



# wwPDB X-ray Structure Validation Summary Report

Aug 6, 2014 – 11:38 PM EDT

PDB ID : 4Q7C  
Title : Structure of AF2299, a CDP-alcohol phosphotransferase  
Authors : Clarke, O.B.; Sciara, G.; Tomasek, D.; Banerjee, S.; Rajashankar, K.R.;  
Shapiro, L.; Mancina, F.; New York Consortium on Membrane Protein Structure (NYCOMPS)  
Deposited on : 2014-04-24  
Resolution : 3.10 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

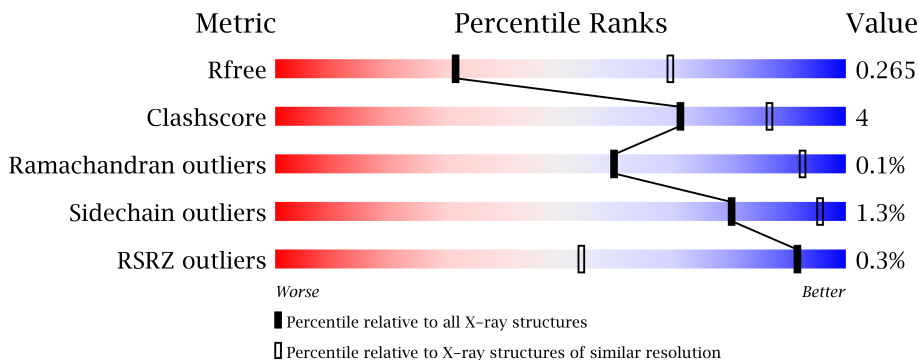
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : **FAILED**  
EDS : stable23489  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable23489

# 1 Overall quality at a glance

The reported resolution of this entry is 3.10 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))
$R_{free}$	66092	1007 (3.18-3.02)
Clashscore	79885	1078 (3.16-3.04)
Ramachandran outliers	78287	1044 (3.16-3.04)
Sidechain outliers	78261	1044 (3.16-3.04)
RSRZ outliers	66119	1008 (3.18-3.02)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	372	
1	B	372	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	MPG	A	403	-	X
3	MPG	A	404	-	X
3	MPG	A	405	-	X
3	MPG	B	403	-	X
3	MPG	B	404	-	X
3	MPG	B	405	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Geometry	Electron density
3	MPG	B	406	-	X
4	TLA	A	407	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5412 atoms, of which 0 are hydrogen and 0 are deuterium.

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 AF2299, a CDP-alcohol phosphotransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	332	Total	C	N	O	S	0	0	0
			2585	1702	413	464	6			
1	B	341	Total	C	N	O	S	0	0	0
			2687	1767	435	479	6			

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-27	MET	-	EXPRESSION TAG	UNP O27985
A	-26	HIS	-	EXPRESSION TAG	UNP O27985
A	-25	HIS	-	EXPRESSION TAG	UNP O27985
A	-24	HIS	-	EXPRESSION TAG	UNP O27985
A	-23	HIS	-	EXPRESSION TAG	UNP O27985
A	-22	HIS	-	EXPRESSION TAG	UNP O27985
A	-21	HIS	-	EXPRESSION TAG	UNP O27985
A	-20	HIS	-	EXPRESSION TAG	UNP O27985
A	-19	HIS	-	EXPRESSION TAG	UNP O27985
A	-18	HIS	-	EXPRESSION TAG	UNP O27985
A	-17	HIS	-	EXPRESSION TAG	UNP O27985
A	-16	SER	-	EXPRESSION TAG	UNP O27985
A	-15	SER	-	EXPRESSION TAG	UNP O27985
A	-14	GLY	-	EXPRESSION TAG	UNP O27985
A	-13	VAL	-	EXPRESSION TAG	UNP O27985
A	-12	ASP	-	EXPRESSION TAG	UNP O27985
A	-11	LEU	-	EXPRESSION TAG	UNP O27985
A	-10	GLY	-	EXPRESSION TAG	UNP O27985
A	-9	THR	-	EXPRESSION TAG	UNP O27985
A	-8	GLU	-	EXPRESSION TAG	UNP O27985
A	-7	ASN	-	EXPRESSION TAG	UNP O27985
A	-6	LEU	-	EXPRESSION TAG	UNP O27985
A	-5	TYR	-	EXPRESSION TAG	UNP O27985
A	-4	PHE	-	EXPRESSION TAG	UNP O27985
A	-3	GLN	-	EXPRESSION TAG	UNP O27985

*Continued on next page...*

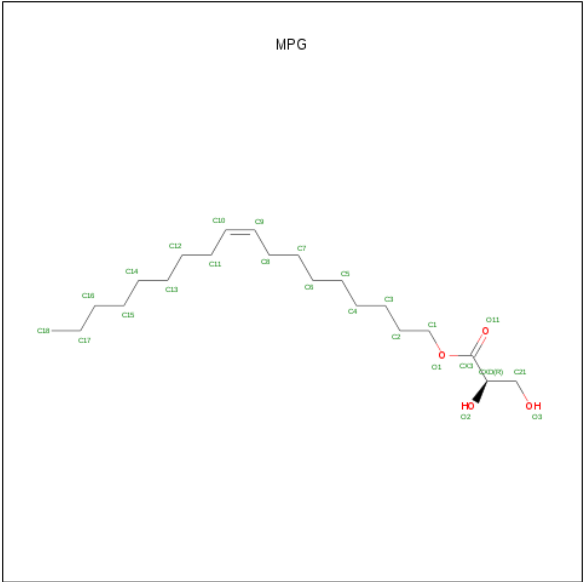
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	EXPRESSION TAG	UNP O27985
A	-1	ASN	-	EXPRESSION TAG	UNP O27985
A	0	ALA	-	EXPRESSION TAG	UNP O27985
B	-27	MET	-	EXPRESSION TAG	UNP O27985
B	-26	HIS	-	EXPRESSION TAG	UNP O27985
B	-25	HIS	-	EXPRESSION TAG	UNP O27985
B	-24	HIS	-	EXPRESSION TAG	UNP O27985
B	-23	HIS	-	EXPRESSION TAG	UNP O27985
B	-22	HIS	-	EXPRESSION TAG	UNP O27985
B	-21	HIS	-	EXPRESSION TAG	UNP O27985
B	-20	HIS	-	EXPRESSION TAG	UNP O27985
B	-19	HIS	-	EXPRESSION TAG	UNP O27985
B	-18	HIS	-	EXPRESSION TAG	UNP O27985
B	-17	HIS	-	EXPRESSION TAG	UNP O27985
B	-16	SER	-	EXPRESSION TAG	UNP O27985
B	-15	SER	-	EXPRESSION TAG	UNP O27985
B	-14	GLY	-	EXPRESSION TAG	UNP O27985
B	-13	VAL	-	EXPRESSION TAG	UNP O27985
B	-12	ASP	-	EXPRESSION TAG	UNP O27985
B	-11	LEU	-	EXPRESSION TAG	UNP O27985
B	-10	GLY	-	EXPRESSION TAG	UNP O27985
B	-9	THR	-	EXPRESSION TAG	UNP O27985
B	-8	GLU	-	EXPRESSION TAG	UNP O27985
B	-7	ASN	-	EXPRESSION TAG	UNP O27985
B	-6	LEU	-	EXPRESSION TAG	UNP O27985
B	-5	TYR	-	EXPRESSION TAG	UNP O27985
B	-4	PHE	-	EXPRESSION TAG	UNP O27985
B	-3	GLN	-	EXPRESSION TAG	UNP O27985
B	-2	SER	-	EXPRESSION TAG	UNP O27985
B	-1	ASN	-	EXPRESSION TAG	UNP O27985
B	0	ALA	-	EXPRESSION TAG	UNP O27985

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

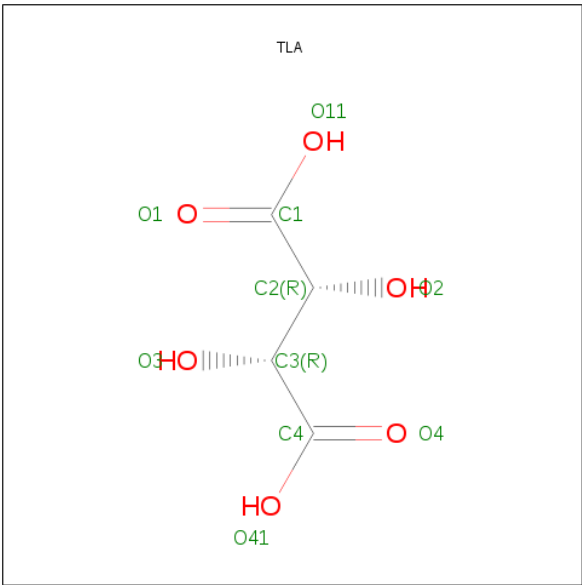
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	2	Total Ca 2 2	0	0
2	A	2	Total Ca 2 2	0	0

- Molecule 3 is 1-MONOOLEOYL-RAC-GLYCEROL (three-letter code: MPG) (formula: C<sub>21</sub>H<sub>40</sub>O<sub>4</sub>).



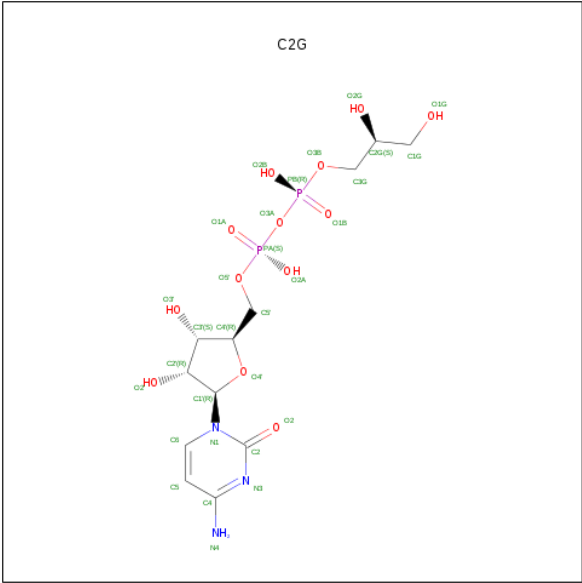
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C 7 7	0	0
3	A	1	Total C 7 7	0	0
3	A	1	Total C 7 7	0	0
3	A	1	Total C 7 7	0	0
3	B	1	Total C 7 7	0	0
3	B	1	Total C 7 7	0	0
3	B	1	Total C 7 7	0	0
3	B	1	Total C 7 7	0	0

- Molecule 4 is L(+)-TARTARIC ACID (three-letter code: TLA) (formula: C<sub>4</sub>H<sub>6</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			10	4	6		
4	B	1	Total	C	O	0	0
			10	4	6		

- Molecule 5 is [CYTIDINE-5'-PHOSPHATE]GLYCERYLPHOSPHORIC ACID ESTER (three-letter code: C2G) (formula: C<sub>12</sub>H<sub>21</sub>N<sub>3</sub>O<sub>13</sub>P<sub>2</sub>).



*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	B	1	Total	C	N	O	P	0	0
			30	12	3	13	2		





## 4 Data and refinement statistics

Xtrriage (Phenix) failed to run properly - this section will therefore be incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	46.42Å 90.85Å 106.61Å 90.00° 92.21° 90.00°	Depositor
Resolution (Å)	69.13 – 3.10 69.13 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.8 (69.13-3.10) 99.8 (69.13-3.10)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.238 , 0.269 0.231 , 0.265	Depositor DCC
$R_{free}$ test set	803 reflections (4.99%)	DCC
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 18.2	EDS
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	5412	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	40.0	wwPDB-VP

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MPG, CA, TLA, C2G

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	0.23	0/2636	0.39	0/3582
1	B	0.24	0/2741	0.40	0/3718
All	All	0.23	0/5377	0.39	0/7300

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2585	0	2606	27	1
1	B	2687	0	2741	21	1
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	28	0	52	2	0
3	B	28	0	52	1	0
4	A	10	0	4	1	0
4	B	10	0	4	0	0
5	A	30	0	19	1	0
5	B	30	0	19	2	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	5412	0	5497	49	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 4.

The worst 5 of 49 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
5:B:408:C2G:O4'	5:B:408:C2G:C1'	1.64	1.24
5:A:408:C2G:C1'	5:A:408:C2G:O4'	1.64	1.13
1:A:69:ARG:NH1	1:A:102:TYR:OH	2.21	0.72
1:B:21:LEU:HD13	1:B:118:LEU:HD13	1.79	0.64
1:A:70:LYS:HD2	1:A:111:TYR:HB2	1.80	0.62

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	Distance(Å)	Clash(Å)
1:A:81:TYR:OH	1:B:45:GLU:OE2[2_645]	2.18	0.02

## 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	328/372 (88%)	319 (97%)	8 (2%)	1 (0%)	50	87
1	B	339/372 (91%)	330 (97%)	9 (3%)	0	100	100
All	All	667/744 (90%)	649 (97%)	17 (2%)	1 (0%)	59	92

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	341	SER

### 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	271/322 (84%)	268 (99%)	3 (1%)	84	96
1	B	283/322 (88%)	279 (99%)	4 (1%)	78	95
All	All	554/644 (86%)	547 (99%)	7 (1%)	80	95

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	120	ASN
1	B	266	PHE
1	B	138	LYS
1	A	248	PHE
1	B	248	PHE

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	106	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

## 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 16 ligands modelled in this entry, 4 are monoatomic - leaving 12 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	MPG	A	403	-	6,6,24	3.47	1 (16%)	5,5,25	0.73	0
3	MPG	A	404	-	6,6,24	3.53	1 (16%)	5,5,25	0.88	0
3	MPG	A	405	-	6,6,24	3.61	1 (16%)	5,5,25	0.80	0
3	MPG	A	406	-	6,6,24	3.42	1 (16%)	5,5,25	0.85	0
4	TLA	A	407	-	9,9,9	0.93	0	12,12,12	1.23	2 (16%)
5	C2G	A	408	2	31,31,31	4.44	8 (25%)	43,46,46	1.54	6 (13%)
3	MPG	B	403	-	6,6,24	3.57	1 (16%)	5,5,25	0.88	0
3	MPG	B	404	-	6,6,24	3.55	1 (16%)	5,5,25	0.90	0
3	MPG	B	405	-	6,6,24	3.57	1 (16%)	5,5,25	0.83	0
3	MPG	B	406	-	6,6,24	3.46	1 (16%)	5,5,25	0.80	0
4	TLA	B	407	-	9,9,9	0.89	0	12,12,12	1.27	3 (25%)
5	C2G	B	408	2	31,31,31	4.43	7 (22%)	43,46,46	1.40	6 (13%)

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	MPG	A	403	-	-	0/4/4/25	0/0/0/0
3	MPG	A	404	-	-	0/4/4/25	0/0/0/0
3	MPG	A	405	-	-	0/4/4/25	0/0/0/0
3	MPG	A	406	-	-	0/4/4/25	0/0/0/0
4	TLA	A	407	-	-	0/12/12/12	0/0/0/0
5	C2G	A	408	2	-	0/22/40/40	0/2/2/2
3	MPG	B	403	-	-	0/4/4/25	0/0/0/0
3	MPG	B	404	-	-	0/4/4/25	0/0/0/0
3	MPG	B	405	-	-	0/4/4/25	0/0/0/0
3	MPG	B	406	-	-	0/4/4/25	0/0/0/0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	TLA	B	407	-	-	0/12/12/12	0/0/0/0
5	C2G	B	408	2	-	0/22/40/40	0/2/2/2

The worst 5 of 23 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	408	C2G	O4'-C1'	18.60	1.64	1.41
5	A	408	C2G	O4'-C1'	18.45	1.64	1.41
5	A	408	C2G	C2'-C1'	-13.61	1.33	1.53
5	B	408	C2G	C2'-C1'	-13.33	1.34	1.53
3	A	405	MPG	C12-C13	-8.84	1.51	1.55

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	408	C2G	C6-C5-C4	6.76	120.73	117.51
5	B	408	C2G	C6-C5-C4	6.09	120.42	117.51
5	A	408	C2G	C3'-C2'-C1'	3.61	106.58	100.92
5	B	408	C2G	C3'-C2'-C1'	2.76	105.25	100.92
5	B	408	C2G	C2-N3-C4	2.71	119.55	115.65

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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<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	332/372 (89%)	-0.04	1 (0%)	91 53	26, 36, 63, 67	0
1	B	341/372 (91%)	-0.06	1 (0%)	91 53	28, 38, 57, 60	0
All	All	673/744 (90%)	-0.05	2 (0%)	91 53	26, 37, 60, 67	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	139	ARG	2.2
1	A	108	LEU	2.0

### 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 carbohydrates 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. 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
3	MPG	A	404	7/25	0.34	7.34	30,30,31,32	0

*Continued on next page...*



*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	MPG	B	406	7/25	0.32	6.74	30,31,32,33	0
3	MPG	B	405	7/25	0.28	5.12	32,34,34,34	0
3	MPG	B	404	7/25	0.34	4.90	28,29,30,30	0
3	MPG	B	403	7/25	0.30	2.99	30,31,32,34	0
4	TLA	A	407	10/10	0.26	2.49	30,32,34,34	0
3	MPG	A	403	7/25	0.30	2.19	30,32,33,33	0
3	MPG	A	405	7/25	0.28	2.17	26,27,28,28	0
3	MPG	A	406	7/25	0.26	1.49	30,32,33,33	0
4	TLA	B	407	10/10	0.33	1.31	34,38,40,40	0
5	C2G	A	408	30/30	0.20	0.74	32,33,34,35	0
5	C2G	B	408	30/30	0.18	0.20	30,32,36,37	0
2	CA	A	402	1/1	0.10	-1.82	32,32,32,32	1
2	CA	B	402	1/1	0.09	-2.11	32,32,32,32	1
2	CA	A	401	1/1	0.09	-2.89	34,34,34,34	0
2	CA	B	401	1/1	0.07	-3.64	32,32,32,32	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.