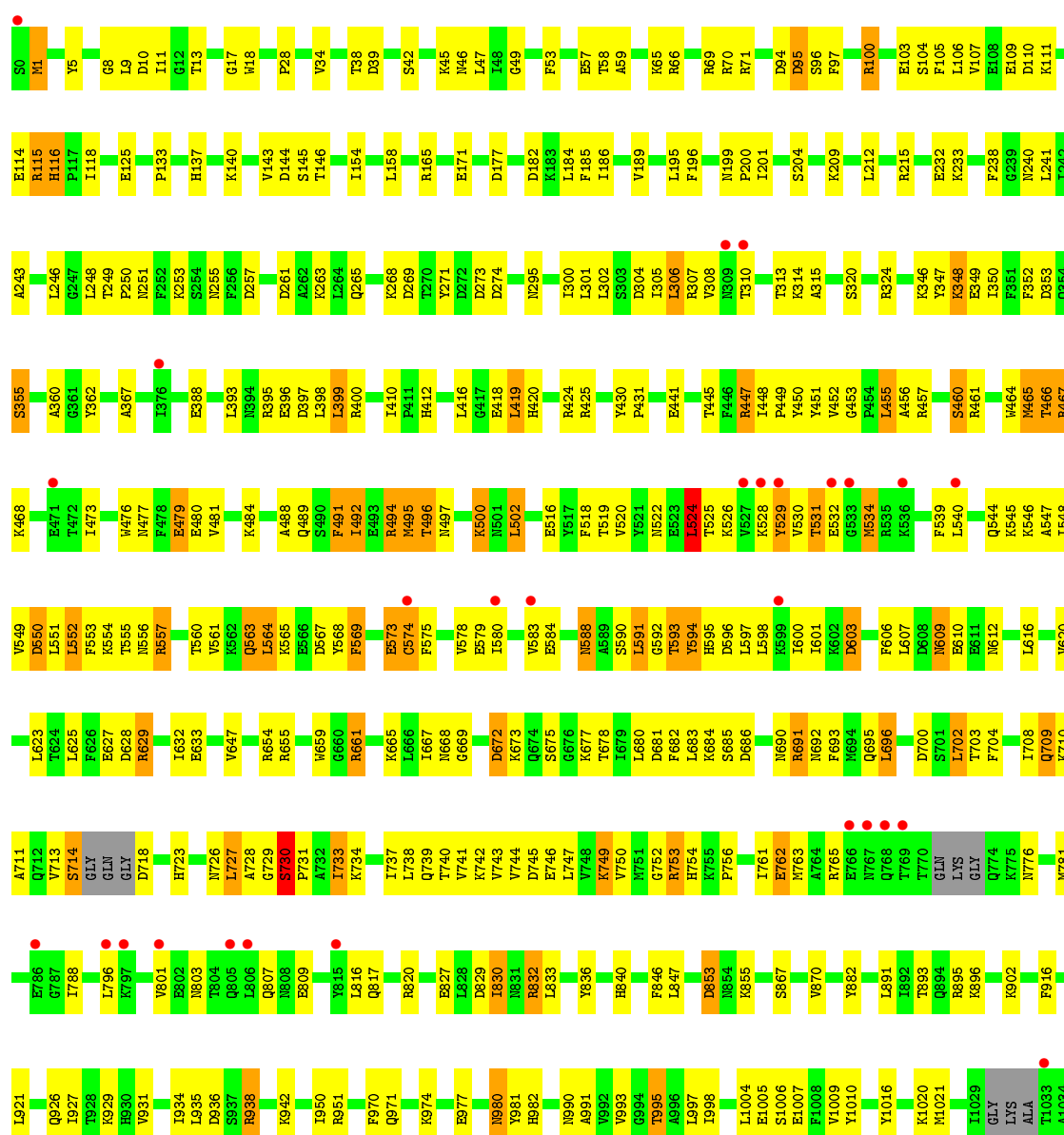
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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.

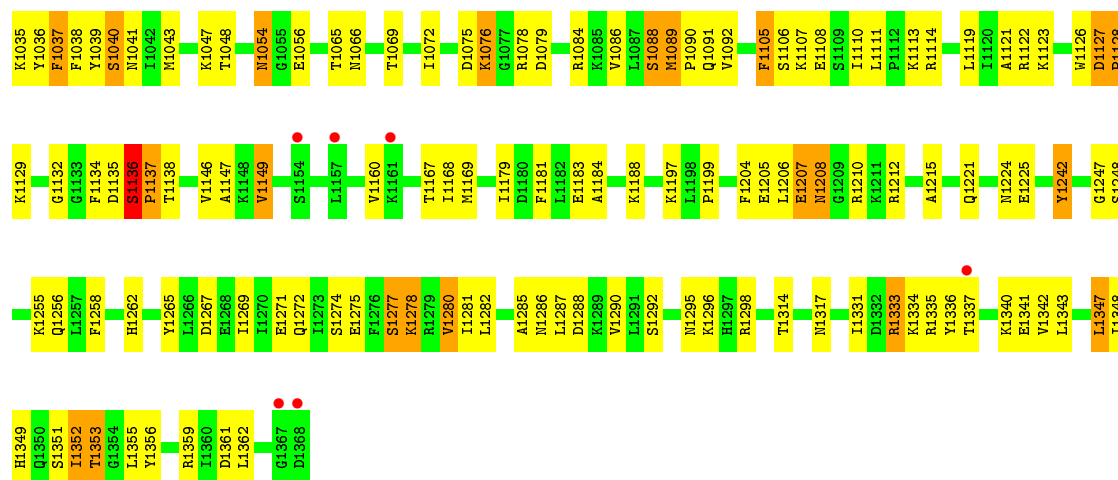
- Molecule 1: CRISPR-associated endonuclease Cas9/Csn1





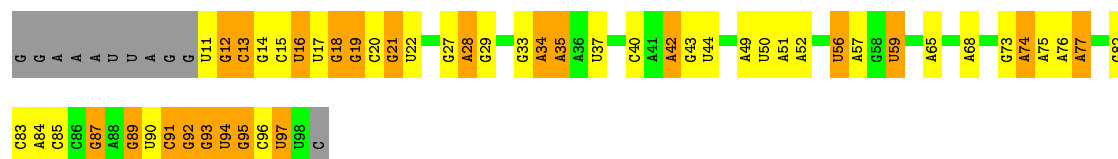
• Molecule 1: CRISPR-associated endonuclease Cas9/Csn1





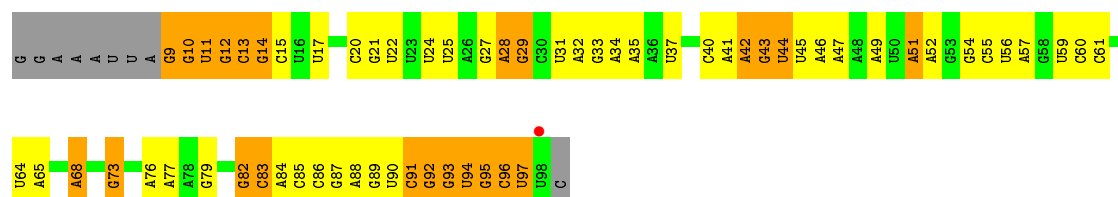
• Molecule 2: RNA (99-MER)

Chain C: 37% 29% 22% 11%



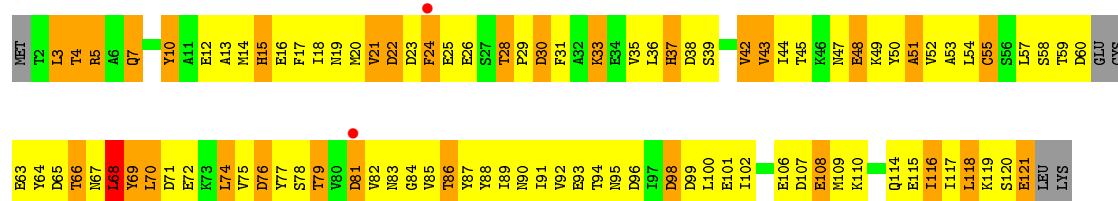
• Molecule 2: RNA (99-MER)

Chain D: 27% 40% 23% 9%



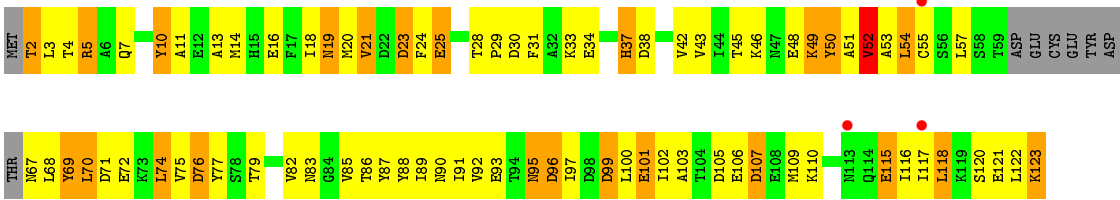
• Molecule 3: AcrIIA2

Chain F: 2% 17% 53% 25%



• Molecule 3: AcrIIA2

Chain E: 28% 46% 19% 7%



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	106.16Å 214.57Å 131.97Å 90.00° 102.89° 90.00°	Depositor
Resolution (Å)	49.51 – 3.31 49.51 – 3.31	Depositor EDS
% Data completeness (in resolution range)	66.1 (49.51-3.31) 66.1 (49.51-3.31)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.60 (at 3.33Å)	Xtriage
Refinement program	PHENIX (1.14_3247: ???)	Depositor
R, R_{free}	0.222 , 0.268 0.223 , 0.270	Depositor DCC
R_{free} test set	2844 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	62.7	Xtriage
Anisotropy	0.292	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 21.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.87	EDS
Total number of atoms	27432	wwPDB-VP
Average B, all atoms (Å ²)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.74% 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	0.53	2/11097 (0.0%)	0.72	6/14959 (0.0%)
1	B	0.58	3/11093 (0.0%)	0.77	13/14954 (0.1%)
2	C	0.51	1/2105 (0.0%)	0.80	1/3279 (0.0%)
2	D	0.54	0/2157	0.79	0/3361
3	E	0.44	0/911	0.73	1/1237 (0.1%)
3	F	0.55	0/932	0.81	4/1269 (0.3%)
All	All	0.55	6/28295 (0.0%)	0.75	25/39059 (0.1%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	21	G	O3'-P	-5.66	1.54	1.61
1	A	200	PRO	N-CD	5.45	1.55	1.47
1	B	200	PRO	N-CD	5.31	1.55	1.47
1	B	1137	PRO	N-CD	5.15	1.55	1.47
1	B	1128	PRO	N-CD	5.11	1.55	1.47

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	42	A	C2'-C3'-O3'	6.80	124.57	113.70
3	F	42	VAL	CB-CA-C	-6.56	98.93	111.40
1	A	1136	SER	C-N-CD	6.22	141.46	128.40
1	B	448	ILE	C-N-CD	6.05	141.11	128.40
1	A	502	LEU	C-N-CD	6.05	141.10	128.40

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	10904	0	10832	634	0
1	B	10900	0	10835	773	0
2	C	1881	0	945	78	0
2	D	1927	0	966	168	0
3	E	900	0	831	188	0
3	F	920	0	824	281	0
All	All	27432	0	25233	1969	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 37.

The worst 5 of 1969 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:B:569:PHE:HE1	1:B:578:VAL:CG1	1.05	1.61
1:B:531:THR:HG21	1:B:575:PHE:CE2	1.09	1.58
1:A:970:PHE:CD1	1:A:1080:PHE:CE1	1.94	1.52
1:B:569:PHE:CE1	1:B:578:VAL:HG11	1.02	1.52
1:B:569:PHE:CD1	1:B:578:VAL:HG21	1.41	1.50

There are no symmetry-related clashes.

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
1	A	1352/1369 (99%)	1309 (97%)	42 (3%)	1 (0%)	51 81

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	1352/1369 (99%)	1314 (97%)	38 (3%)	0	100	100
3	E	111/123 (90%)	99 (89%)	10 (9%)	2 (2%)	8	35
3	F	114/123 (93%)	98 (86%)	12 (10%)	4 (4%)	3	22
All	All	2929/2984 (98%)	2820 (96%)	102 (4%)	7 (0%)	47	76

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	F	21	VAL
3	F	66	THR
3	F	51	ALA
3	F	68	LEU
3	E	52	VAL

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	1169/1228 (95%)	1071 (92%)	98 (8%)	11	36
1	B	1168/1228 (95%)	1065 (91%)	103 (9%)	10	34
3	E	99/114 (87%)	73 (74%)	26 (26%)	0	2
3	F	100/114 (88%)	69 (69%)	31 (31%)	0	1
All	All	2536/2684 (94%)	2278 (90%)	258 (10%)	7	27

5 of 258 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	424	ARG
1	B	573	GLU
3	E	37	HIS
1	B	460	SER
1	B	496	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1317	ASN
1	B	544	GLN
1	B	1286	ASN
1	B	255	ASN
1	B	489	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	C	87/99 (87%)	30 (34%)	7 (8%)
2	D	90/99 (90%)	31 (34%)	8 (8%)
All	All	177/198 (89%)	61 (34%)	15 (8%)

5 of 61 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	C	12	G
2	C	13	C
2	C	16	U
2	C	18	G
2	C	19	G

5 of 15 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	C	94	U
2	D	9	G
2	D	89	G
2	C	93	G
2	D	42	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 carbohydrates 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	1360/1369 (99%)	-0.23	24 (1%) 68 67	32, 64, 102, 160	0
1	B	1360/1369 (99%)	-0.16	34 (2%) 57 55	23, 57, 113, 171	0
2	C	88/99 (88%)	-0.37	0 100 100	39, 67, 146, 176	0
2	D	90/99 (90%)	-0.34	1 (1%) 80 81	32, 57, 137, 159	0
3	E	115/123 (93%)	0.03	3 (2%) 56 53	52, 83, 109, 120	0
3	F	118/123 (95%)	-0.03	2 (1%) 70 68	44, 75, 101, 109	0
All	All	3131/3182 (98%)	-0.19	64 (2%) 65 64	23, 63, 110, 176	0

The worst 5 of 64 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	767	ASN	9.4
1	A	574	CYS	7.5
1	A	2	ASP	7.2
1	A	576	ASP	7.0
1	B	310	THR	7.0

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 carbohydrates 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.