



Full wwPDB X-ray Structure Validation Report ⓘ

Jun 16, 2024 – 10:27 PM EDT

PDB ID : 5MQM
Title : Glycoside hydrolase BT_0986
Authors : Basle, A.; Ndeh, D.; Rogowski, A.; Cartmell, A.; Luis, A.S.; Venditto, I.;
Labourel, A.; Gilbert, H.J.
Deposited on : 2016-12-20
Resolution : 2.11 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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 : 2.37.1
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) : 2.37.1

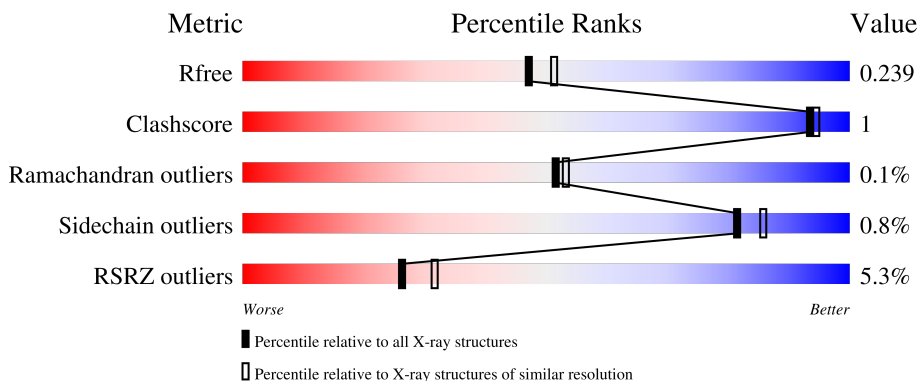
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.11 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	6241 (2.14-2.10)
Clashscore	141614	6778 (2.14-2.10)
Ramachandran outliers	138981	6705 (2.14-2.10)
Sidechain outliers	138945	6706 (2.14-2.10)
RSRZ outliers	127900	6112 (2.14-2.10)

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 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 $\leq 5\%$. 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	1105	

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 8661 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.

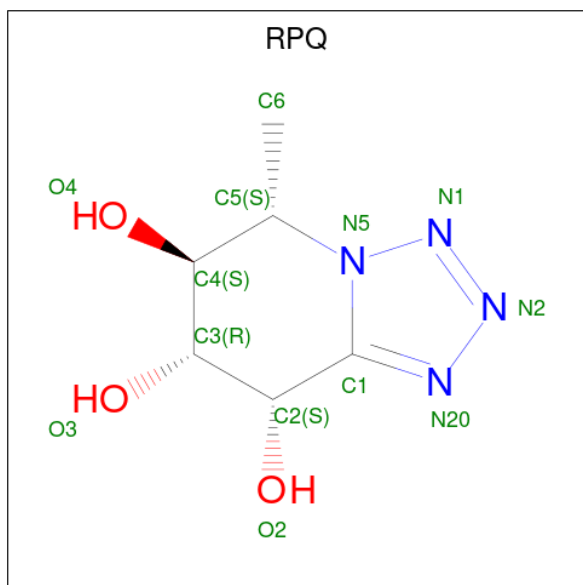
- Molecule 1 is a protein called Glycosyl hydrolases family 2, sugar binding domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1038	8276	5270	1424	1541	41	0	1	0

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total	Ca	0	0
			3	3		

- Molecule 3 is D-rhamnopyranose tetrazole (three-letter code: RPQ) (formula: C₆H₁₀N₄O₃).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	13	6	4	3	0	0

- Molecule 4 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: C₈H₁₈O₅).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			13	8	5		
4	A	1	Total	C	O	0	0
			13	8	5		

- Molecule 5 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



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

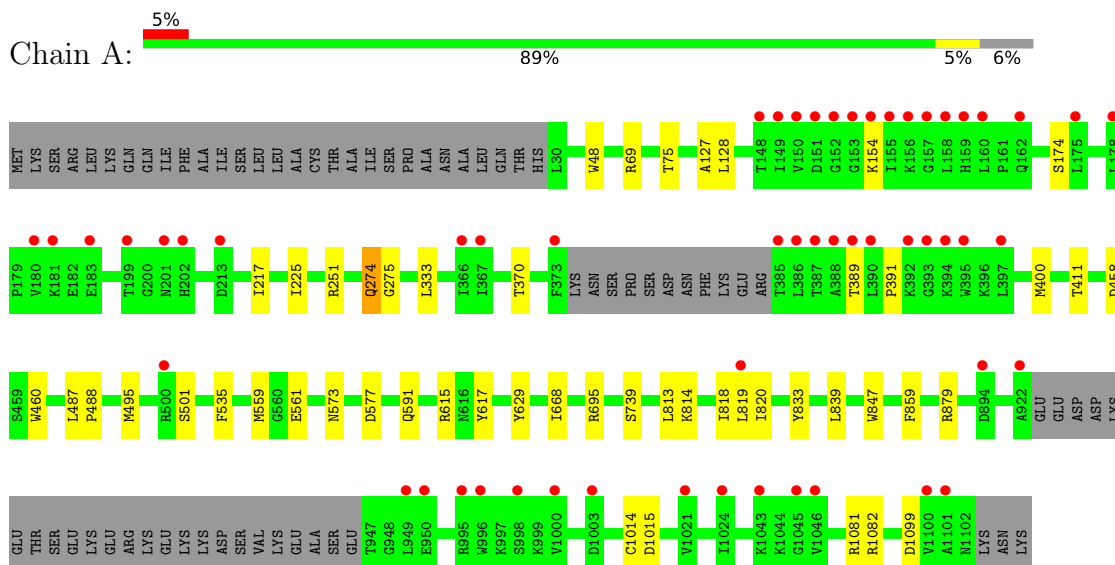
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	323	Total 323	O 323	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: Glycosyl hydrolases family 2, sugar binding domain



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	64.05Å 84.81Å 120.63Å 90.00° 99.20° 90.00°	Depositor
Resolution (Å)	50.69 – 2.11 50.69 – 2.11	Depositor EDS
% Data completeness (in resolution range)	98.8 (50.69-2.11) 98.8 (50.69-2.11)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.69 (at 2.10Å)	Xtrriage
Refinement program	REFMAC 5.8.0155	Depositor
R, R_{free}	0.199 , 0.240 0.203 , 0.239	Depositor DCC
R_{free} test set	3563 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å ²)	35.0	Xtrriage
Anisotropy	0.084	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 38.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	8661	wwPDB-VP
Average B, all atoms (Å ²)	41.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.20% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: PG4, CA, PGE, RPQ

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	0.57	0/8491	0.76	8/11524 (0.1%)

There are no bond length outliers.

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1082	ARG	NE-CZ-NH1	5.60	123.10	120.30
1	A	69	ARG	NE-CZ-NH1	5.46	123.03	120.30
1	A	251	ARG	NE-CZ-NH1	5.41	123.01	120.30
1	A	577	ASP	CB-CG-OD1	5.27	123.04	118.30
1	A	615	ARG	NE-CZ-NH2	-5.24	117.68	120.30
1	A	695	ARG	NE-CZ-NH2	-5.10	117.75	120.30
1	A	879	ARG	NE-CZ-NH2	-5.06	117.77	120.30
1	A	1081	ARG	NE-CZ-NH1	5.04	122.82	120.30

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	8276	0	8108	23	0
2	A	3	0	0	0	0
3	A	13	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	26	0	36	2	0
5	A	20	0	28	0	0
6	A	323	0	0	0	0
All	All	8661	0	8172	23	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 1.

All (23) 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:814:LYS:HG3	1:A:820:ILE:HD12	1.76	0.67
1:A:128:LEU:HD13	1:A:460:TRP:HA	1.84	0.59
1:A:488:PRO:HD2	1:A:495:MET:HE1	1.87	0.57
1:A:174:SER:HA	1:A:370:THR:HG22	1.91	0.53
1:A:847:TRP:HA	1:A:859:PHE:O	2.12	0.50
1:A:217:ILE:HD13	1:A:225:ILE:HD13	1.94	0.49
1:A:819:LEU:HD11	1:A:833:TYR:CB	2.43	0.49
1:A:48:TRP:HB2	1:A:75:THR:OG1	2.13	0.48
1:A:274:GLN:NE2	1:A:275:GLY:O	2.47	0.47
1:A:400:MET:HE1	1:A:487:LEU:C	2.38	0.44
1:A:488:PRO:CD	1:A:495:MET:HE1	2.47	0.44
1:A:154:LYS:HA	1:A:389:THR:HG22	2.00	0.44
1:A:411:THR:HG22	4:A:1205:PG4:C4	2.48	0.43
1:A:127:ALA:HA	1:A:128:LEU:HA	1.82	0.43
1:A:411:THR:HG22	4:A:1205:PG4:H42	2.01	0.43
1:A:559:MET:HA	1:A:591:GLN:O	2.19	0.42
1:A:48:TRP:HA	1:A:629:TYR:O	2.19	0.42
1:A:1014:CYS:HA	1:A:1015:ASP:HA	1.89	0.42
1:A:617:TYR:OH	1:A:668:ILE:HG23	2.20	0.41
1:A:333:LEU:HD11	1:A:501:SER:HB3	2.03	0.41
1:A:813:LEU:HD22	1:A:818:ILE:HD13	2.02	0.41
1:A:839:LEU:HD23	1:A:847:TRP:CG	2.55	0.41
1:A:174:SER:HA	1:A:370:THR:CG2	2.51	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	1033/1105 (94%)	995 (96%)	37 (4%)	1 (0%)	51	53

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	391	PRO

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	874/934 (94%)	867 (99%)	7 (1%)	81	86

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	274	GLN
1	A	458	ASP
1	A	535	PHE
1	A	561	GLU
1	A	573	ASN
1	A	739	SER
1	A	1099	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 3 are monoatomic - leaving 5 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	PG4	A	1205	-	12,12,12	0.58	0	11,11,11	0.53	0
3	RPQ	A	1204	2	11,14,14	2.21	4 (36%)	11,21,21	3.24	4 (36%)
4	PG4	A	1206	-	12,12,12	0.61	0	11,11,11	0.77	0
5	PGE	A	1208	-	9,9,9	0.45	0	8,8,8	0.44	0
5	PGE	A	1207	-	9,9,9	0.50	0	8,8,8	0.21	0

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
4	PG4	A	1205	-	-	4/10/10/10	-
3	RPQ	A	1204	2	-	-	0/1/2/2
4	PG4	A	1206	-	-	4/10/10/10	-
5	PGE	A	1208	-	-	5/7/7/7	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	PGE	A	1207	-	-	4/7/7/7	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1204	RPQ	N1-N2	4.23	1.40	1.30
3	A	1204	RPQ	N1-N5	3.61	1.41	1.34
3	A	1204	RPQ	N20-N2	3.53	1.40	1.34
3	A	1204	RPQ	C1-C2	-2.45	1.47	1.51

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1204	RPQ	N2-N1-N5	-8.33	99.24	106.02
3	A	1204	RPQ	C4-C3-C2	4.57	117.18	110.24
3	A	1204	RPQ	N20-N2-N1	3.39	113.92	110.09
3	A	1204	RPQ	C3-C4-C5	2.04	114.75	111.05

There are no chirality outliers.

All (17) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1206	PG4	C6-C5-O3-C4
4	A	1205	PG4	O3-C5-C6-O4
4	A	1206	PG4	O1-C1-C2-O2
5	A	1207	PGE	O1-C1-C2-O2
5	A	1208	PGE	O1-C1-C2-O2
4	A	1205	PG4	O4-C7-C8-O5
5	A	1208	PGE	O3-C5-C6-O4
4	A	1206	PG4	C5-C6-O4-C7
4	A	1205	PG4	C8-C7-O4-C6
5	A	1207	PGE	C1-C2-O2-C3
5	A	1208	PGE	C1-C2-O2-C3
4	A	1205	PG4	O1-C1-C2-O2
5	A	1208	PGE	C3-C4-O3-C5
5	A	1207	PGE	O2-C3-C4-O3
4	A	1206	PG4	C1-C2-O2-C3
5	A	1207	PGE	O3-C5-C6-O4
5	A	1208	PGE	O2-C3-C4-O3

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1205	PG4	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

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, 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	1038/1105 (93%)	0.29	55 (5%) 26 31	24, 38, 66, 101	0

All (55) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1003	ASP	6.3
1	A	159	HIS	5.4
1	A	155	ILE	5.2
1	A	156	LYS	5.2
1	A	150	VAL	4.9
1	A	390	LEU	4.6
1	A	152	GLY	4.6
1	A	202	HIS	4.6
1	A	160	LEU	4.5
1	A	180	VAL	4.4
1	A	158	LEU	4.2
1	A	949	LEU	4.2
1	A	149	ILE	4.2
1	A	157	GLY	4.1
1	A	392	LYS	4.0
1	A	387	THR	3.9
1	A	950	GLU	3.9
1	A	393	GLY	3.8
1	A	181	LYS	3.8
1	A	199	THR	3.7
1	A	1100	VAL	3.7
1	A	1021	VAL	3.6
1	A	373	PHE	3.6
1	A	397	LEU	3.6
1	A	154	LYS	3.4
1	A	1101	ALA	3.3
1	A	153	GLY	3.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	395	TRP	3.1
1	A	1046	VAL	3.0
1	A	1024	ILE	3.0
1	A	367	ILE	2.9
1	A	162	GLN	2.8
1	A	183	GLU	2.8
1	A	151	ASP	2.7
1	A	213	ASP	2.7
1	A	996	TRP	2.6
1	A	201	ASN	2.6
1	A	388	ALA	2.5
1	A	1000	VAL	2.5
1	A	389	THR	2.4
1	A	178	LEU	2.4
1	A	894	ASP	2.4
1	A	819	LEU	2.3
1	A	385	THR	2.3
1	A	998	SER	2.3
1	A	922	ALA	2.3
1	A	386	LEU	2.2
1	A	366	ILE	2.2
1	A	175	LEU	2.2
1	A	1045	GLY	2.2
1	A	394	LYS	2.2
1	A	500	ARG	2.1
1	A	148	THR	2.1
1	A	995	ARG	2.1
1	A	1043	LYS	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 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(\AA^2)	Q<0.9
4	PG4	A	1205	13/13	0.85	0.18	44,51,57,58	0
4	PG4	A	1206	13/13	0.87	0.19	46,53,57,59	0
5	PGE	A	1207	10/10	0.88	0.16	49,53,62,64	0
5	PGE	A	1208	10/10	0.88	0.13	55,57,61,61	0
3	RPQ	A	1204	13/13	0.91	0.15	38,41,43,43	0
2	CA	A	1203	1/1	0.95	0.25	59,59,59,59	0
2	CA	A	1201	1/1	0.98	0.13	39,39,39,39	0
2	CA	A	1202	1/1	0.99	0.07	35,35,35,35	0

6.5 Other polymers [i](#)

There are no such residues in this entry.