wwPDB X-ray Structure Validation Summary Report (i)

Nov 15, 2023 – 10:55 PM JST

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*
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 (i)) were used in the production of this report:

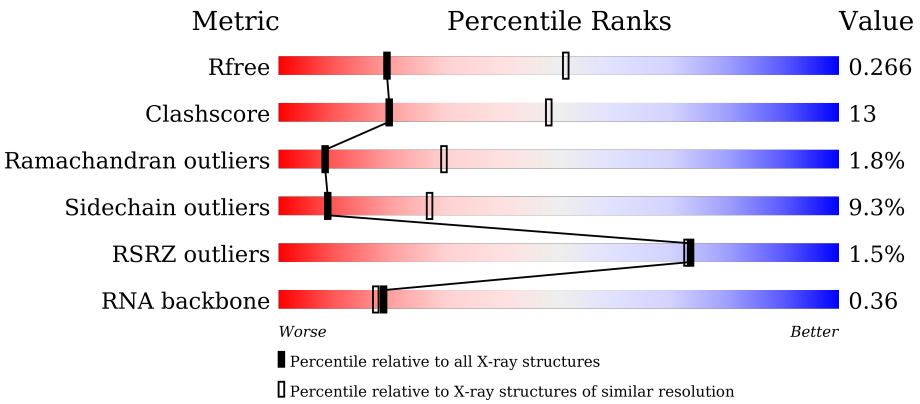
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.90 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 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	A	84	40% 44% 12% • •
2	C	28	68% 32%
3	D	12	8% 50% 42% 8%
4	B	1368	% 66% 27% • •

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 (i)

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
			Total	C	N	O	P			
1	A	82	1737	779	318	558	82	0	0	0

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

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

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

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

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

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

There are 11 discrepancies between the modelled and reference sequences:

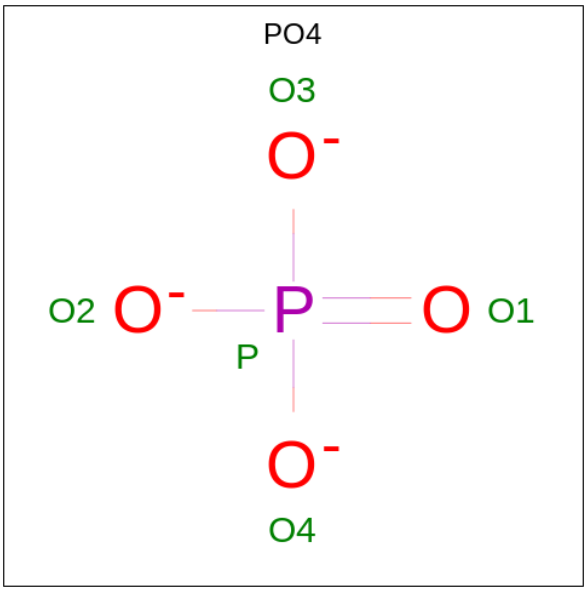
Chain	Residue	Modelled	Actual	Comment	Reference
B	10	ALA	ASP	engineered mutation	UNP Q99ZW2
B	80	LEU	CYS	engineered mutation	UNP Q99ZW2
B	262	THR	ALA	engineered mutation	UNP Q99ZW2
B	324	LEU	ARG	engineered mutation	UNP Q99ZW2
B	409	ILE	SER	engineered mutation	UNP Q99ZW2
B	480	LYS	GLU	engineered mutation	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_4P).



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

- Molecule 6 is water.

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

Continued on next page...

Continued from previous page...

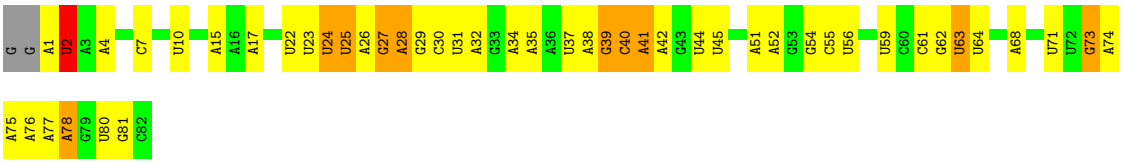
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	B	15	15	15	0	0

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.

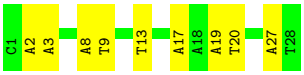
• Molecule 1: sgRNA

Chain A: 40% 44% 12%



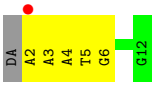
• Molecule 2: DNA (28-MER)

Chain C: 68% 32%



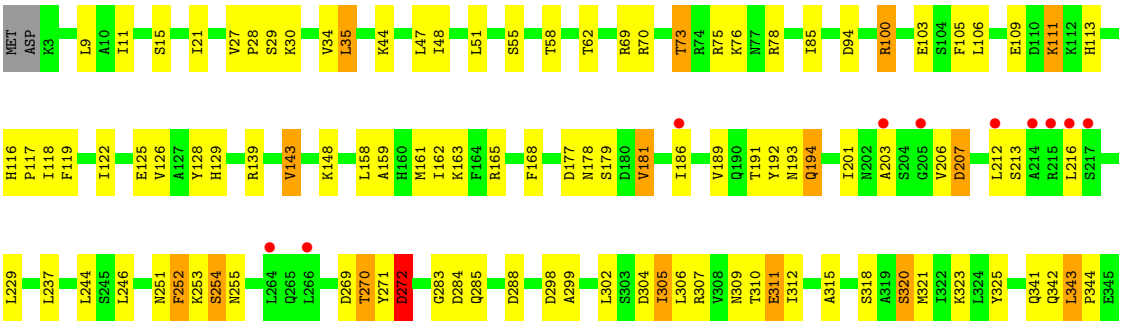
• Molecule 3: non-targeted DNA

Chain D: 8% 50% 42% 8%



• Molecule 4: CRISPR-associated endonuclease Cas9/Csn1

Chain B: % 66% 27%



K346 Y347 K348 I363 Y373 I376 I379 L380 G385 T386 L389 L393 D397 L398 Q402 P411 H412 Q413 I414 H415 L416 L419 I422 L423 R424 E427 D428 F429 K439 L444 T445 F446 R447 Y450 G453 P454 N459 S460 R461 M465

K468 S469 E470 E471 F478 V482 D483 K484 G485 Q489 S490 R494 M495 K506 S512 L513 L514 T519 V520 Y521 T525 K528 Y529 V530 T531 F539 L540 S541 G542 D543 V549 K554 T555 N556 T560 V561 K562 Q563 I572 F575 I580 E584

A589 Y594 H595 D596 L597 L598 K599 I600 I601 K604 D605 F606 L607 D608 N612 E613 L616 T624 L625 F626 R629 E630 M631 R635 F643 L651 Y656 T657 K665 I670 R671 D672 K673 Q674 S675 G676 K677 T678 D681 N690 R691 N692 F693 I694 Q695 L696

I697 T703 I708 Q712 V713 S714 GLY GLN GLY ASP S719 I724 A725 I733 K734 K735 G736 I737 L738 Q739 T740 V743 V744 D745 E746 M751 H754 V760 M763 A764 R765 GLU ASN GLN THR THR GLN LYS GLY Q774 K775 R778 M781 I784 E785 I788

G792 S793 Q794 I795 L796 K797 V801 E802 N803 T804 Q805 L806 Q807 L811 Y812 L813 Y814 Y815 L816 Q817 R820 V824 D825 Q826 E827 L828 D829 I830 S834 D835 Y836 A840 I841 V842 P843 Q844 S845 F846 L847 K848 D849 D850 S851 I852 D853 N854 K855 V856 R859 N863 R864

G865 K866 V870 K877 Y882 L886 L887 S909 E910 L911 D912 I917 Q920 E923 T924 R925 Q926 I927 T928 Q933 I934 R938 K942 D947 K948 R951 I956 K959 L962 V963 R967 Q971 F972 Y973 K974 V975 H976 E977 I978 N979 N980 Y981

A984 H985 D986 A987 Y988 L989 N990 A991 V992 V993 K1000 Y1001 L1004 F1008 G1011 ASP TYR LYS VAL TYR ASP VAL ARG LYS MET ILE ALA LYS SER GLU GLN GLU ILE G1030 T1033 F1037 I1042 M1043 N1044 K1047 T1048 E1049 I1050 THR LEU ALA ASN GLY GLU ILE ARG

K1059 L1062 E1068 T1069 G1070 E1071 I1072 V1073 W1074 V1083 K1096 I1110 L1111 K1118 L1119 I1120 K1130 F1134 D1135 S1136 P1137 V1143 L1144 K1151 G1152 K1153 K1161 E1162 Y1187 K1191 L1194 K1197 Y1201 E1205 L1206 E1207 N1208 G1209 R1210 K1211 R1212 M1213 L1214

K1222 E1225 Y1237 H1241 Y1242 E1243 K1244 LEU LYS GLY SER P1249 N1252 Q1256 L1257 F1258 H1264 I1270 S1277 K1278 R1279 V1280 I1281 L1287 V1290 L1291 S1292 N1295 R1298 I1302 N1308 I1309 I1310 H1311 L1312 F1313 T1314 L1315 T1316 N1317 Y1326 F1327 D1332

S1338 T1339 V1342 L1343 D1344 A1345 I1348 S1351 I1352 T1353 Y1356 R1359 I1360 D1361 L1362 S1363 Q1364 L1365 G1366 G1367 ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	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
$< I/\sigma(I) >$ [1]	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^2)	56.5	Xtriage
Anisotropy	0.095	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 42.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	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^2)	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.*

[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 (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%)	8 29

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 (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
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	719	SER
4	B	864	ARG
4	B	1241	HIS
4	B	754	HIS
4	B	834	SER

Sometimes 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	1044	ASN
4	B	1308	ASN
4	B	1317	ASN
4	B	1261	GLN
4	B	595	HIS

5.3.3 RNA (i)

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 monosaccharides 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	1405	-	4,4,4	0.62	0	6,6,6	0.46	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	1404	-	4,4,4	0.55	0	6,6,6	0.54	0
5	PO4	B	1403	-	4,4,4	0.73	0	6,6,6	0.40	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, 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	82/84 (97%)	-0.50	0	100	100	28, 49, 86, 93	0
2	C	28/28 (100%)	-0.73	0	100	100	39, 54, 71, 76	0
3	D	11/12 (91%)	-0.06	1 (9%)	9	6	36, 54, 121, 148	0
4	B	1323/1368 (96%)	-0.11	20 (1%)	73	73	30, 62, 109, 137	0
All	All	1444/1492 (96%)	-0.14	21 (1%)	73	73	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 monosaccharides 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(Å^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.