



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:30 PM GMT

PDB ID : 4IJS
Title : Crystal structure of nucleocapsid protein encoded by the prototypic member of orthobunyavirus
Authors : Li, B.B.; Wang, Q.; Lou, Z.Y.
Deposited on : 2012-12-23
Resolution : 3.20 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
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
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) : trunk26865

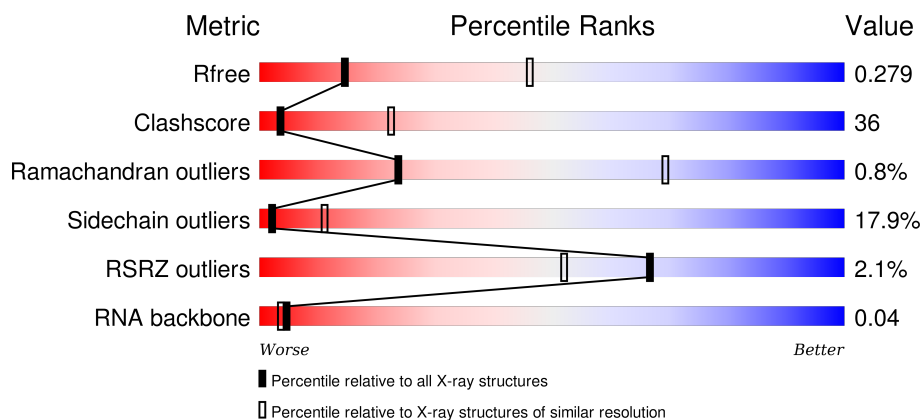
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.20 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}	91344	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)
RNA backbone	2183	1079 (3.70-2.70)

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	248	
1	B	248	
1	C	248	
1	D	248	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	10	<div><div></div><div></div><div></div><div></div></div> <div>10%50%30%10%</div>
2	F	10	<div><div></div><div></div><div></div><div></div></div> <div>20%50%20%10%</div>
2	G	10	<div><div></div><div></div><div></div><div></div></div> <div>10%50%30%10%</div>
2	H	10	<div><div></div><div></div><div></div><div></div></div> <div>20%70%10%</div>

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 8298 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 Nucleoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	232	Total	C	N	O	S	0	0	0
			1865	1205	315	339	6			
1	B	232	Total	C	N	O	S	0	0	0
			1865	1205	315	339	6			
1	C	232	Total	C	N	O	S	0	0	0
			1865	1205	315	339	6			
1	D	232	Total	C	N	O	S	0	0	0
			1865	1205	315	339	6			

There are 68 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-14	MET	-	EXPRESSION TAG	UNP P16495
A	-13	ARG	-	EXPRESSION TAG	UNP P16495
A	-12	GLY	-	EXPRESSION TAG	UNP P16495
A	-11	SER	-	EXPRESSION TAG	UNP P16495
A	-10	HIS	-	EXPRESSION TAG	UNP P16495
A	-9	HIS	-	EXPRESSION TAG	UNP P16495
A	-8	HIS	-	EXPRESSION TAG	UNP P16495
A	-7	HIS	-	EXPRESSION TAG	UNP P16495
A	-6	HIS	-	EXPRESSION TAG	UNP P16495
A	-5	HIS	-	EXPRESSION TAG	UNP P16495
A	-4	GLY	-	EXPRESSION TAG	UNP P16495
A	-3	SER	-	EXPRESSION TAG	UNP P16495
A	-2	ILE	-	EXPRESSION TAG	UNP P16495
A	-1	GLU	-	EXPRESSION TAG	UNP P16495
A	0	GLY	-	EXPRESSION TAG	UNP P16495
A	1	ARG	-	EXPRESSION TAG	UNP P16495
A	83	SER	ASN	ENGINEERED MUTATION	UNP P16495
B	-14	MET	-	EXPRESSION TAG	UNP P16495
B	-13	ARG	-	EXPRESSION TAG	UNP P16495
B	-12	GLY	-	EXPRESSION TAG	UNP P16495
B	-11	SER	-	EXPRESSION TAG	UNP P16495

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-10	HIS	-	EXPRESSION TAG	UNP P16495
B	-9	HIS	-	EXPRESSION TAG	UNP P16495
B	-8	HIS	-	EXPRESSION TAG	UNP P16495
B	-7	HIS	-	EXPRESSION TAG	UNP P16495
B	-6	HIS	-	EXPRESSION TAG	UNP P16495
B	-5	HIS	-	EXPRESSION TAG	UNP P16495
B	-4	GLY	-	EXPRESSION TAG	UNP P16495
B	-3	SER	-	EXPRESSION TAG	UNP P16495
B	-2	ILE	-	EXPRESSION TAG	UNP P16495
B	-1	GLU	-	EXPRESSION TAG	UNP P16495
B	0	GLY	-	EXPRESSION TAG	UNP P16495
B	1	ARG	-	EXPRESSION TAG	UNP P16495
B	83	SER	ASN	ENGINEERED MUTATION	UNP P16495
C	-14	MET	-	EXPRESSION TAG	UNP P16495
C	-13	ARG	-	EXPRESSION TAG	UNP P16495
C	-12	GLY	-	EXPRESSION TAG	UNP P16495
C	-11	SER	-	EXPRESSION TAG	UNP P16495
C	-10	HIS	-	EXPRESSION TAG	UNP P16495
C	-9	HIS	-	EXPRESSION TAG	UNP P16495
C	-8	HIS	-	EXPRESSION TAG	UNP P16495
C	-7	HIS	-	EXPRESSION TAG	UNP P16495
C	-6	HIS	-	EXPRESSION TAG	UNP P16495
C	-5	HIS	-	EXPRESSION TAG	UNP P16495
C	-4	GLY	-	EXPRESSION TAG	UNP P16495
C	-3	SER	-	EXPRESSION TAG	UNP P16495
C	-2	ILE	-	EXPRESSION TAG	UNP P16495
C	-1	GLU	-	EXPRESSION TAG	UNP P16495
C	0	GLY	-	EXPRESSION TAG	UNP P16495
C	1	ARG	-	EXPRESSION TAG	UNP P16495
C	83	SER	ASN	ENGINEERED MUTATION	UNP P16495
D	-14	MET	-	EXPRESSION TAG	UNP P16495
D	-13	ARG	-	EXPRESSION TAG	UNP P16495
D	-12	GLY	-	EXPRESSION TAG	UNP P16495
D	-11	SER	-	EXPRESSION TAG	UNP P16495
D	-10	HIS	-	EXPRESSION TAG	UNP P16495
D	-9	HIS	-	EXPRESSION TAG	UNP P16495
D	-8	HIS	-	EXPRESSION TAG	UNP P16495
D	-7	HIS	-	EXPRESSION TAG	UNP P16495
D	-6	HIS	-	EXPRESSION TAG	UNP P16495
D	-5	HIS	-	EXPRESSION TAG	UNP P16495
D	-4	GLY	-	EXPRESSION TAG	UNP P16495
D	-3	SER	-	EXPRESSION TAG	UNP P16495

Continued on next page...

Continued from previous page...

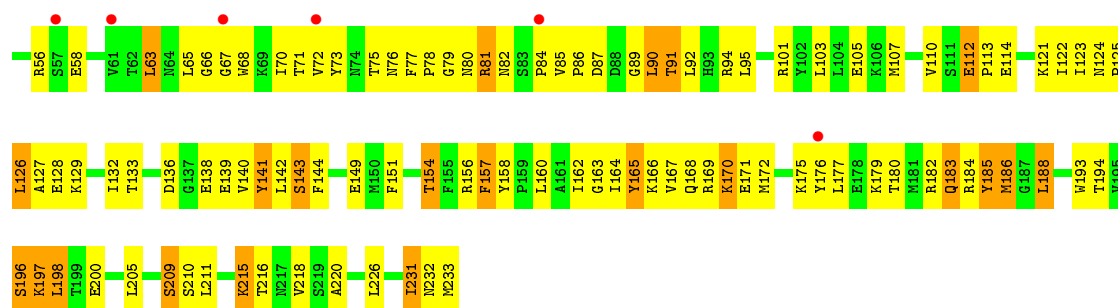
Chain	Residue	Modelled	Actual	Comment	Reference
D	-2	ILE	-	EXPRESSION TAG	UNP P16495
D	-1	GLU	-	EXPRESSION TAG	UNP P16495
D	0	GLY	-	EXPRESSION TAG	UNP P16495
D	1	ARG	-	EXPRESSION TAG	UNP P16495
D	83	SER	ASN	ENGINEERED MUTATION	UNP P16495

- Molecule 2 is a RNA chain called RNA (5'-R(P*AP*AP*AP*AP*AP*AP*AP*AP*A)-3').

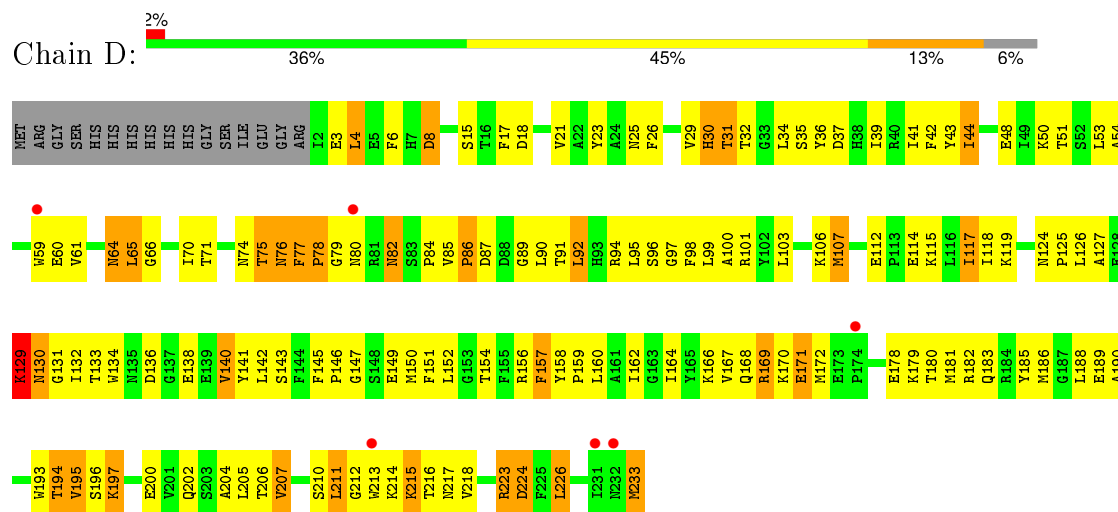
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	10	Total	C	N	O	P	0	0	0
			220	100	50	60	10			
2	F	9	Total	C	N	O	P	0	0	0
			198	90	45	54	9			
2	G	9	Total	C	N	O	P	0	0	0
			198	90	45	54	9			
2	H	9	Total	C	N	O	P	0	0	0
			198	90	45	54	9			

- Molecule 3 is water.

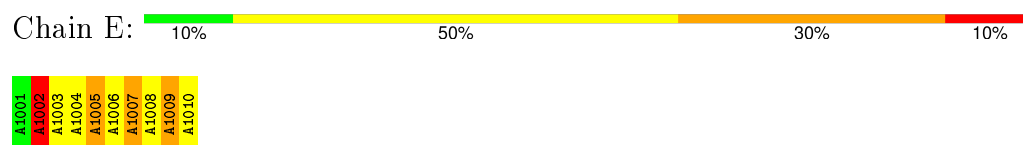
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	10	Total	O	0	0
			10	10		
3	B	6	Total	O	0	0
			6	6		
3	C	3	Total	O	0	0
			3	3		
3	D	2	Total	O	0	0
			2	2		
3	E	1	Total	O	0	0
			1	1		
3	H	2	Total	O	0	0
			2	2		



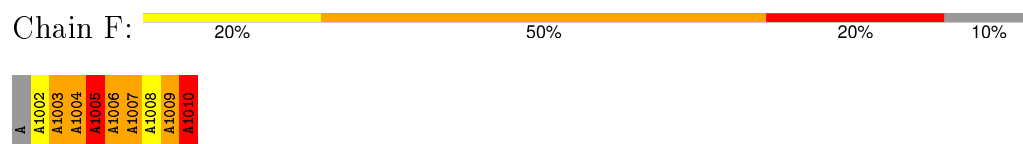
• Molecule 1: Nucleoprotein



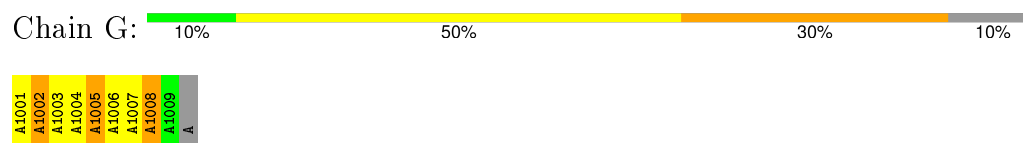
• Molecule 2: RNA (5'-R(P*AP*AP*AP*AP*AP*AP*AP*AP*AP*A)-3')



• Molecule 2: RNA (5'-R(P*AP*AP*AP*AP*AP*AP*AP*AP*AP*A)-3')



• Molecule 2: RNA (5'-R(P*AP*AP*AP*AP*AP*AP*AP*AP*AP*A)-3')



• Molecule 2: RNA (5'-R(P*AP*AP*AP*AP*AP*AP*AP*AP*AP*A)-3')



A1001
A1002
A1003
A1004
A1005
A1006
A1007
A1008
A1009
A

4 Data and refinement statistics

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, α , β , γ	106.14Å 106.14Å 485.19Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.75 – 3.20 47.17 – 3.19	Depositor EDS
% Data completeness (in resolution range)	90.4 (40.75-3.20) 97.5 (47.17-3.19)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.76 (at 3.19Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.220 , 0.276 0.248 , 0.279	Depositor DCC
R_{free} test set	1189 reflections (5.16%)	DCC
Wilson B-factor (Å ²)	93.4	Xtriage
Anisotropy	0.500	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 70.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 23462 reflections	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	8298	wwPDB-VP
Average B, all atoms (Å ²)	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.35% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.80	1/1909 (0.1%)	0.89	7/2581 (0.3%)
1	B	0.75	1/1909 (0.1%)	0.85	6/2581 (0.2%)
1	C	0.60	0/1909	0.77	3/2581 (0.1%)
1	D	0.75	1/1909 (0.1%)	0.86	5/2581 (0.2%)
2	E	1.15	2/249 (0.8%)	1.57	6/386 (1.6%)
2	F	1.01	2/224 (0.9%)	1.52	1/347 (0.3%)
2	G	0.93	1/224 (0.4%)	1.09	1/347 (0.3%)
2	H	0.65	0/224	1.09	0/347
All	All	0.76	8/8557 (0.1%)	0.92	29/11751 (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	A	0	2
1	B	0	1
1	C	0	1
All	All	0	4

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	G	1005	A	N9-C4	7.59	1.42	1.37
2	E	1005	A	N9-C4	6.75	1.41	1.37
2	E	1005	A	N3-C4	6.17	1.38	1.34
1	D	78	PRO	N-CD	5.59	1.55	1.47
1	B	86	PRO	N-CD	5.44	1.55	1.47

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	68	TRP	N-CA-C	-8.59	87.81	111.00
2	E	1007	A	O5'-P-OP1	-8.52	98.04	105.70
1	A	34	LEU	CB-CA-C	7.86	125.12	110.20
1	C	215	LYS	CB-CA-C	-6.93	96.54	110.40
1	A	34	LEU	N-CA-C	-6.68	92.95	111.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	33	GLY	Peptide
1	A	65	LEU	Peptide
1	B	82	ASN	Peptide
1	C	12	ASN	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	1865	0	1878	136	0
1	B	1865	0	1878	148	1
1	C	1865	0	1878	136	0
1	D	1865	0	1878	149	0
2	E	220	0	111	2	0
2	F	198	0	100	16	0
2	G	198	0	100	15	0
2	H	198	0	100	8	0
3	A	10	0	0	0	0
3	B	6	0	0	1	0
3	C	3	0	0	0	0
3	D	2	0	0	0	0
3	E	1	0	0	0	0
3	H	2	0	0	0	0
All	All	8298	0	7923	583	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 36.

The worst 5 of 583 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:C:157:PHE:CE2	1:C:185:TYR:CE1	1.94	1.55
1:C:157:PHE:CZ	1:C:185:TYR:CD1	2.01	1.47
1:D:169:ARG:NH2	1:D:215:LYS:HG2	1.08	1.39
1:D:169:ARG:NH2	1:D:215:LYS:CG	1.82	1.38
1:C:157:PHE:CE2	1:C:185:TYR:HE1	1.35	1.35

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:B:174:PRO:O	1:B:233:MET:CE[3_545]	1.93	0.27

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	230/248 (93%)	203 (88%)	24 (10%)	3 (1%)	15	59
1	B	230/248 (93%)	203 (88%)	26 (11%)	1 (0%)	39	80
1	C	230/248 (93%)	205 (89%)	23 (10%)	2 (1%)	21	67
1	D	230/248 (93%)	194 (84%)	35 (15%)	1 (0%)	39	80
All	All	920/992 (93%)	805 (88%)	108 (12%)	7 (1%)	24	69

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	82	ASN
1	B	169	ARG
1	C	78	PRO
1	A	228	LYS
1	C	220	ALA

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	200/215 (93%)	174 (87%)	26 (13%)	5	24
1	B	200/215 (93%)	167 (84%)	33 (16%)	3	13
1	C	200/215 (93%)	155 (78%)	45 (22%)	1	5
1	D	200/215 (93%)	161 (80%)	39 (20%)	2	9
All	All	800/860 (93%)	657 (82%)	143 (18%)	2	11

5 of 143 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	20	GLU
1	C	126	LEU
1	D	197	LYS
1	C	31	THR
1	C	81	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	82	ASN
1	B	12	ASN
1	B	64	ASN
1	C	76	ASN
1	C	232	ASN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	E	9/10 (90%)	9 (100%)	1 (11%)
2	F	8/10 (80%)	7 (87%)	2 (25%)
2	G	8/10 (80%)	5 (62%)	0
2	H	8/10 (80%)	6 (75%)	1 (12%)

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
All	All	33/40 (82%)	27 (81%)	4 (12%)

5 of 27 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	E	1002	A
2	E	1003	A
2	E	1004	A
2	E	1005	A
2	E	1006	A

All (4) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	E	1002	A
2	F	1005	A
2	F	1006	A
2	H	1002	A

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](#)

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 ⓘ

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	232/248 (93%)	-0.04	1 (0%) 93 90	26, 50, 78, 101	0
1	B	232/248 (93%)	0.06	6 (2%) 59 45	25, 62, 90, 122	0
1	C	232/248 (93%)	0.07	7 (3%) 54 39	23, 76, 104, 127	0
1	D	232/248 (93%)	0.10	6 (2%) 59 45	27, 69, 103, 125	0
2	E	10/10 (100%)	0.19	0 100 100	52, 75, 84, 95	0
2	F	9/10 (90%)	0.28	0 100 100	75, 86, 96, 106	0
2	G	9/10 (90%)	0.04	0 100 100	86, 96, 108, 111	0
2	H	9/10 (90%)	0.26	0 100 100	83, 100, 114, 120	0
All	All	965/1032 (93%)	0.05	20 (2%) 67 52	23, 66, 101, 127	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	67	GLY	4.3
1	D	213	TRP	3.5
1	B	12	ASN	3.4
1	D	232	ASN	3.3
1	B	86	PRO	3.0

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

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

6.3 Carbohydrates ⓘ

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.