wwPDB X-ray Structure Validation Summary Report (i)

Dec 23, 2024 – 04:12 PM JST

PDB ID	:	8Y0C
Title	:	Crystal structure of FnCas12a in complex with pre-crRNA and 18nt target DNA
Authors	:	Chen, J.; Liu, L.
Deposited on	:	2024-01-22
Resolution	:	3.45 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 (i) 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 (1)) were used in the production of this report:

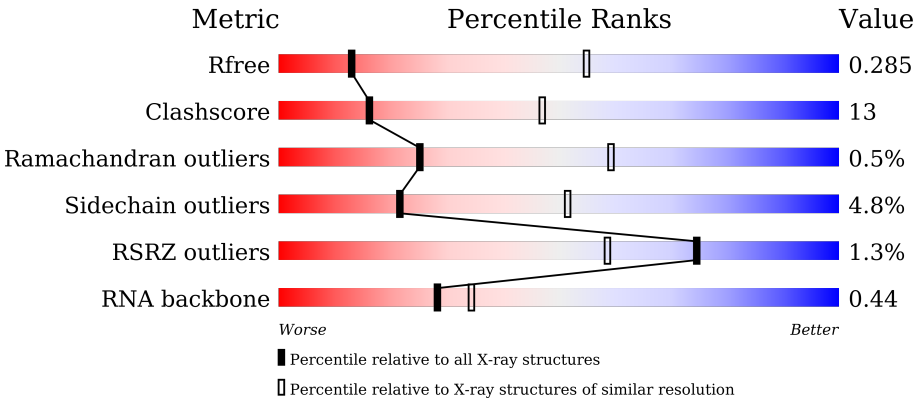
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.45 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}	164625	1597 (3.52-3.40)
Clashscore	180529	1041 (3.50-3.42)
Ramachandran outliers	177936	1026 (3.50-3.42)
Sidechain outliers	177891	1027 (3.50-3.42)
RSRZ outliers	164620	1596 (3.52-3.40)
RNA backbone	3690	1057 (3.92-3.00)

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 <=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	B	58	17% 38% 17% 28%
2	C	27	67% 30% .
3	D	11	55% 27% 18%
4	A	1300	% 64% 31% . .

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 12057 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 (42-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	B	42	890	400	155	294	41	0	0	0

- Molecule 2 is a DNA chain called DNA (27-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	27	550	265	95	163	27	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(P*AP*GP*TP*CP*CP*TP*TP*TP*AP*CP *T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	D	11	221	107	34	69	11	0	0	0

- Molecule 4 is a protein called CRISPR-associated endonuclease Cas12a.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	A	1260	10392	6681	1711	1978	22	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1006	ALA	GLU	conflict	UNP A0Q7Q2

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	2	Total Mg 2 2	0	0

- Molecule 6 is water.

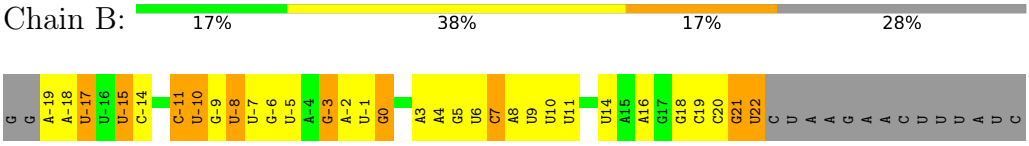
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	2	Total O 2 2	0	0

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.

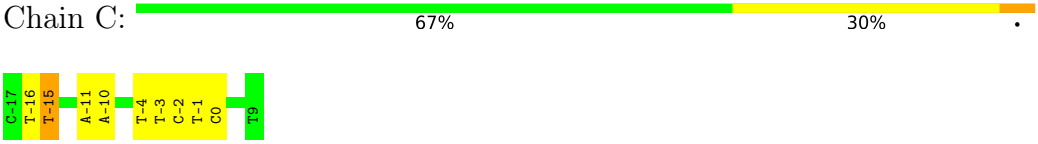
• Molecule 1: RNA (42-MER)

Chain B:



• Molecule 2: DNA (27-MER)

Chain C:

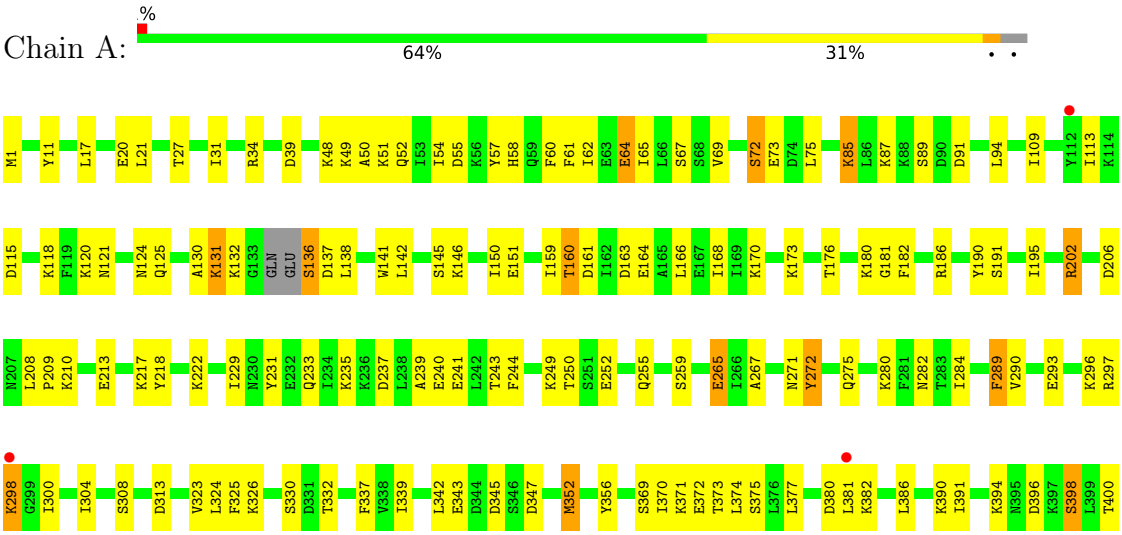


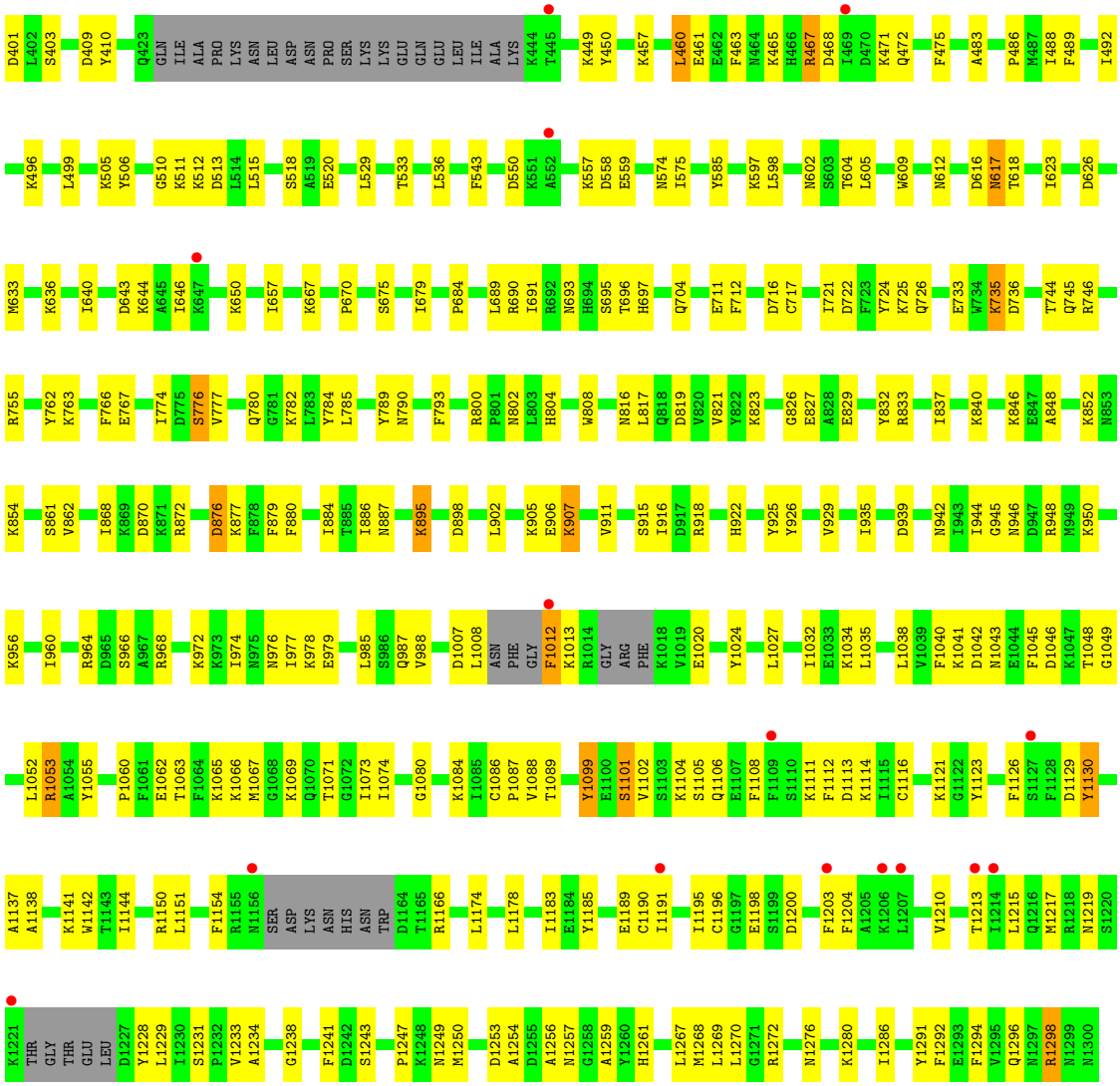
• Molecule 3: DNA (5'-D(P*AP*GP*TP*CP*CP*TP*TP*TP*AP*CP*T)-3')

Chain D:

• Molecule 4: CRISPR-associated endonuclease Cas12a

Chain A:




D401 L402 S403 D409 Y410 Q423 GLN ILE ALA PRO LYS ASN LEU ASP ASN PRO SER LYS LYS GLU GLN GLU LEU ILE ALA LYS K444 T445 K449 Y450 K457 L460 E461 E462 F463 N464 K465 H466 R467 D468 I469 D470 K471 Q472 F475 A483 P486 M487 I488 F489 I492
K496 L499 K505 Y506 G510 K511 K512 D513 L514 L515 S518 A519 E520 L529 T533 L536 F543 D550 K551 A552 K557 D558 E559 N574 I575 Y585 K597 L598 N602 S603 T604 L605 W609 N612 D616 N617 T618 I623 D626
M633 K636 I640 D643 K644 A645 I646 K647 K650 I657 K667 P670 S675 I679 P684 L689 R690 I691 R692 N693 H694 S695 T696 H697 Q704 E711 F712 D716 C717 I721 D722 F723 Y724 K725 Q726 E733 W734 K735 D736 T744 Q745 R746
R755 Y762 K763 F766 E767 I774 D775 S776 V777 Q780 G781 K782 L783 Y784 L785 Y789 N790 F793 R800 P801 N802 L803 H804 W808 N816 L817 Q818 D819 V820 V821 Y822 K823 G826 E827 A828 E829 Y832 R833 I837 K840 K846 E847 A848 K852 N853
K854 S861 V862 I868 K869 D870 K871 R872 D876 K877 F878 F879 F880 I884 T885 I886 N887 K895 D898 L902 K905 E906 K907 V911 S915 I916 D917 R918 H922 Y925 Y926 V929 I935 D939 N942 I943 I944 G945 N946 D947 R948 M949 K950
K956 I960 R964 D965 S966 A967 R968 K972 K973 I974 N975 N976 I977 K978 E979 L985 S986 Q987 V988 D1007 L1008 ASN PHE GLY F1012 K1013 R1014 GLY ARG PHE K1018 V1019 E1020 Y1024 L1027 I1032 E1033 K1034 L1035 L1038 V1039 F1040 K1041 D1042 N1043 E1044 F1045 D1046 K1047 T1048 G1049
L1052 R1053 A1054 Y1055 P1060 F1061 E1062 T1063 F1064 K1065 K1066 M1067 G1068 K1069 Q1070 T1071 G1072 I1073 I1074 G1080 K1084 I1085 C1086 P1087 V1088 T1089 Y1099 E1100 S1101 V1102 S1103 K1104 S1105 Q1106 E1107 F1108 F1109 S1110 K1111 F1112 D1113 K1114 I1115 C1116 K1121 G1122 Y1123 F1126 S1127 F1128 D1129 Y1130
A1137 A1138 K1141 W1142 T1143 I1144 R1150 L1151 F1154 R1155 N1156 SER ASP LYS ASN HIS ASN TRP D1164 T1165 R1166 L1174 L1178 I1183 E1184 Y1185 E1189 C1190 I1191 I1195 C1196 G1197 E1198 S1199 D1200 F1203 F1204 A1205 K1206 L1207 V1210 T1213 I1214 L1215 Q1216 M1217 R1218 N1219 S1220
K1221 THR GLY THR GLU LEU D1227 Y1228 L1229 I1230 S1231 P1232 V1233 A1234 G1238 F1241 D1242 S1243 P1247 K1248 N1249 M1250 D1253 A1254 D1255 A1256 N1257 G1258 A1259 Y1260 H1261 L1267 M1268 L1269 L1270 G1271 R1272 N1276 K1280 I1286 Y1291 F1292 E1293 F1294 V1295 Q1296 N1297 R1298 N1299 N1300

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	124.28Å 124.28Å 268.52Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.22 – 3.45 47.22 – 3.45	Depositor EDS
% Data completeness (in resolution range)	88.9 (47.22-3.45) 88.8 (47.22-3.45)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.32 (at 3.48Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.216 , 0.285 0.217 , 0.285	Depositor DCC
R_{free} test set	1394 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å^2)	48.2	Xtriage
Anisotropy	0.077	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 40.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.39$, $< L^2 > = 0.21$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	12057	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.45% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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	B	0.89	0/995	1.80	36/1548 (2.3%)
2	C	1.15	1/615 (0.2%)	1.17	2/946 (0.2%)
3	D	1.15	1/245 (0.4%)	1.25	1/375 (0.3%)
4	A	0.55	0/10595	0.71	3/14220 (0.0%)
All	All	0.64	2/12450 (0.0%)	0.90	42/17089 (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
4	A	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	-3	DT	C3'-O3'	-7.79	1.33	1.44
3	D	0	DC	C3'-O3'	-5.93	1.36	1.44

The worst 5 of 42 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	20	C	C2-N1-C1'	9.34	129.07	118.80
1	B	-10	U	N3-C2-O2	-8.90	115.97	122.20
1	B	4	A	C8-N9-C4	8.30	109.12	105.80
1	B	-11	C	C6-N1-C2	-8.05	117.08	120.30
1	B	-19	A	N1-C6-N6	-7.47	114.12	118.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	A	789	TYR	Peptide

5.2 Too-close contacts (i)

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	B	890	0	449	17	0
2	C	550	0	308	4	0
3	D	221	0	127	7	0
4	A	10392	0	10350	289	1
5	B	2	0	0	0	0
6	A	2	0	0	1	0
All	All	12057	0	11234	301	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 13.

The worst 5 of 301 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)
4:A:371:LYS:HE2	4:A:489:PHE:HB3	1.47	0.93
1:B:19:C:N3	1:B:21:G:N2	2.20	0.89
4:A:518:SER:O	6:A:1401:HOH:O	1.95	0.85
4:A:960:ILE:HG22	4:A:977:ILE:HD12	1.63	0.81
2:C:0:DC:H4'	4:A:827:GLU:HG3	1.63	0.81

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)
4:A:72:SER:OG	4:A:1046:ASP:OD1[5_454]	2.11	0.09

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
4	A	1246/1300 (96%)	1185 (95%)	55 (4%)	6 (0%)	25 59

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	574	ASN
4	A	160	THR
4	A	626	ASP
4	A	1013	LYS
4	A	1137	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
4	A	1144/1183 (97%)	1089 (95%)	55 (5%)	21 51

5 of 55 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	A	746	ARG
4	A	854	LYS
4	A	1298	ARG
4	A	1130	TYR
4	A	755	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such

sidechains are listed below:

Mol	Chain	Res	Type
4	A	843	HIS
4	A	901	ASN
4	A	1043	ASN
4	A	617	ASN
4	A	77	GLN

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	B	41/58 (70%)	6 (14%)	2 (4%)

5 of 6 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	B	-10	U
1	B	-9	G
1	B	-7	U
1	B	-6	G
1	B	21	G

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	B	-10	U
1	B	0	G

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 2 ligands modelled in this entry, 2 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, 95^{th} 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	B	42/58 (72%)	-0.74	0	100	100	19, 27, 101, 139	0
2	C	27/27 (100%)	-0.63	0	100	100	21, 39, 70, 98	0
3	D	11/11 (100%)	-0.61	0	100	100	33, 38, 78, 95	0
4	A	1260/1300 (96%)	-0.20	18 (1%)	73	57	14, 44, 112, 152	0
All	All	1340/1396 (95%)	-0.23	18 (1%)	74	59	14, 43, 112, 152	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	1207	LEU	5.3
4	A	1203	PHE	3.7
4	A	1213	THR	2.9
4	A	1214	ILE	2.9
4	A	298	LYS	2.7

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 monosaccharides 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. The B-factors column lists the minimum, median, 95^{th} 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(Å^2)	Q<0.9
5	MG	B	101	1/1	0.95	0.28	2,2,2,2	0
5	MG	B	102	1/1	0.96	0.20	9,9,9,9	0

6.5 Other polymers (i)

There are no such residues in this entry.