



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 24, 2020 – 10:24 am BST

PDB ID : 6K4P  
Title : Crystal structure of xCas9 in complex with sgRNA and DNA (TGG PAM)  
Authors : Chen, W.; Zhang, H.; Zhang, Y.; Wang, Y.; Gan, J.; Ji, Q.  
Deposited on : 2019-05-25  
Resolution : 2.90 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](mailto: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

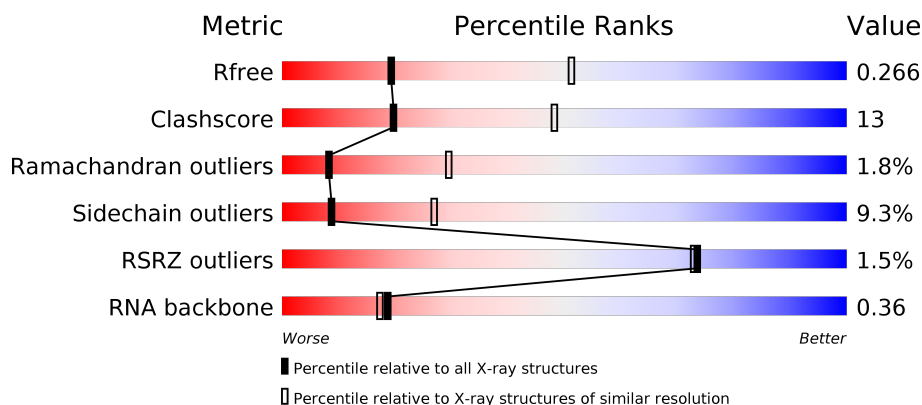
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)
RNA backbone	3102	1007 (3.16-2.64)

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  $\geq 3$ , 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	84	<div> <div>40%</div> <div>44%</div> <div>12%</div> <div>..</div> </div>
2	C	28	<div> <div>68%</div> <div>32%</div> </div>
3	D	12	<div> <div>8%</div> <div>50%</div> <div>42%</div> <div>8%</div> </div>
4	B	1368	<div> <div>%</div> <div>66%</div> <div>27%</div> <div>..</div> </div>

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
5	PO4	B	1404	-	-	X	-

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 13124 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 sgRNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	82	Total	C	N	O	P	0	0	0
			1737	779	318	558	82			

- Molecule 2 is a DNA chain called DNA (28-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	28	Total	C	N	O	P	0	0	0
			567	276	96	168	27			

- Molecule 3 is a DNA chain called non-targeted DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	11	Total	C	N	O	P	0	0	0
			227	110	43	64	10			

- Molecule 4 is a protein called CRISPR-associated endonuclease Cas9/Csn1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	1323	Total	C	N	O	S	0	0	0
			10546	6730	1827	1970	19			

There are 11 discrepancies between the modelled and reference sequences:

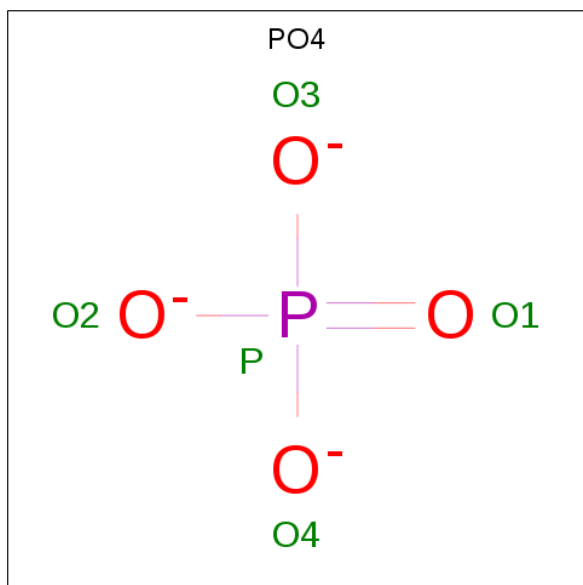
Chain	Residue	Modelled	Actual	Comment	Reference
B	10	ALA	ASP	engineered mutation	UNP Q99ZW2
B	80	LEU	CYS	engineered mutation	UNP Q99ZW2
B	262	THR	ALA	engineered mutation	UNP Q99ZW2
B	324	LEU	ARG	engineered mutation	UNP Q99ZW2
B	409	ILE	SER	engineered mutation	UNP Q99ZW2
B	480	LYS	GLU	engineered mutation	UNP Q99ZW2
B	543	ASP	GLU	engineered mutation	UNP Q99ZW2
B	574	GLU	CYS	engineered mutation	UNP Q99ZW2

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	694	ILE	MET	engineered mutation	UNP Q99ZW2
B	840	ALA	HIS	engineered mutation	UNP Q99ZW2
B	1219	VAL	GLU	engineered mutation	UNP Q99ZW2

- Molecule 5 is PHOSPHATE ION (three-letter code: PO4) (formula: O<sub>4</sub>P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	O	P	0	0
			5	4	1		
5	B	1	Total	O	P	0	0
			5	4	1		
5	B	1	Total	O	P	0	0
			5	4	1		
5	B	1	Total	O	P	0	0
			5	4	1		
5	B	1	Total	O	P	0	0
			5	4	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	5	Total	O	0	0
			5	5		
6	C	1	Total	O	0	0
			1	1		
6	D	1	Total	O	0	0
			1	1		

*Continued on next page...*

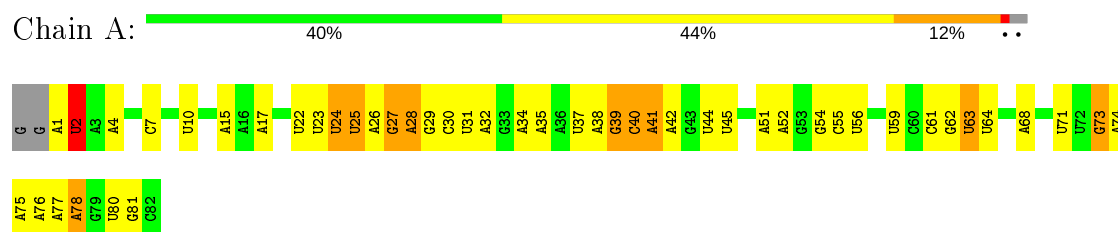
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	15	Total	O	0	0
			15	15		

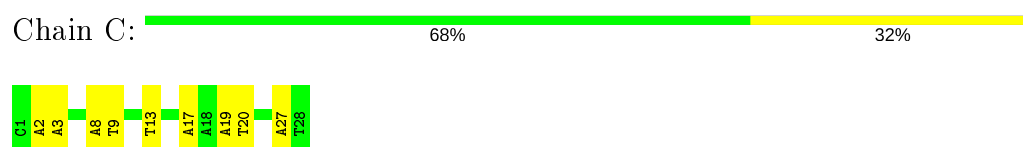
### 3 Residue-property plots [i](#)

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 ( $\text{RSRZ} > 2$ ). 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.

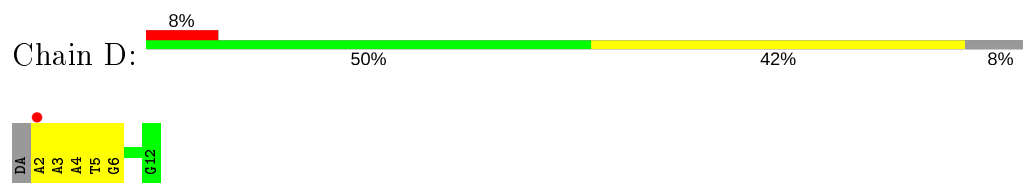
- Molecule 1: sgRNA



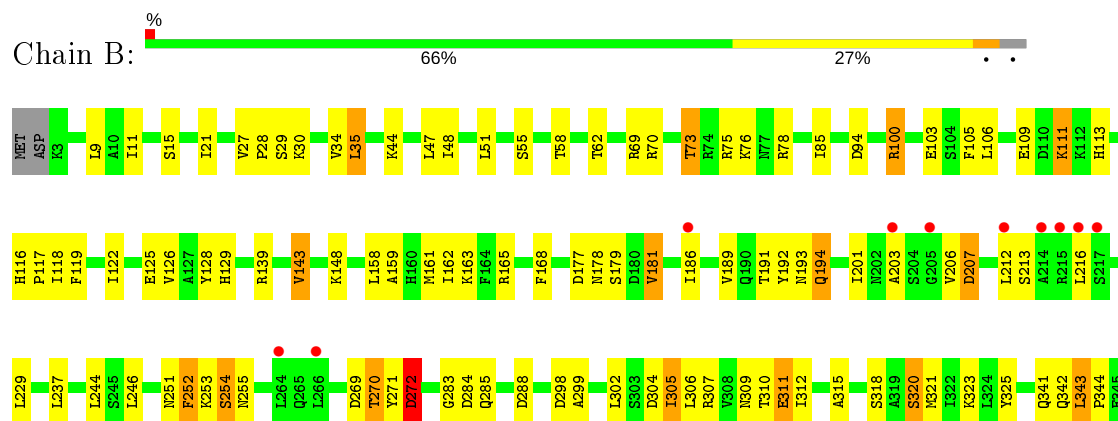
- Molecule 2: DNA (28-MER)

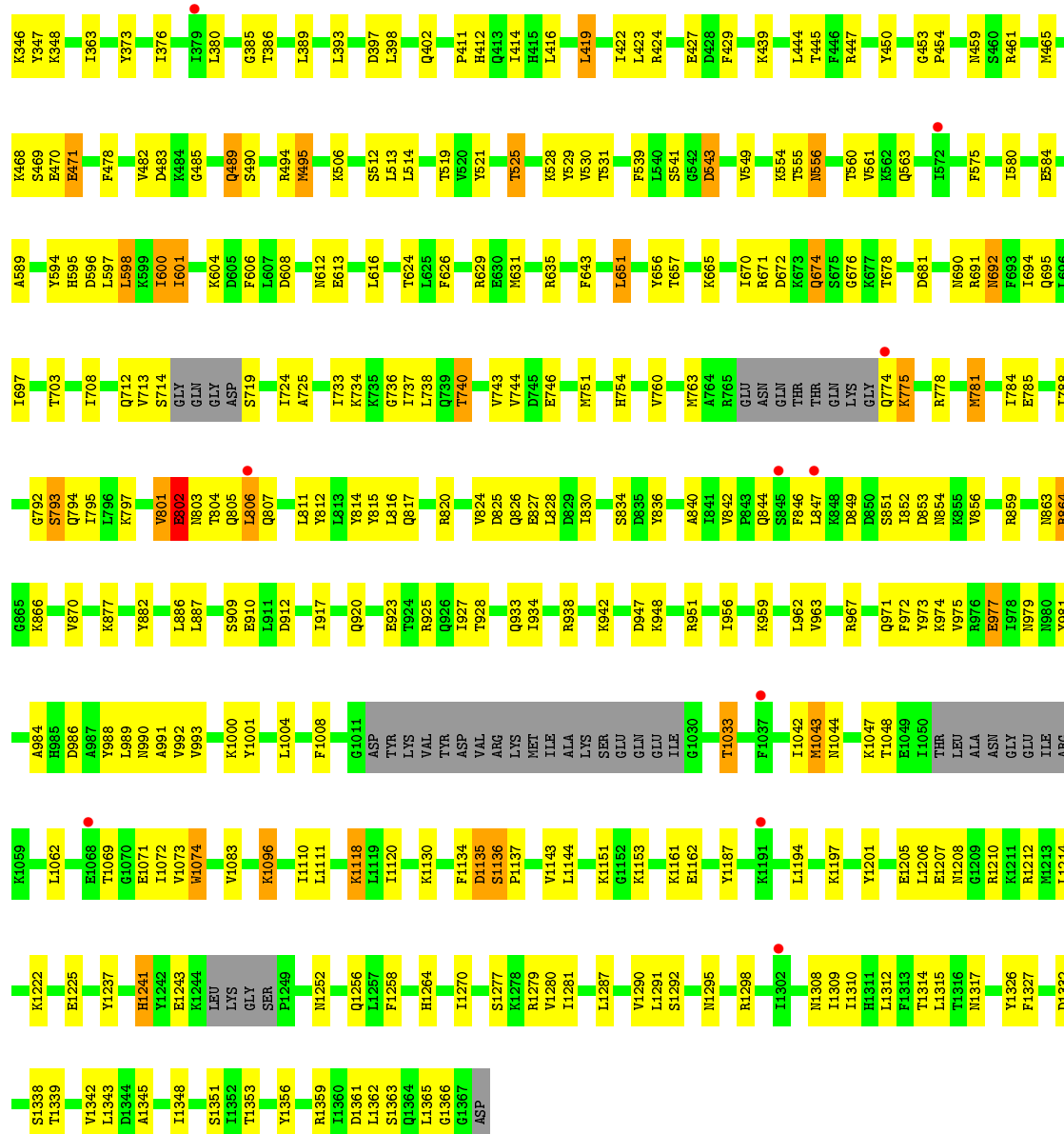


- Molecule 3: non-targeted DNA



- Molecule 4: CRISPR-associated endonuclease Cas9/Csn1







## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	177.78Å 69.67Å 189.02Å 90.00° 109.52° 90.00°	Depositor
Resolution (Å)	46.77 – 2.90 46.72 – 2.90	Depositor EDS
% Data completeness (in resolution range)	91.4 (46.77-2.90) 91.5 (46.72-2.90)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.49 (at 2.91Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, $R_{free}$	0.203 , 0.269 0.207 , 0.266	Depositor DCC
$R_{free}$ test set	2285 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	56.5	Xtriage
Anisotropy	0.095	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 42.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	13124	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.65% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: PO4

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	0.86	10/1947 (0.5%)	0.93	0/3031
2	C	0.70	1/634 (0.2%)	0.91	0/976
3	D	1.03	2/255 (0.8%)	0.79	0/393
4	B	0.73	0/10734	0.92	0/14468
All	All	0.75	13/13570 (0.1%)	0.92	0/18868

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	5	DT	O3'-P	-9.27	1.50	1.61
1	A	26	A	O3'-P	-9.15	1.50	1.61
1	A	24	U	O3'-P	-8.39	1.51	1.61
1	A	25	U	O3'-P	-6.83	1.52	1.61
1	A	7	C	O3'-P	-6.26	1.53	1.61

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	A	1737	0	868	34	0
2	C	567	0	322	13	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	D	227	0	127	7	0
4	B	10546	0	10471	290	0
5	B	25	0	0	2	0
6	A	5	0	0	0	0
6	B	15	0	0	1	0
6	C	1	0	0	0	0
6	D	1	0	0	0	0
All	All	13124	0	11788	314	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 13.

The worst 5 of 314 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:B:781:MET:HG3	4:B:803:ASN:HB3	1.15	1.09
4:B:781:MET:HG3	4:B:803:ASN:CB	1.82	1.09
4:B:106:LEU:O	4:B:111:LYS:HE3	1.55	1.06
4:B:963:VAL:HG21	4:B:990:ASN:ND2	1.88	0.88
4:B:781:MET:HG2	4:B:806:LEU:HD12	1.57	0.87

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
4	B	1311/1368 (96%)	1147 (88%)	140 (11%)	24 (2%)	<b>8</b> <b>29</b>

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	B	1136	SER
4	B	252	PHE
4	B	284	ASP
4	B	311	GLU
4	B	471	GLU

### 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	B	1116/1226 (91%)	1012 (91%)	104 (9%)	9 27

5 of 104 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	556	ASN
4	B	740	THR
4	B	1151	LYS
4	B	584	GLU
4	B	651	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	595	HIS
4	B	807	GLN
4	B	1221	GLN
4	B	489	GLN
4	B	1261	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	80/84 (95%)	19 (23%)	2 (2%)

5 of 19 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	2	U
1	A	4	A
1	A	10	U
1	A	17	A
1	A	29	G

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	28	A
1	A	29	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](#)

5 ligands are modelled in this entry.

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	$\# Z  > 2$
5	PO4	B	1403	-	4,4,4	0.73	0	6,6,6	0.40	0
5	PO4	B	1404	-	4,4,4	0.55	0	6,6,6	0.54	0
5	PO4	B	1401	-	4,4,4	0.62	0	6,6,6	0.53	0
5	PO4	B	1402	-	4,4,4	0.73	0	6,6,6	0.45	0
5	PO4	B	1405	-	4,4,4	0.62	0	6,6,6	0.46	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 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	B	1404	PO4	2	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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	82/84 (97%)	-0.50	0 <span>100</span> <span>100</span>	28, 49, 86, 93	0
2	C	28/28 (100%)	-0.73	0 <span>100</span> <span>100</span>	39, 54, 71, 76	0
3	D	11/12 (91%)	-0.06	1 (9%) <span>9</span> <span>6</span>	36, 54, 121, 148	0
4	B	1323/1368 (96%)	-0.11	20 (1%) <span>73</span> <span>73</span>	30, 62, 109, 137	0
All	All	1444/1492 (96%)	-0.14	21 (1%) <span>73</span> <span>73</span>	28, 61, 108, 148	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	216	LEU	6.9
4	B	215	ARG	5.4
4	B	217	SER	4.3
4	B	264	LEU	3.2
4	B	214	ALA	3.2

### 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<sup>th</sup> 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
5	PO4	B	1402	5/5	0.84	0.22	71,86,92,94	0
5	PO4	B	1405	5/5	0.87	0.13	107,112,117,122	0
5	PO4	B	1404	5/5	0.94	0.10	64,69,74,82	0
5	PO4	B	1403	5/5	0.95	0.11	78,81,85,86	0
5	PO4	B	1401	5/5	0.96	0.16	52,58,61,63	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.