



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 02:01 am GMT

PDB ID : 5C5J  
Title : Polymerase Nucleotide complex  
Authors : Nair, D.T.; Kottur, J.  
Deposited on : 2015-06-20  
Resolution : 2.10 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

<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	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

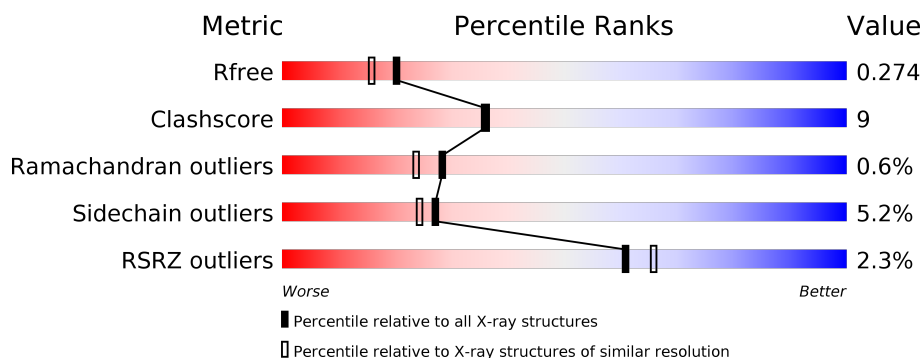
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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))
$R_{free}$	100719	4243 (2.10-2.10)
Clashscore	112137	4788 (2.10-2.10)
Ramachandran outliers	110173	4740 (2.10-2.10)
Sidechain outliers	110143	4741 (2.10-2.10)
RSRZ outliers	101464	4275 (2.10-2.10)

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. 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	352	<div> <div>2%</div> <div> <div></div> <div>78%</div> <div>16%</div> <div>• •</div> </div> </div>
1	F	352	<div> <div>2%</div> <div> <div></div> <div>74%</div> <div>21%</div> <div>• •</div> </div> </div>
2	B	18	<div> <div></div> <div> <div></div> <div>72%</div> <div>28%</div> </div> </div>
2	C	18	<div> <div>6%</div> <div> <div></div> <div>61%</div> <div>22%</div> <div>11%</div> <div>6%</div> </div> </div>
2	G	18	<div> <div>6%</div> <div> <div></div> <div>50%</div> <div>39%</div> <div>11%</div> </div> </div>
2	H	18	<div> <div>6%</div> <div> <div></div> <div>39%</div> <div>39%</div> <div>22%</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 7204 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 DNA polymerase IV.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	F	342	Total	C	N	O	S	0	0	0
			2687	1695	494	484	14			
1	A	342	Total	C	N	O	S	0	0	0
			2687	1695	494	484	14			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	0	GLY	-	expression tag	UNP W8STT9
F	1	SER	-	expression tag	UNP W8STT9
A	0	GLY	-	expression tag	UNP W8STT9
A	1	SER	-	expression tag	UNP W8STT9

- Molecule 2 is a DNA chain called DNA (5'-D(\*TP\*CP\*TP\*AP\*GP\*GP\*GP\*TP\*CP\*CP\*TP\*AP\*GP\*GP\*AP\*CP\*CP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	18	Total	C	N	O	P	0	0	0
			364	174	66	107	17			
2	H	14	Total	C	N	O	P	0	0	0
			283	135	54	81	13			
2	B	18	Total	C	N	O	P	0	0	0
			364	174	66	107	17			
2	C	17	Total	C	N	O	P	0	0	0
			343	164	64	99	16			

- Molecule 3 is 8-OXO-GUANOSINE-5'-TRIPHOSPHATE (three-letter code: 8GT) (formula: C<sub>10</sub>H<sub>16</sub>N<sub>5</sub>O<sub>15</sub>P<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total 32	C 10	N 5	O 14	P 3	0	0

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Mg 1 1	0	0

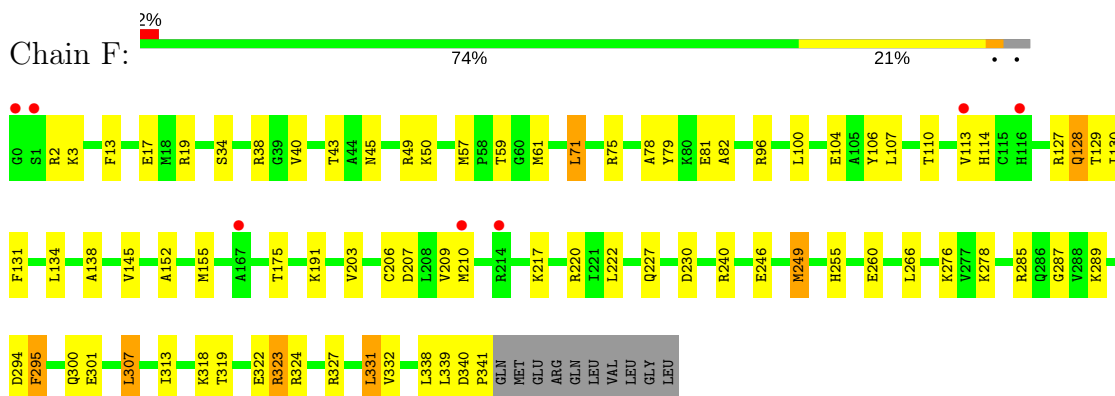
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	F	186	Total O 186 186	0	0
5	A	151	Total O 151 151	0	0
5	G	28	Total O 28 28	0	0
5	H	26	Total O 26 26	0	0
5	B	25	Total O 25 25	0	0
5	C	27	Total O 27 27	0	0

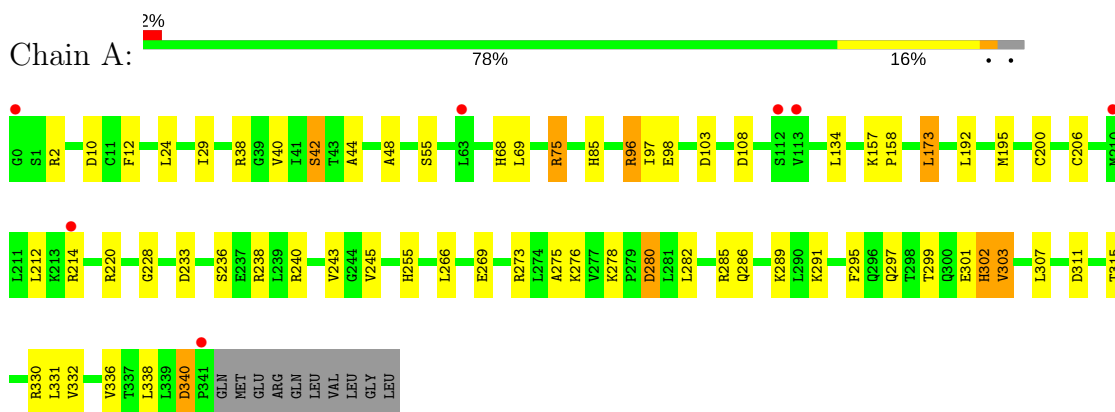
### 3 Residue-property plots [i](#)

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.

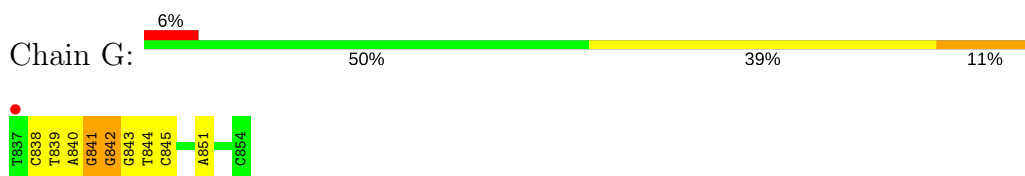
#### • Molecule 1: DNA polymerase IV



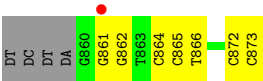
#### • Molecule 1: DNA polymerase IV



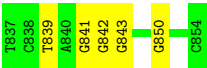
#### • Molecule 2: DNA (5'-D(\*TP\*CP\*TP\*AP\*GP\*GP\*GP\*TP\*CP\*CP\*TP\*AP\*GP\*GP\*AP\*CP\*CP\*C)-3')



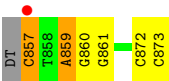
#### • Molecule 2: DNA (5'-D(\*TP\*CP\*TP\*AP\*GP\*GP\*GP\*TP\*CP\*CP\*TP\*AP\*GP\*GP\*AP\*CP\*CP\*C)-3')



• Molecule 2: DNA (5'-D(\*TP\*CP\*TP\*AP\*GP\*GP\*GP\*TP\*CP\*CP\*TP\*AP\*GP\*GP\*AP\*CP\*CP\*C)-3')



• Molecule 2: DNA (5'-D(\*TP\*CP\*TP\*AP\*GP\*GP\*GP\*TP\*CP\*CP\*TP\*AP\*GP\*GP\*AP\*CP\*CP\*C)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	86.45Å 57.38Å 110.64Å 90.00° 94.51° 90.00°	Depositor
Resolution (Å)	36.91 – 2.10 39.11 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.5 (36.91-2.10) 99.7 (39.11-2.10)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.26 (at 2.10Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, $R_{free}$	0.222 , 0.274 0.221 , 0.274	Depositor DCC
$R_{free}$ test set	3219 reflections (5.08%)	DCC
Wilson B-factor (Å <sup>2</sup> )	41.9	Xtriage
Anisotropy	0.400	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 42.9	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	0.94	EDS
Total number of atoms	7204	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.53% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG, 8GT

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	0.39	0/2738	0.60	1/3699 (0.0%)
1	F	0.45	0/2738	0.66	1/3699 (0.0%)
2	B	0.98	0/407	1.11	1/626 (0.2%)
2	C	0.97	1/384 (0.3%)	1.00	1/590 (0.2%)
2	G	1.15	1/407 (0.2%)	1.20	2/626 (0.3%)
2	H	1.02	0/317	0.95	0/487
All	All	0.61	2/6991 (0.0%)	0.76	6/9727 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	857	DC	C1'-N1	5.59	1.56	1.49
2	G	842	DG	O5'-C5'	-5.45	1.28	1.42

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	G	841	DG	OP1-P-OP2	8.03	131.65	119.60
2	B	850	DG	O4'-C1'-N9	7.59	113.31	108.00
2	G	842	DG	O5'-P-OP1	-7.00	99.40	105.70
1	A	75	ARG	NE-CZ-NH1	-5.72	117.44	120.30
2	C	859	DA	O5'-P-OP2	-5.10	101.11	105.70

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	2687	0	2739	44	0
1	F	2687	0	2739	55	1
2	B	364	0	204	8	0
2	C	343	0	190	5	1
2	G	364	0	204	7	0
2	H	283	0	156	6	0
3	A	32	0	10	3	0
4	A	1	0	0	0	0
5	A	151	0	0	7	1
5	B	25	0	0	0	0
5	C	27	0	0	1	0
5	F	186	0	0	14	0
5	G	28	0	0	1	0
5	H	26	0	0	0	0
All	All	7204	0	6242	118	2

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 9.

The worst 5 of 118 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:240:ARG:NH2	2:B:843:DG:OP1	2.01	0.94
1:F:128:GLN:O	1:F:130:ILE:N	2.11	0.83
1:F:285:ARG:NH1	2:H:866:DT:OP2	2.12	0.82
1:F:57:MET:HG3	1:F:61:MET:HE2	1.62	0.81
1:F:191:LYS:NZ	5:F:402:HOH:O	2.15	0.80

All (2) 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:F:240:ARG:NH1	2:C:857:DC:O5'[2_555]	2.10	0.10
5:A:609:HOH:O	5:A:613:HOH:O[2_554]	2.15	0.05

## 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	
1	A	340/352 (97%)	334 (98%)	6 (2%)	0	100	100
1	F	340/352 (97%)	323 (95%)	13 (4%)	4 (1%)	15	9
All	All	680/704 (97%)	657 (97%)	19 (3%)	4 (1%)	28	24

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	129	THR
1	F	128	GLN
1	F	295	PHE
1	F	113	VAL

### 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	286/297 (96%)	268 (94%)	18 (6%)	21	17
1	F	286/297 (96%)	274 (96%)	12 (4%)	34	33
All	All	572/594 (96%)	542 (95%)	30 (5%)	27	24

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	40	VAL
1	A	75	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	336	VAL
1	A	68	HIS
1	A	96	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	227	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 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	$\# Z  > 2$
3	8GT	A	401	4	26,34,35	5.22	16 (61%)	30,54,56	1.99	6 (20%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical

component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	8GT	A	401	4	-	0/18/34/38	0/3/3/3

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	401	8GT	C2'-C3'	-10.57	1.24	1.52
3	A	401	8GT	O4'-C1'	-4.79	1.31	1.42
3	A	401	8GT	C5'-C4'	-3.49	1.40	1.51
3	A	401	8GT	O6-C6	-2.24	1.18	1.24
3	A	401	8GT	C2-N3	2.56	1.48	1.35

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	401	8GT	C2'-C1'-N9	-5.35	110.36	115.79
3	A	401	8GT	N3-C2-N1	-4.99	120.17	127.46
3	A	401	8GT	C5-C6-N1	-2.50	119.93	123.48
3	A	401	8GT	C6-N1-C2	2.34	119.42	116.06
3	A	401	8GT	O4'-C1'-N9	3.65	111.56	108.15

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	401	8GT	3	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	342/352 (97%)	0.17	7 (2%) 65 70	30, 50, 65, 78	0
1	F	342/352 (97%)	0.18	7 (2%) 65 70	26, 43, 62, 78	0
2	B	18/18 (100%)	-0.15	0 100 100	40, 53, 68, 69	0
2	C	17/18 (94%)	-0.04	1 (5%) 23 29	43, 49, 55, 63	0
2	G	18/18 (100%)	0.06	1 (5%) 25 31	31, 48, 73, 77	0
2	H	14/18 (77%)	0.38	1 (7%) 17 21	33, 43, 79, 79	0
All	All	751/776 (96%)	0.16	17 (2%) 61 66	26, 47, 65, 79	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	857	DC	4.6
1	F	0	GLY	4.2
2	G	837	DT	3.7
1	F	116	HIS	3.4
1	A	341	PRO	3.4

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	MG	A	402	1/1	0.98	0.14	-1.14	46,46,46,46	0
3	8GT	A	401	32/33	0.95	0.12	-1.32	48,54,63,63	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.