



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 02:44 AM UTC

PDB ID : 5XVN / pdb_00005xvn
Title : E. far Cas1-Cas2/prespacer binary complex
Authors : Xiao, Y.; Ng, S.; Nam, K.H.; Ke, A.
Deposited on : 2017-06-28
Resolution : 3.25 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

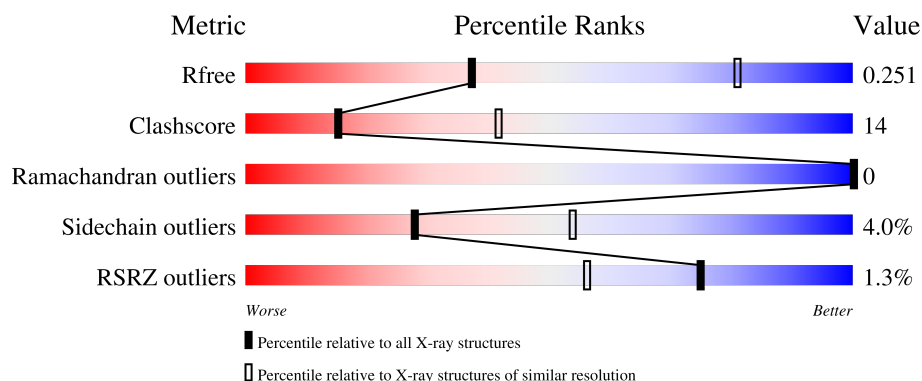
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.25 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.



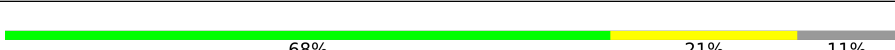
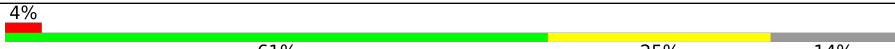
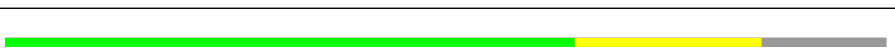
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	180053	1605 (3.30-3.22)
Clashscore	190562	1660 (3.30-3.22)
Ramachandran outliers	187476	1630 (3.30-3.22)
Sidechain outliers	187428	1629 (3.30-3.22)
RSRZ outliers	180081	1605 (3.30-3.22)

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	288	
1	B	288	
1	C	288	
1	D	288	
1	I	288	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	J	288	 76% 22% .
1	K	288	 % 78% 18% ..
1	L	288	 78% 20% .
2	E	109	 3% 63% 30% . 5%
2	F	109	 2% 70% 24% 6% .
2	M	109	 61% 36% ..
2	N	109	 3% 61% 29% 5% 5%
3	G	28	 68% 21% 11%
3	H	28	 4% 61% 25% 14%
3	O	28	 4% 64% 29% 7%
3	P	28	 64% 21% 14%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 24089 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 Cas1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	285	Total	C	N	O	S	0	0	0
			2337	1503	404	420	10			
1	B	288	Total	C	N	O	S	0	0	0
			2359	1515	409	425	10			
1	C	285	Total	C	N	O	S	0	0	0
			2337	1503	404	420	10			
1	D	259	Total	C	N	O	S	0	0	0
			2113	1360	363	381	9			
1	I	285	Total	C	N	O	S	0	0	0
			2337	1503	404	420	10			
1	J	288	Total	C	N	O	S	0	0	0
			2359	1515	409	425	10			
1	K	285	Total	C	N	O	S	0	0	0
			2337	1503	404	420	10			
1	L	288	Total	C	N	O	S	0	0	0
			2359	1515	409	425	10			

- Molecule 2 is a protein called CRISPR-associated endoribonuclease Cas2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	104	Total	C	N	O	S	0	0	0
			862	548	155	153	6			
2	F	108	Total	C	N	O	S	0	0	0
			899	571	162	159	7			
2	M	108	Total	C	N	O	S	0	0	0
			899	571	162	159	7			
2	N	104	Total	C	N	O	S	0	0	0
			862	548	155	153	6			

- Molecule 3 is a DNA chain called DNA (28-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	25	Total	C	N	O	P	0	0	0
			511	243	90	153	25			
3	H	24	Total	C	N	O	P	0	0	0
			491	233	88	146	24			
3	O	26	Total	C	N	O	P	0	0	0
			530	252	93	159	26			
3	P	24	Total	C	N	O	P	0	0	0
			491	233	88	146	24			

- Molecule 4 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	1	Total	Mg	0	0
			1	1		
4	F	1	Total	Mg	0	0
			1	1		
4	M	1	Total	Mg	0	0
			1	1		
4	N	1	Total	Mg	0	0
			1	1		

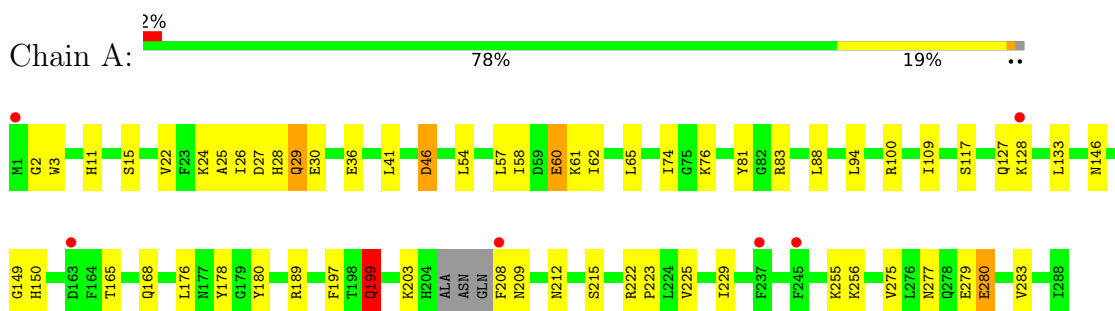
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	1	Total	O	0	0
			1	1		
5	L	1	Total	O	0	0
			1	1		

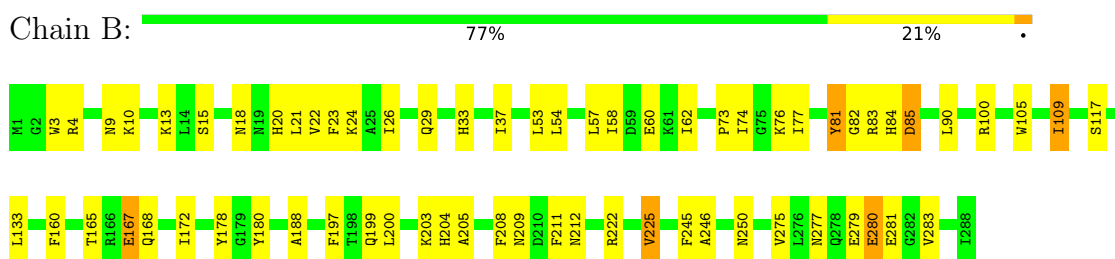
3 Residue-property plots [i](#)

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.

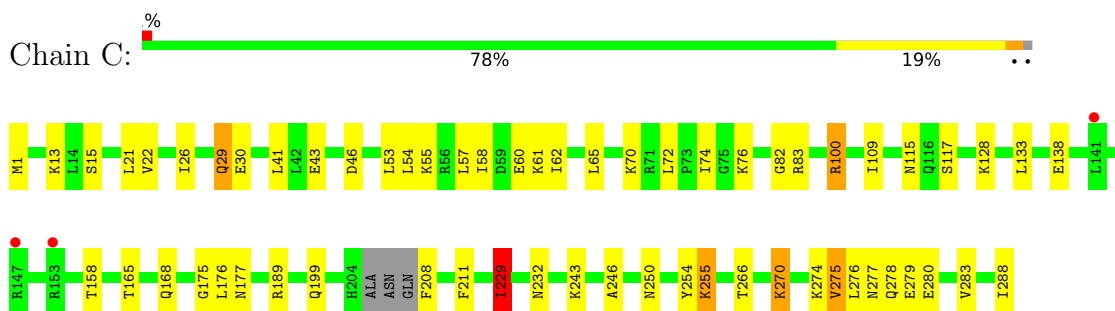
• Molecule 1: CRISPR-associated endonuclease Cas1



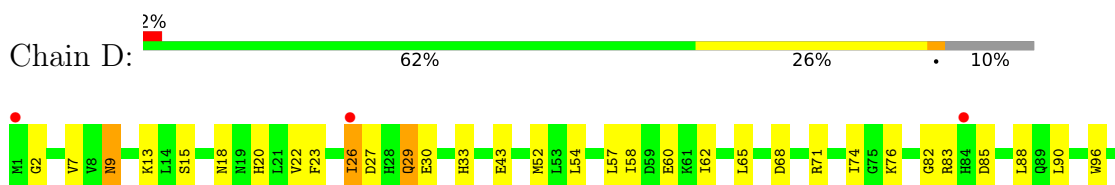
• Molecule 1: CRISPR-associated endonuclease Cas1

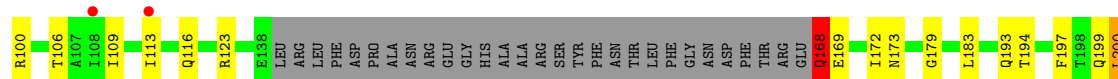


• Molecule 1: CRISPR-associated endonuclease Cas1

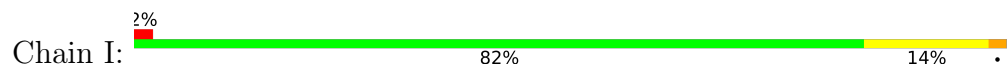


• Molecule 1: CRISPR-associated endonuclease Cas1

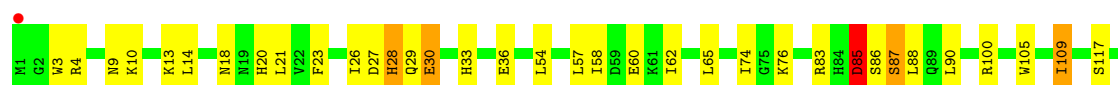




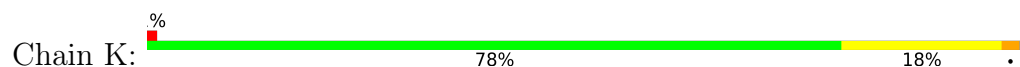
● Molecule 1: CRISPR-associated endonuclease Cas1



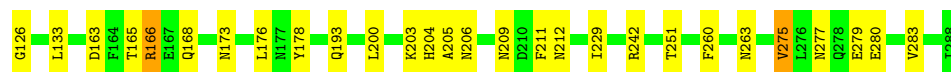
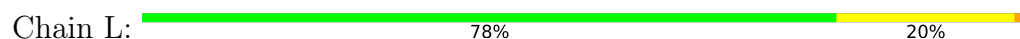
● Molecule 1: CRISPR-associated endonuclease Cas1



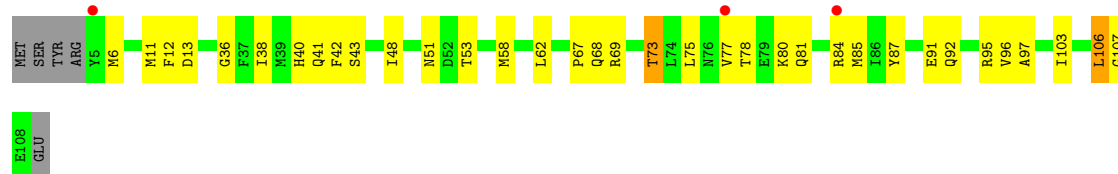
● Molecule 1: CRISPR-associated endonuclease Cas1



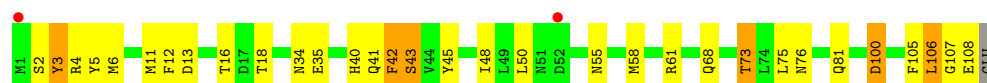
● Molecule 1: CRISPR-associated endonuclease Cas1



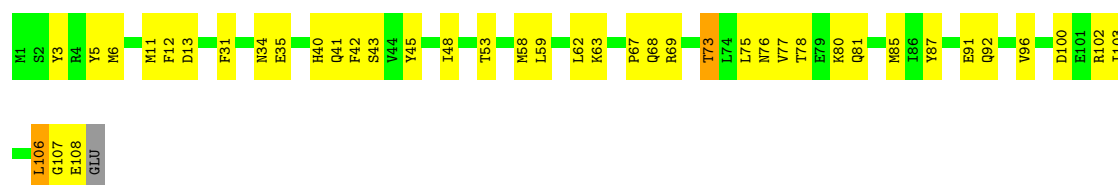
● Molecule 2: CRISPR-associated endoribonuclease Cas2



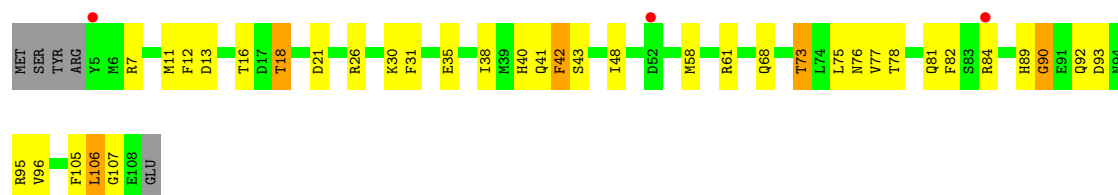
- Molecule 2: CRISPR-associated endoribonuclease Cas2



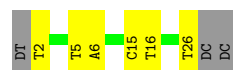
- Molecule 2: CRISPR-associated endoribonuclease Cas2



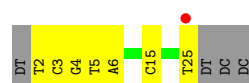
- Molecule 2: CRISPR-associated endoribonuclease Cas2



- Molecule 3: DNA (28-MER)



- Molecule 3: DNA (28-MER)



● Molecule 3: DNA (28-MER)



● Molecule 3: DNA (28-MER)



4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α , β , γ	160.71Å 160.71Å 187.97Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.41 – 3.25 49.41 – 3.25	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.41-3.25) 99.9 (49.41-3.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.38 (at 3.25Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.206 , 0.249 0.211 , 0.251	Depositor DCC
R_{free} test set	3714 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å ²)	80.3	Xtriage
Anisotropy	0.122	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 46.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.027 for h,-k,-l	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	24089	wwPDB-VP
Average B, all atoms (Å ²)	85.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 12.29% 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	0.68	0/2384	0.99	6/3214 (0.2%)
1	B	0.72	0/2407	1.01	5/3247 (0.2%)
1	C	0.67	0/2384	0.99	6/3214 (0.2%)
1	D	0.73	0/2153	1.06	9/2903 (0.3%)
1	I	0.68	1/2384 (0.0%)	1.00	8/3214 (0.2%)
1	J	0.70	0/2407	1.06	11/3247 (0.3%)
1	K	0.72	0/2384	0.98	5/3214 (0.2%)
1	L	0.77	1/2407 (0.0%)	1.03	5/3247 (0.2%)
2	E	0.79	0/876	1.04	1/1173 (0.1%)
2	F	0.83	1/914 (0.1%)	1.10	3/1223 (0.2%)
2	M	0.86	0/914	1.07	2/1223 (0.2%)
2	N	0.83	0/876	1.13	2/1173 (0.2%)
3	G	0.44	0/571	0.83	0/879
3	H	0.44	0/549	0.80	0/845
3	O	0.46	0/592	0.82	0/911
3	P	0.45	0/549	0.81	0/845
All	All	0.71	3/24751 (0.0%)	1.01	63/33772 (0.2%)

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	B	0	1
1	D	0	1
1	I	0	1
1	J	0	1
All	All	0	4

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	L	166	ARG	CA-C	6.49	1.60	1.52
1	I	25	ALA	CA-C	-5.91	1.46	1.53
2	F	3	TYR	N-CA	5.76	1.53	1.46

The worst 5 of 63 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	J	168	GLN	N-CA-C	14.67	129.94	109.11
1	I	168	GLN	N-CA-C	11.38	122.79	108.45
1	B	281	GLU	N-CA-C	9.99	126.69	113.30
1	J	281	GLU	N-CA-C	9.95	126.63	113.30
2	N	42	PHE	N-CA-C	9.82	122.89	111.11

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	81	TYR	Peptide
1	D	168	GLN	Peptide
1	I	168	GLN	Peptide
1	J	85	ASP	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	2337	0	2366	65	0
1	B	2359	0	2386	66	0
1	C	2337	0	2365	67	1
1	D	2113	0	2155	104	0
1	I	2337	0	2366	50	0
1	J	2359	0	2386	70	0
1	K	2337	0	2365	52	1
1	L	2359	0	2386	53	0
2	E	862	0	876	58	0
2	F	899	0	915	49	0
2	M	899	0	915	52	0
2	N	862	0	877	42	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	G	511	0	283	17	0
3	H	491	0	271	15	0
3	O	530	0	294	17	0
3	P	491	0	271	7	0
4	E	1	0	0	0	0
4	F	1	0	0	0	0
4	M	1	0	0	0	0
4	N	1	0	0	0	0
5	D	1	0	0	0	0
5	L	1	0	0	2	0
All	All	24089	0	23477	649	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 14.

The worst 5 of 649 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:D:237:PHE:CE2	1:D:241:LYS:HE3	1.40	1.55
2:F:6:MET:HE1	2:F:55:ASN:ND2	1.35	1.42
1:D:237:PHE:CE2	1:D:241:LYS:CE	2.10	1.34
1:A:61:LYS:NZ	1:A:81:TYR:O	1.70	1.22
1:D:229:ILE:HA	1:D:240:MET:HE1	1.20	1.18

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:C:138:GLU:OE1	1:K:127:GLN:OE1[1_545]	2.13	0.07

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	281/288 (98%)	268 (95%)	13 (5%)	0	100	100
1	B	286/288 (99%)	276 (96%)	10 (4%)	0	100	100
1	C	281/288 (98%)	266 (95%)	15 (5%)	0	100	100
1	D	255/288 (88%)	242 (95%)	13 (5%)	0	100	100
1	I	281/288 (98%)	269 (96%)	12 (4%)	0	100	100
1	J	286/288 (99%)	272 (95%)	14 (5%)	0	100	100
1	K	281/288 (98%)	269 (96%)	12 (4%)	0	100	100
1	L	286/288 (99%)	275 (96%)	11 (4%)	0	100	100
2	E	102/109 (94%)	96 (94%)	6 (6%)	0	100	100
2	F	106/109 (97%)	104 (98%)	2 (2%)	0	100	100
2	M	106/109 (97%)	101 (95%)	5 (5%)	0	100	100
2	N	102/109 (94%)	101 (99%)	1 (1%)	0	100	100
All	All	2653/2740 (97%)	2539 (96%)	114 (4%)	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 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	253/255 (99%)	245 (97%)	8 (3%)	34	58
1	B	255/255 (100%)	249 (98%)	6 (2%)	43	64
1	C	253/255 (99%)	246 (97%)	7 (3%)	38	60
1	D	230/255 (90%)	222 (96%)	8 (4%)	32	56
1	I	253/255 (99%)	240 (95%)	13 (5%)	21	49
1	J	255/255 (100%)	245 (96%)	10 (4%)	28	54
1	K	253/255 (99%)	242 (96%)	11 (4%)	26	52
1	L	255/255 (100%)	246 (96%)	9 (4%)	32	56
2	E	94/99 (95%)	89 (95%)	5 (5%)	20	48

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	F	98/99 (99%)	91 (93%)	7 (7%)	13	39
2	M	98/99 (99%)	93 (95%)	5 (5%)	21	49
2	N	94/99 (95%)	88 (94%)	6 (6%)	16	42
All	All	2391/2436 (98%)	2296 (96%)	95 (4%)	28	54

5 of 95 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	J	87	SER
1	K	229	ILE
1	J	167	GLU
1	K	30	GLU
1	L	7	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 75 such sidechains are listed below:

Mol	Chain	Res	Type
1	K	173	ASN
2	N	41	GLN
1	K	232	ASN
1	L	209	ASN
1	C	263	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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	285/288 (98%)	0.06	6 (2%) 63 44	54, 82, 120, 147	0
1	B	288/288 (100%)	-0.17	0 100 100	53, 75, 107, 152	0
1	C	285/288 (98%)	0.24	3 (1%) 78 60	74, 99, 129, 173	0
1	D	259/288 (89%)	0.36	7 (2%) 56 38	66, 118, 156, 185	0
1	I	285/288 (98%)	0.13	6 (2%) 63 44	54, 91, 120, 138	0
1	J	288/288 (100%)	-0.25	1 (0%) 90 82	53, 76, 105, 170	0
1	K	285/288 (98%)	0.00	3 (1%) 78 60	49, 80, 118, 137	0
1	L	288/288 (100%)	-0.29	0 100 100	42, 64, 93, 157	0
2	E	104/109 (95%)	-0.05	3 (2%) 53 36	54, 77, 101, 124	0
2	F	108/109 (99%)	0.07	2 (1%) 66 48	35, 76, 109, 117	0
2	M	108/109 (99%)	-0.18	0 100 100	40, 63, 97, 120	0
2	N	104/109 (95%)	-0.17	3 (2%) 53 36	34, 62, 85, 115	0
3	G	25/28 (89%)	-0.47	0 100 100	57, 88, 105, 127	0
3	H	24/28 (85%)	-0.41	1 (4%) 40 26	62, 85, 120, 132	0
3	O	26/28 (92%)	-0.27	1 (3%) 44 29	50, 77, 141, 155	0
3	P	24/28 (85%)	-0.52	0 100 100	43, 71, 101, 127	0
All	All	2786/2852 (97%)	-0.02	36 (1%) 75 56	34, 82, 128, 185	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	N	5	TYR	5.1
2	E	5	TYR	3.9
2	F	52	ASP	3.9
1	I	208	PHE	3.3
1	A	1	MET	3.3

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](#)

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. 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	B-factors($\text{\AA}^2$)	Q<0.9
4	MG	E	201	1/1	0.93	0.10	45,45,45,45	0
4	MG	F	201	1/1	0.95	0.08	43,43,43,43	0
4	MG	N	201	1/1	0.95	0.14	31,31,31,31	0
4	MG	M	201	1/1	0.96	0.14	51,51,51,51	0

6.5 Other polymers [i](#)

There are no such residues in this entry.