



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 5, 2017 – 06:20 PM EDT

PDB ID : 5NG6
Title : Crystal structure of FnCas12a bound to a crRNA
Authors : Swarts, D.C.; van der Oost, J.; Jinek, M.
Deposited on : 2017-03-16
Resolution : 3.34 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

<http://wwpdb.org/validation/2016/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)	:	1.9-1692
EDS	:	rb-20029077
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20029077

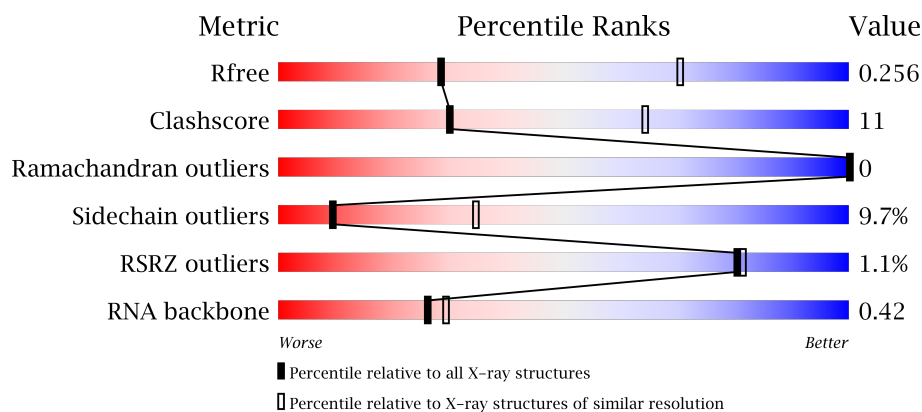
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.34 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}	100719	1167 (3.40-3.28)
Clashscore	112137	1239 (3.40-3.28)
Ramachandran outliers	110173	1219 (3.40-3.28)
Sidechain outliers	110143	1218 (3.40-3.28)
RSRZ outliers	101464	1176 (3.40-3.28)
RNA backbone	2435	1121 (3.88-2.80)

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 on the lower 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 $\leq 5\%$. 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	1302	<div> <div>67%</div> <div>26%</div> <div>• •</div> </div>
1	C	1302	<div> <div>• %</div> <div>69%</div> <div>25%</div> <div>• •</div> </div>
1	E	1302	<div> <div>2%</div> <div>66%</div> <div>27%</div> <div>• •</div> </div>
1	G	1302	<div> <div>• %</div> <div>65%</div> <div>27%</div> <div>• •</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	B	43	
2	D	43	
2	F	43	
2	H	43	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	MG	C	1401	-	-	-	X

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 43813 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 Cpf1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1268	Total	C	N	O	S	0	0	0
			10476	6737	1729	1989	21			
1	C	1268	Total	C	N	O	S	0	0	0
			10476	6737	1729	1989	21			
1	E	1262	Total	C	N	O	S	0	0	0
			10425	6708	1721	1976	20			
1	G	1253	Total	C	N	O	S	0	0	0
			10355	6665	1705	1964	21			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	SER	-	expression tag	UNP A0Q7Q2
A	0	ASN	-	expression tag	UNP A0Q7Q2
A	1	ALA	-	expression tag	UNP A0Q7Q2
C	-1	SER	-	expression tag	UNP A0Q7Q2
C	0	ASN	-	expression tag	UNP A0Q7Q2
C	1	ALA	-	expression tag	UNP A0Q7Q2
E	-1	SER	-	expression tag	UNP A0Q7Q2
E	0	ASN	-	expression tag	UNP A0Q7Q2
E	1	ALA	-	expression tag	UNP A0Q7Q2
G	-1	SER	-	expression tag	UNP A0Q7Q2
G	0	ASN	-	expression tag	UNP A0Q7Q2
G	1	ALA	-	expression tag	UNP A0Q7Q2

- Molecule 2 is a RNA chain called crRNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	25	Total	C	N	O	P	0	0	0
			507	227	83	173	24			
2	D	25	Total	C	N	O	P	0	0	0
			507	227	83	173	24			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	25	Total	C	N	O	P	0	0	0
			507	227	83	173	24			
2	H	25	Total	C	N	O	P	0	0	0
			507	227	83	173	24			

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

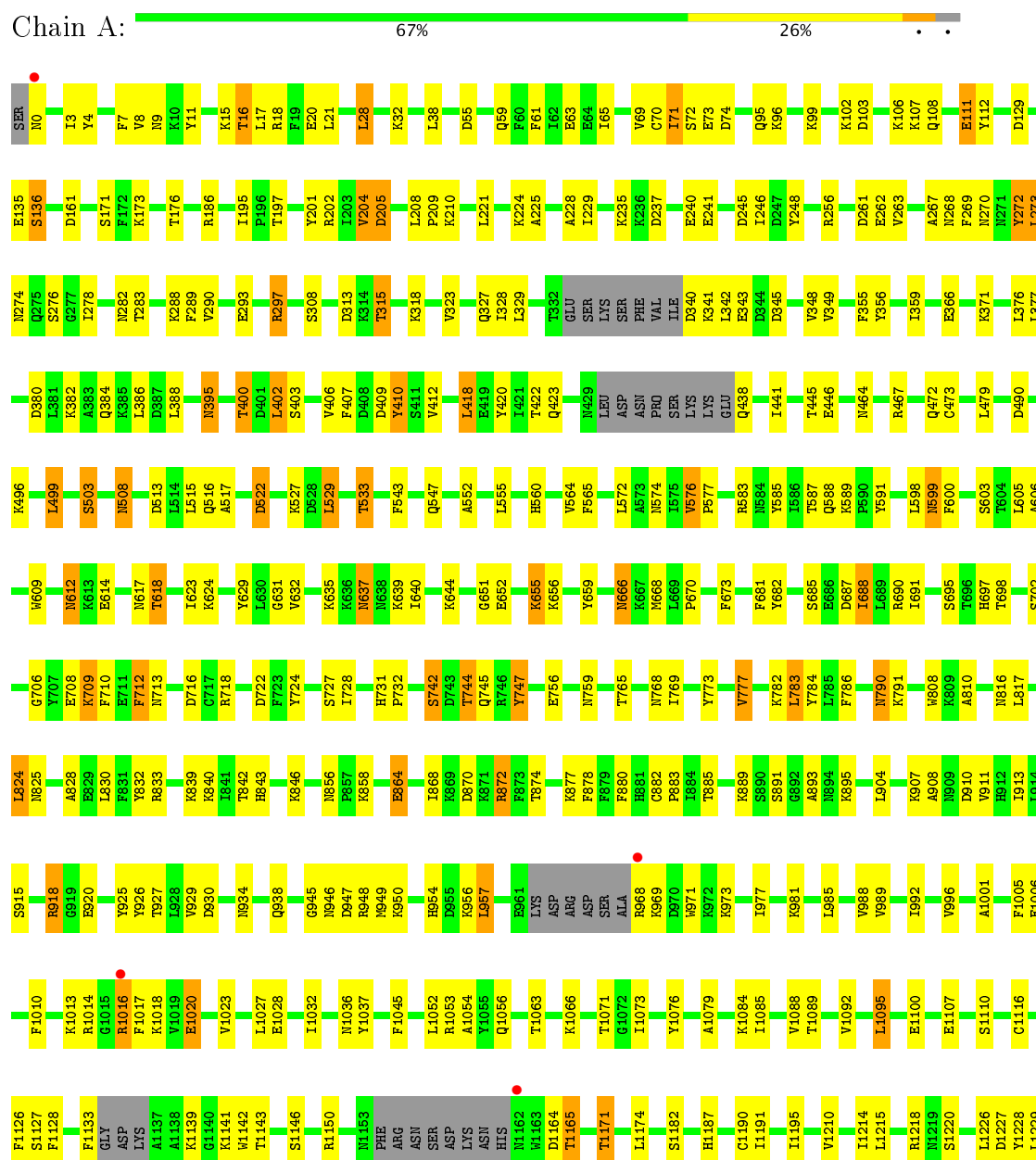
- Molecule 4 is water.

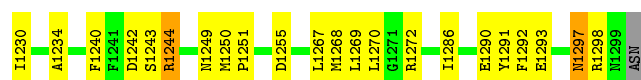
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	5	Total	O	0	0
			5	5		
4	B	10	Total	O	0	0
			10	10		
4	D	6	Total	O	0	0
			6	6		
4	E	5	Total	O	0	0
			5	5		
4	F	10	Total	O	0	0
			10	10		
4	G	3	Total	O	0	0
			3	3		
4	H	6	Total	O	0	0
			6	6		

3 Residue-property plots

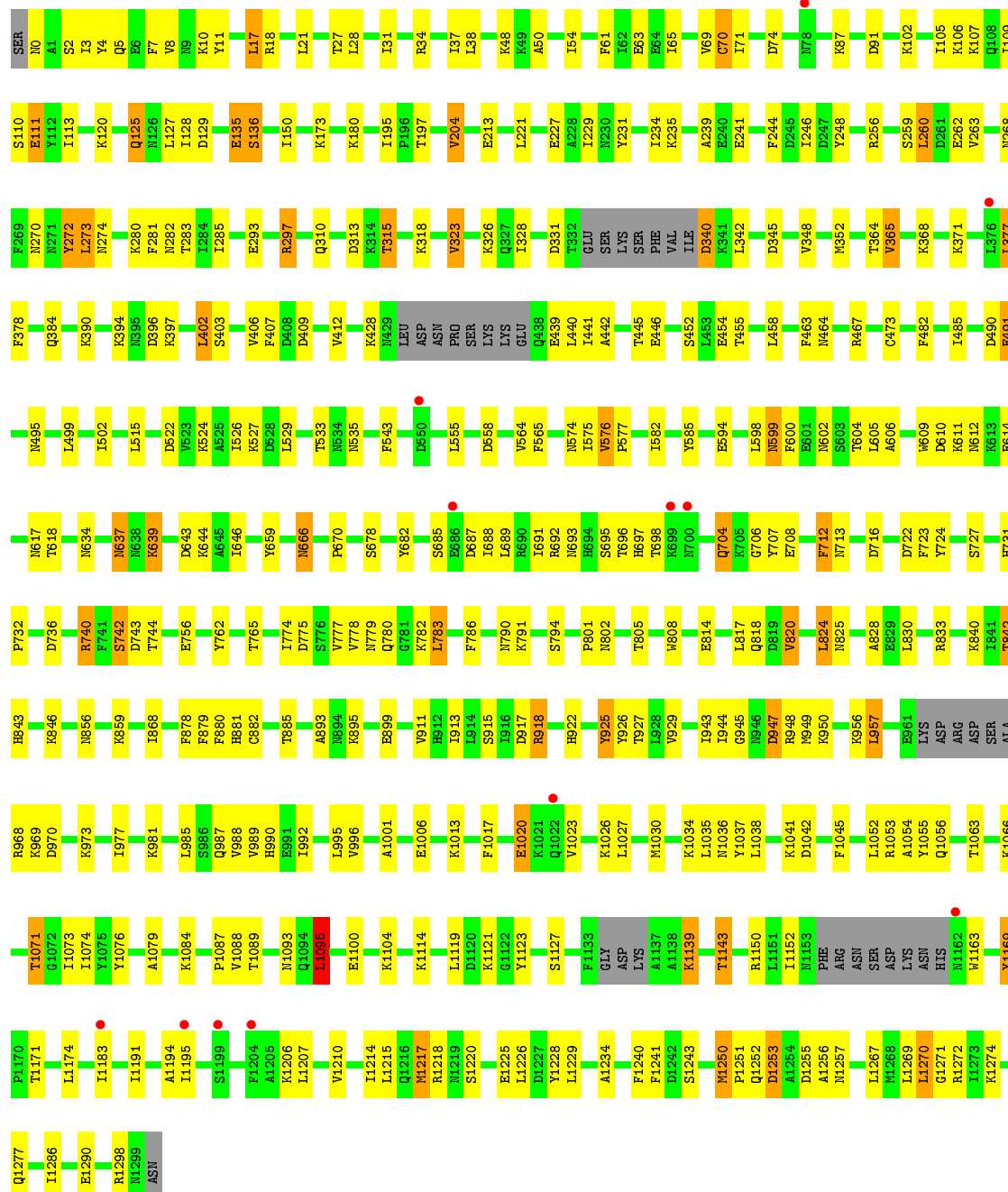
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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 Cpf1



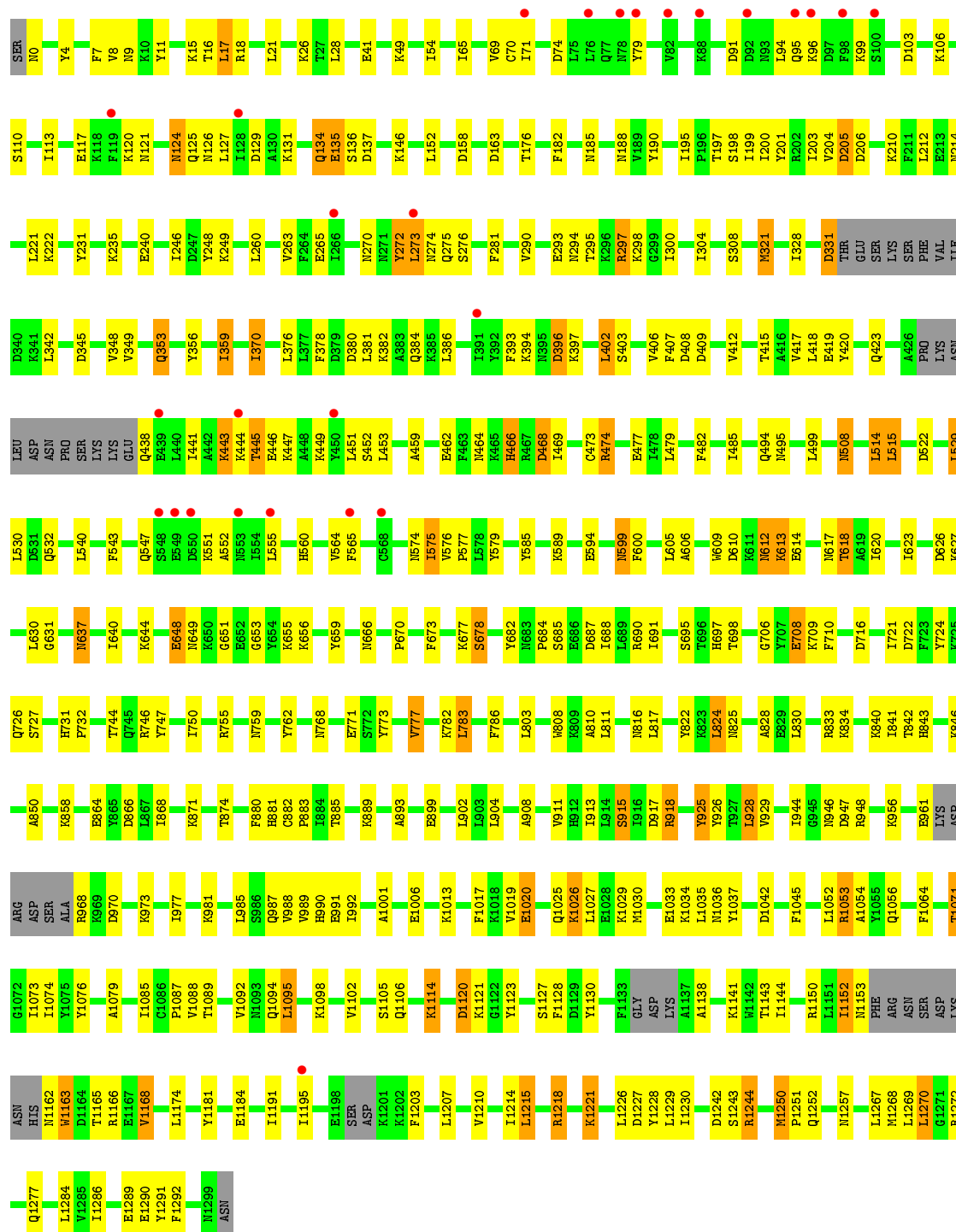


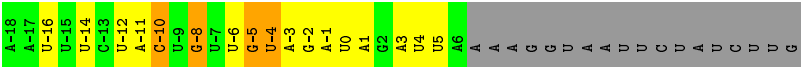
• Molecule 1: CRISPR-associated endonuclease Cpf1



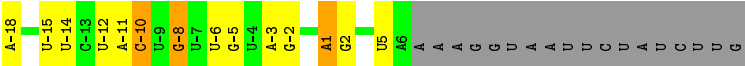
• Molecule 1: CRISPR-associated endonuclease Cpf1



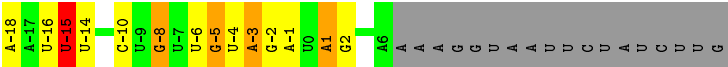
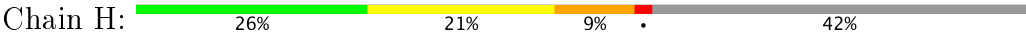




● Molecule 2: crRNA



● Molecule 2: crRNA



4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2	Depositor
Cell constants a, b, c, α , β , γ	214.68Å 277.06Å 135.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.24 – 3.34 49.24 – 3.34	Depositor EDS
% Data completeness (in resolution range)	99.6 (49.24-3.34) 99.8 (49.24-3.34)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.20	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.75 (at 3.33Å)	Xtriage
Refinement program	PHENIX (1.10 _2155: ???)	Depositor
R, R_{free}	0.239 , 0.258 0.236 , 0.256	Depositor DCC
R_{free} test set	5695 reflections (4.86%)	DCC
Wilson B-factor (Å ²)	85.7	Xtriage
Anisotropy	0.128	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 38.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	43813	wwPDB-VP
Average B, all atoms (Å ²)	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.74% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 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 >5	RMSZ	# Z >5
1	A	0.41	0/10683	0.48	2/14342 (0.0%)
1	C	0.38	0/10683	0.47	2/14342 (0.0%)
1	E	0.38	0/10630	0.47	2/14267 (0.0%)
1	G	0.38	0/10558	0.48	2/14169 (0.0%)
2	B	1.03	0/565	0.95	0/878
2	D	0.86	0/565	0.88	0/878
2	F	1.02	0/565	0.88	0/878
2	H	0.93	0/565	0.95	2/878 (0.2%)
All	All	0.44	0/44814	0.51	10/60632 (0.0%)

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	H	-15	U	C2-N1-C1'	7.35	126.52	117.70
1	A	1095	LEU	CA-CB-CG	6.77	130.88	115.30
1	G	1095	LEU	CA-CB-CG	6.24	129.66	115.30
1	E	1095	LEU	CA-CB-CG	6.17	129.49	115.30
1	C	1095	LEU	CA-CB-CG	6.01	129.13	115.30

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	10476	0	10455	222	1
1	C	10476	0	10455	216	0
1	E	10425	0	10402	241	0
1	G	10355	0	10344	226	0
2	B	507	0	254	16	0
2	D	507	0	254	11	0
2	F	507	0	254	10	0
2	H	507	0	254	10	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
3	E	1	0	0	0	0
3	F	1	0	0	0	0
3	G	1	0	0	0	0
3	H	1	0	0	0	0
4	A	5	0	0	1	0
4	B	10	0	0	3	0
4	D	6	0	0	0	0
4	E	5	0	0	0	0
4	F	10	0	0	0	0
4	G	3	0	0	0	0
4	H	6	0	0	0	0
All	All	43813	0	42672	919	1

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 919 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:A:293:GLU:OE2	1:A:297:ARG:NH1	2.09	0.85
1:A:872:ARG:NH1	2:B:-13:C:OP2	2.11	0.84
1:G:1272:ARG:NH2	1:G:1290:GLU:OE2	2.11	0.82
2:B:-17:A:OP1	4:B:301:HOH:O	1.98	0.82
1:A:920:GLU:O	1:A:1016:ARG:NH2	2.11	0.82

All (1) 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)
1:A:235:LYS:NZ	1:A:706:GLY:O[2_535]	2.15	0.05

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	1256/1302 (96%)	1232 (98%)	24 (2%)	0	100	100
1	C	1256/1302 (96%)	1237 (98%)	19 (2%)	0	100	100
1	E	1248/1302 (96%)	1225 (98%)	23 (2%)	0	100	100
1	G	1239/1302 (95%)	1217 (98%)	22 (2%)	0	100	100
All	All	4999/5208 (96%)	4911 (98%)	88 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	
1	A	1153/1185 (97%)	1041 (90%)	112 (10%)	9	34
1	C	1153/1185 (97%)	1050 (91%)	103 (9%)	11	39
1	E	1146/1185 (97%)	1034 (90%)	112 (10%)	9	34
1	G	1140/1185 (96%)	1022 (90%)	118 (10%)	8	31
All	All	4592/4740 (97%)	4147 (90%)	445 (10%)	9	34

5 of 445 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	1163	TRP
1	E	408	ASP
1	G	863	PHE
1	C	1220	SER
1	E	135	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	271	ASN
1	C	405	GLN
1	G	532	GLN
1	A	843	HIS
1	A	1245	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	23/43 (53%)	8 (34%)	0
2	D	23/43 (53%)	8 (34%)	1 (4%)
2	F	23/43 (53%)	7 (30%)	0
2	H	23/43 (53%)	7 (30%)	0
All	All	92/172 (53%)	30 (32%)	1 (1%)

5 of 30 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	-10	C
2	B	-8	G
2	B	-6	U
2	B	-5	G
2	B	-4	U

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	D	-5	G

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 8 ligands modelled in this entry, 8 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 [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	1268/1302 (97%)	-0.02	4 (0%) 93 94	43, 67, 93, 144	0
1	C	1268/1302 (97%)	0.05	12 (0%) 84 85	51, 75, 105, 152	0
1	E	1262/1302 (96%)	0.18	27 (2%) 64 64	50, 78, 125, 166	0
1	G	1253/1302 (96%)	0.07	12 (0%) 82 83	48, 81, 126, 166	0
2	B	25/43 (58%)	-0.07	0 100 100	42, 53, 69, 99	0
2	D	25/43 (58%)	0.04	0 100 100	57, 70, 87, 93	0
2	F	25/43 (58%)	0.01	0 100 100	47, 54, 84, 112	0
2	H	25/43 (58%)	0.03	0 100 100	48, 60, 76, 97	0
All	All	5151/5380 (95%)	0.07	55 (1%) 80 81	42, 74, 117, 166	0

The worst 5 of 55 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	550	ASP	5.3
1	E	78	ASN	4.7
1	C	1162	ASN	4.5
1	E	266	ILE	4.0
1	E	568	CYS	3.8

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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	LLDF	B-factors(Å ²)	Q<0.9
3	MG	C	1401	1/1	0.92	0.47	9.10	56,56,56,56	0
3	MG	G	1401	1/1	0.94	0.26	1.82	56,56,56,56	0
3	MG	E	1401	1/1	0.94	0.15	-2.32	78,78,78,78	0
3	MG	A	1401	1/1	0.95	0.16	-2.45	56,56,56,56	0
3	MG	B	200	1/1	0.93	0.07	-	61,61,61,61	0
3	MG	D	200	1/1	0.97	0.08	-	61,61,61,61	0
3	MG	H	200	1/1	0.95	0.12	-	61,61,61,61	0
3	MG	F	200	1/1	0.99	0.06	-	62,62,62,62	0

6.5 Other polymers

There are no such residues in this entry.