Full wwPDB X-ray Structure Validation Report ⓘ

May 16, 2020 – 11:46 am BST

PDB ID	:	2YKG
Title	:	Structural insights into RNA recognition by RIG-I
Authors	:	Luo, D.; Pyle, A.M.
Deposited on	:	2011-05-27
Resolution	:	2.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

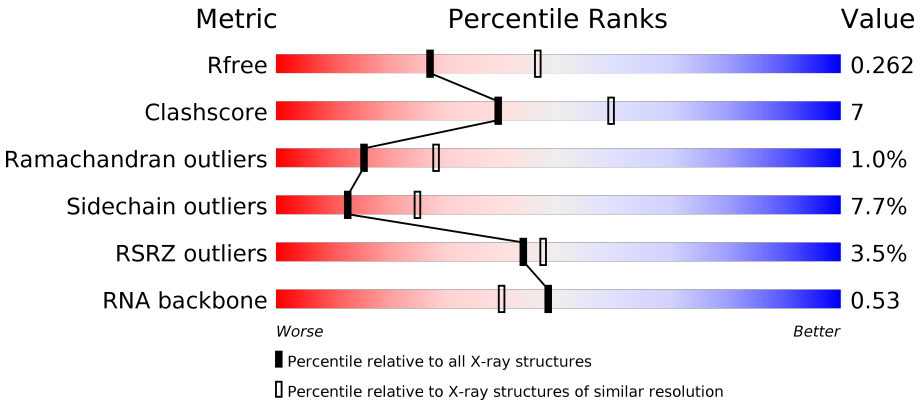
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)
RNA backbone	3102	1008 (2.84-2.16)

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 <=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	696	3% 73% 16% • 9%
2	C	10	70% 20% 10%
2	D	10	70% 30%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5523 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 PROBABLE ATP-DEPENDENT RNA HELICASE DDX58.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	636	Total	C	N	O	S	5	1	0
			4994	3201	846	917	30			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	306	ASN	GLN	conflict	UNP O95786
A	419	THR	ASN	conflict	UNP O95786
A	828	ASP	GLU	conflict	UNP O95786

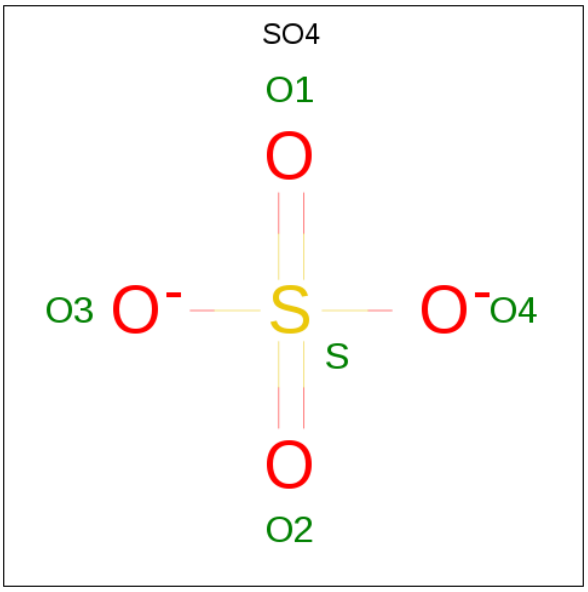
- Molecule 2 is a RNA chain called 5'-R(*GP*CP*GP*CP*GP*CP*GP*CP*GP*CP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	10	Total	C	N	O	P	0	0	0
			212	95	40	68	9			
2	D	10	Total	C	N	O	P	0	0	0
			212	95	40	68	9			

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Zn	0	0
			1	1		

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O S 5 4 1	0	0

- Molecule 5 is water.

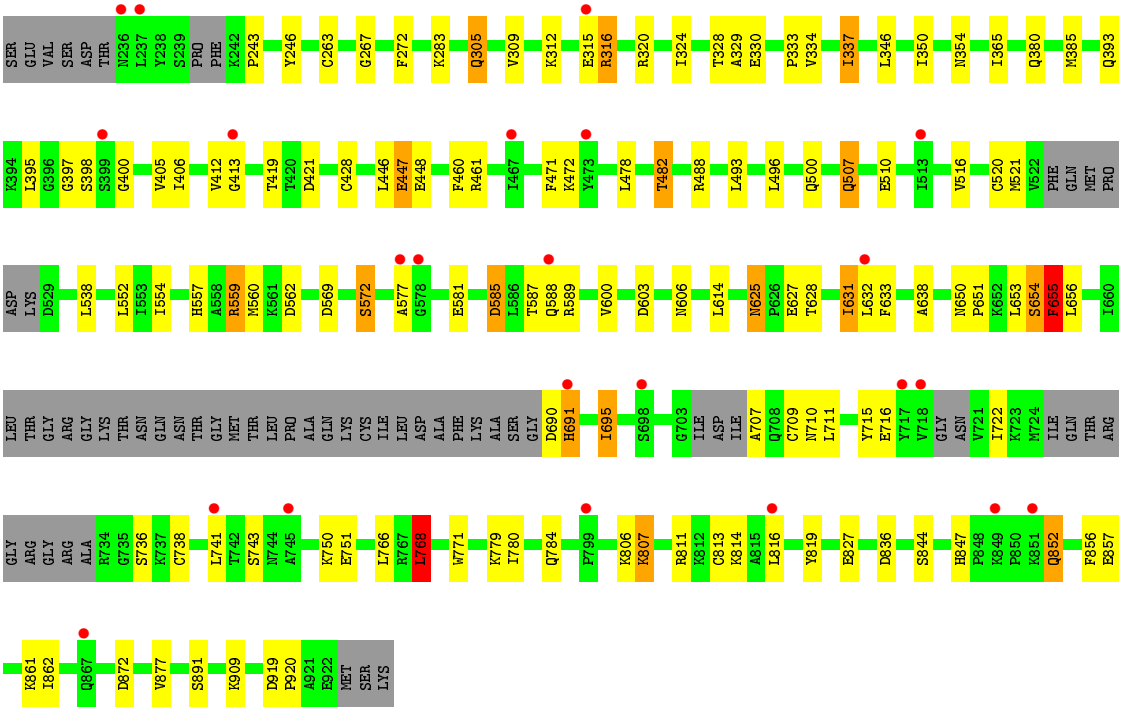
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	80	Total O 80 80	0	0
5	C	11	Total O 11 11	0	0
5	D	8	Total O 8 8	0	0

3 Residue-property plots (i)

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: PROBABLE ATP-DEPENDENT RNA HELICASE DDX58

Chain A:

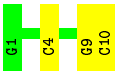


- Molecule 2: 5'-R(*GP*CP*GP*CP*GP*CP*GP*CP*GP*CP)-3'

Chain C:

- Molecule 2: 5'-R(*GP*CP*GP*CP*GP*CP*GP*CP*GP*CP)-3'

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	47.62Å 76.21Å 219.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.00 – 2.50 43.69 – 2.50	Depositor EDS
% Data completeness (in resolution range)	93.7 (45.00-2.50) 93.7 (43.69-2.50)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.78 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.220 , 0.273 0.214 , 0.262	Depositor DCC
R_{free} test set	1341 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	54.1	Xtriage
Anisotropy	0.097	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 50.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	5523	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.50% 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: ZN, SO4

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	0.77	1/5093 (0.0%)	0.69	4/6881 (0.1%)
2	C	0.86	0/236	1.30	3/367 (0.8%)
2	D	0.84	0/236	1.45	1/367 (0.3%)
All	All	0.77	1/5565 (0.0%)	0.78	8/7615 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	807	LYS	CE-NZ	-39.46	0.50	1.49

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	807	LYS	CD-CE-NZ	22.21	162.79	111.70
2	C	10	C	O4'-C1'-N1	7.00	113.80	108.20
1	A	768	LEU	CA-CB-CG	5.91	128.89	115.30
1	A	654	SER	N-CA-C	5.58	126.07	111.00
2	D	4	C	O4'-C1'-N1	5.42	112.53	108.20
2	C	2	C	O4'-C1'-N1	5.30	112.44	108.20
1	A	771	TRP	CA-CB-CG	5.13	123.44	113.70
2	C	3	G	O4'-C1'-N9	5.01	112.21	108.20

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	4994	0	4909	78	0
2	C	212	0	112	0	0
2	D	212	0	112	0	0
3	A	1	0	0	0	0
4	A	5	0	0	1	0
5	A	80	0	0	1	0
5	C	11	0	0	0	0
5	D	8	0	0	0	0
All	All	5523	0	5133	78	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 7.

All (78) 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:560:MET:HE3	1:A:600:VAL:HG12	1.41	1.00
1:A:654:SER:H	1:A:655:PHE:HB3	1.28	0.98
1:A:560:MET:CE	1:A:600:VAL:HG12	1.94	0.96
1:A:471:PHE:HB2	1:A:560:MET:HE3	1.46	0.95
1:A:654:SER:N	1:A:655:PHE:HB3	1.87	0.88
1:A:690:ASP:HA	1:A:691:HIS:CB	2.07	0.84
1:A:471:PHE:HB2	1:A:560:MET:CE	2.10	0.81
1:A:653:LEU:HB3	1:A:656:LEU:HD12	1.61	0.81
1:A:780:ILE:O	1:A:784:GLN:HG3	1.81	0.80
1:A:428:CYS:SG	1:A:779:LYS:NZ	2.54	0.79
1:A:557:HIS:CE1	1:A:716:GLU:HG2	2.18	0.78
1:A:654:SER:CA	1:A:655:PHE:HB3	2.17	0.73
1:A:461:ARG:HB2	1:A:741:LEU:HD23	1.71	0.73
1:A:585:ASP:O	1:A:589:ARG:HG3	1.89	0.72
1:A:460:PHE:HE1	1:A:751:GLU:HG2	1.55	0.71
1:A:560:MET:HE3	1:A:600:VAL:CG1	2.21	0.70
1:A:263:CYS:SG	5:A:2007:HOH:O	2.00	0.67
1:A:412:VAL:HG23	1:A:413:GLY:N	2.09	0.66
1:A:631:ILE:HD12	1:A:633:PHE:CE1	2.34	0.62
1:A:560:MET:HE1	1:A:600:VAL:HG12	1.80	0.61
1:A:656:LEU:HB3	1:A:695:ILE:HD11	1.82	0.61
1:A:460:PHE:CE1	1:A:751:GLU:HG2	2.36	0.59
1:A:246:TYR:HD2	1:A:448:GLU:OE1	1.86	0.59

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:654:SER:H	1:A:655:PHE:CB	2.09	0.58
1:A:559:ARG:HD3	1:A:562:ASP:OD2	2.04	0.57
1:A:447:GLU:H	1:A:447:GLU:CD	2.08	0.57
1:A:603:ASP:O	1:A:606:ASN:HB2	2.06	0.56
1:A:419:THR:HG22	1:A:421:ASP:H	1.69	0.56
1:A:496:LEU:HD23	1:A:516:VAL:HG13	1.88	0.54
1:A:844:SER:HB3	1:A:862:ILE:HG22	1.90	0.54
1:A:625:ASN:ND2	1:A:627:GLU:H	2.06	0.54
1:A:632:LEU:HD11	1:A:715:TYR:HB2	1.90	0.53
1:A:315:GLU:O	1:A:316:ARG:CB	2.57	0.53
1:A:329:ALA:HB1	1:A:330:GLU:HA	1.90	0.53
1:A:312:LYS:O	1:A:315:GLU:HG2	2.09	0.52
1:A:478:LEU:O	1:A:482:THR:CG2	2.57	0.52
1:A:690:ASP:CA	1:A:691:HIS:CB	2.84	0.52
1:A:334:VAL:HG11	1:A:354:ASN:OD1	2.10	0.52
1:A:807:LYS:HD2	1:A:816:LEU:HD13	1.92	0.52
1:A:813:CYS:O	1:A:814:LYS:HB2	2.10	0.50
1:A:412:VAL:CG2	1:A:413:GLY:N	2.74	0.49
1:A:569:ASP:O	1:A:572:SER:HB3	2.13	0.49
1:A:507:GLN:H	1:A:507:GLN:CD	2.15	0.49
1:A:554:ILE:HD11	1:A:638:ALA:HB1	1.95	0.48
1:A:267:GLY:N	4:A:1923:SO4:O1	2.33	0.47
1:A:333:PRO:O	1:A:337:ILE:HD12	2.15	0.47
1:A:768:LEU:C	1:A:768:LEU:HD12	2.35	0.47
1:A:460:PHE:CE1	1:A:751:GLU:CG	2.98	0.47
1:A:710:ASN:HA	1:A:736:SER:HB3	1.97	0.46
1:A:807:LYS:HG2	1:A:819:TYR:CE1	2.50	0.46
1:A:872:ASP:O	1:A:891:SER:OG	2.30	0.45
1:A:393:GLN:HG3	1:A:398:SER:HB2	1.99	0.45
1:A:628:THR:HA	1:A:710:ASN:HD21	1.81	0.45
1:A:380:GLN:HA	1:A:385:MET:HE3	1.99	0.45
1:A:521:MET:H	1:A:521:MET:HG3	1.59	0.44
1:A:631:ILE:HD13	1:A:632:LEU:N	2.33	0.44
1:A:632:LEU:HD11	1:A:715:TYR:CB	2.48	0.44
1:A:405:VAL:O	1:A:406:ILE:HD13	2.18	0.43
1:A:559:ARG:HH11	1:A:562:ASP:CG	2.22	0.43
1:A:478:LEU:O	1:A:482:THR:HG23	2.16	0.43
1:A:625:ASN:C	1:A:625:ASN:HD22	2.21	0.43
1:A:324:ILE:HB	1:A:346:LEU:CD2	2.48	0.43
1:A:243:PRO:HB3	1:A:272:PHE:HE2	1.83	0.43
1:A:560:MET:HE1	1:A:600:VAL:O	2.19	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:919:ASP:HA	1:A:920:PRO:HD2	1.86	0.43
1:A:650:ASN:HA	1:A:651:PRO:HD2	1.89	0.43
1:A:305:GLN:O	1:A:309:VAL:HG23	2.19	0.42
1:A:246:TYR:CD2	1:A:448:GLU:OE1	2.69	0.42
1:A:478:LEU:O	1:A:482:THR:HG22	2.18	0.42
1:A:655:PHE:C	1:A:655:PHE:CD1	2.92	0.42
1:A:472:LYS:HG3	1:A:552:LEU:HD22	2.01	0.42
1:A:560:MET:HE1	1:A:600:VAL:C	2.40	0.42
1:A:710:ASN:O	1:A:736:SER:HA	2.20	0.42
1:A:707:ALA:C	1:A:709:CYS:H	2.22	0.41
1:A:398:SER:O	1:A:400:GLY:HA3	2.22	0.40
1:A:852:GLN:HG2	1:A:857:GLU:HB2	2.02	0.40
1:A:847:HIS:HD2	1:A:861:LYS:HG2	1.87	0.40
1:A:856:PHE:HB2	1:A:877:VAL:HB	2.03	0.40

There are no symmetry-related clashes.

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	623/696 (90%)	583 (94%)	34 (6%)	6 (1%)	15 28

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	655	PHE
1	A	316	ARG
1	A	397	GLY
1	A	691	HIS
1	A	577	ALA
1	A	722	ILE

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	536/624 (86%)	495 (92%)	41 (8%)	13 25

All (41) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	283	LYS
1	A	305	GLN
1	A	320	ARG
1	A	328	THR
1	A	337	ILE
1	A	350	ILE
1	A	365	ILE
1	A	395	LEU
1	A	446	LEU
1	A	447	GLU
1	A	482	THR
1	A	488	ARG
1	A	493	LEU
1	A	500	GLN
1	A	507	GLN
1	A	510	GLU
1	A	520	CYS
1	A	538	LEU
1	A	559	ARG
1	A	572	SER
1	A	581	GLU
1	A	585	ASP
1	A	587	THR
1	A	588	GLN
1	A	614	LEU
1	A	625	ASN
1	A	631	ILE
1	A	655	PHE
1	A	695	ILE
1	A	711	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	738	CYS
1	A	743	SER
1	A	750	LYS
1	A	766	LEU
1	A	768	LEU
1	A	806	LYS
1	A	811	ARG
1	A	827	GLU
1	A	836	ASP
1	A	852	GLN
1	A	909	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	619	GLN
1	A	625	ASN
1	A	645	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	C	9/10 (90%)	1 (11%)	0
2	D	9/10 (90%)	2 (22%)	0
All	All	18/20 (90%)	3 (16%)	0

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	C	10	C
2	D	9	G
2	D	10	C

There are no RNA pucker outliers to report.

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 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	SO4	A	1923	-	4,4,4	0.16	0	6,6,6	0.30	0

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1923	SO4	1	0

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, 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	A	636/696 (91%)	0.28	23 (3%) 42 46	22, 57, 91, 116	3 (0%)
2	C	10/10 (100%)	-0.59	0 100 100	35, 56, 110, 133	0
2	D	10/10 (100%)	-0.72	0 100 100	44, 56, 93, 107	0
All	All	656/716 (91%)	0.25	23 (3%) 44 47	22, 57, 93, 133	3 (0%)

All (23) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	399	SER	4.8
1	A	718	VAL	3.6
1	A	577	ALA	3.5
1	A	851	LYS	3.2
1	A	236	ASN	2.9
1	A	513	ILE	2.8
1	A	578	GLY	2.7
1	A	816	LEU	2.6
1	A	741	LEU	2.6
1	A	413	GLY	2.5
1	A	632	LEU	2.5
1	A	467	ILE	2.5
1	A	691	HIS	2.4
1	A	717	TYR	2.4
1	A	799	PRO	2.3
1	A	315	GLU	2.3
1	A	698	SER	2.2
1	A	588	GLN	2.2
1	A	237	LEU	2.2
1	A	867	GLN	2.1
1	A	849	LYS	2.1
1	A	745	ALA	2.0
1	A	473	TYR	2.0

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)

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
3	ZN	A	927	1/1	0.96	0.14	72,72,72,72	0
4	SO4	A	1923	5/5	0.98	0.12	63,63,64,64	0

6.5 Other polymers (i)

There are no such residues in this entry.