



Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 03:34 pm BST

PDB ID : 5T9A
Title : Crystal structure of BuGH2Cwt
Authors : Pluvinae, B.; Boraston, A.B.; Abbott, W.D.
Deposited on : 2016-09-09
Resolution : 2.50 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 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

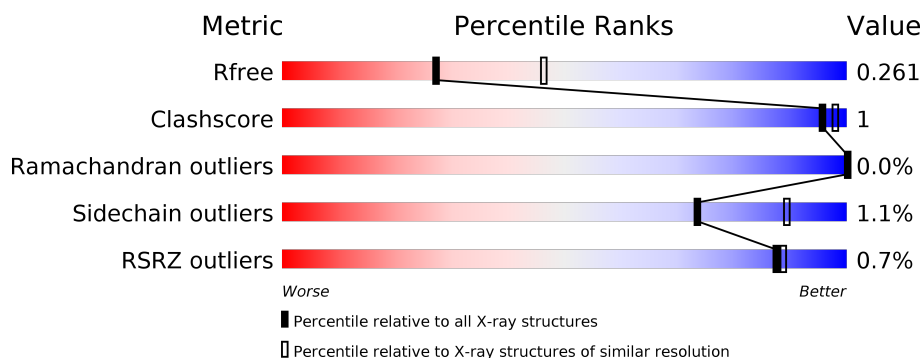
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 ≥ 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	846	<div> <div>%</div> <div> <div></div> <div>92%</div> <div></div> </div> <div>• •</div> </div>
1	B	846	<div> <div>%</div> <div> <div></div> <div>93%</div> <div></div> </div> <div>• •</div> </div>
1	C	846	<div> <div>%</div> <div> <div></div> <div>93%</div> <div></div> </div> <div>• •</div> </div>
1	D	846	<div> <div></div> <div>92%</div> <div></div> </div> <div>• •</div>

2 Entry composition [i](#)

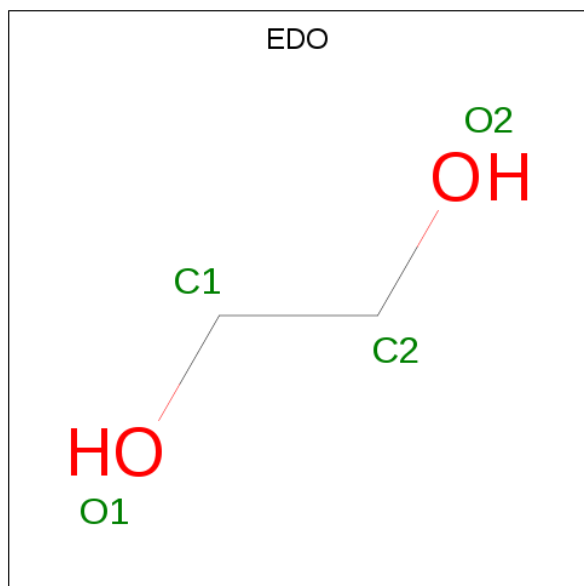
There are 4 unique types of molecules in this entry. The entry contains 26748 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Glycoside Hydrolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	811	Total	C	N	O	S	0	2	0
			6502	4149	1097	1231	25			
1	B	812	Total	C	N	O	S	0	0	0
			6488	4139	1096	1228	25			
1	C	811	Total	C	N	O	S	0	0	0
			6441	4113	1086	1217	25			
1	D	811	Total	C	N	O	S	0	0	0
			6524	4165	1101	1233	25			

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



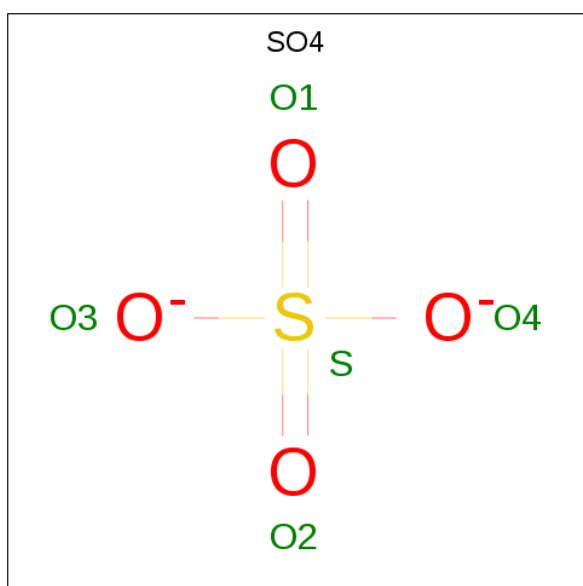
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			4	2	2		
2	A	1	Total	C	O	0	0
			4	2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			4	2	2		
2	A	1	Total	C	O	0	0
			4	2	2		
2	C	1	Total	C	O	0	0
			4	2	2		
2	D	1	Total	C	O	0	0
			4	2	2		
2	D	1	Total	C	O	0	0
			4	2	2		

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	O	S	0	0
			5	4	1		
3	C	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is water.

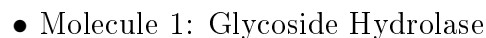
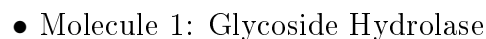
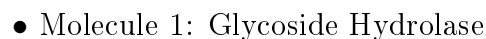
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	186	Total	O	0	0
			186	186		
4	B	207	Total	O	0	0
			207	207		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	151	Total 151	O 151	0	0
4	D	211	Total 211	O 211	0	0

- Molecule 1: Glycoside Hydrolase





4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	75.14Å 121.35Å 124.52Å 70.90° 87.47° 78.96°	Depositor
Resolution (Å)	117.64 – 2.50 67.79 – 2.50	Depositor EDS
% Data completeness (in resolution range)	87.5 (117.64-2.50) 87.5 (67.79-2.50)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.95 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.208 , 0.260 0.212 , 0.261	Depositor DCC
R_{free} test set	6086 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å ²)	29.1	Xtriage
Anisotropy	0.049	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 32.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	26748	wwPDB-VP
Average B, all atoms (Å ²)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.44% 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: SO4, EDO

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	0.41	0/6678	0.62	1/9068 (0.0%)
1	B	0.42	0/6664	0.62	0/9050
1	C	0.41	0/6617	0.61	0/8992
1	D	0.42	0/6700	0.63	0/9088
All	All	0.42	0/26659	0.62	1/36198 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	472	ARG	NE-CZ-NH2	5.36	122.98	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	6502	0	6149	15	0
1	B	6488	0	6134	13	0
1	C	6441	0	6061	12	0
1	D	6524	0	6217	14	0
2	A	16	0	24	0	0
2	C	4	0	6	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	8	0	12	0	0
3	B	5	0	0	0	0
3	C	5	0	0	0	0
4	A	186	0	0	0	0
4	B	207	0	0	2	0
4	C	151	0	0	1	0
4	D	211	0	0	1	0
All	All	26748	0	24603	54	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 (54) 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:C:780:ILE:HD13	1:C:811:THR:HG23	1.79	0.65
1:B:163:ALA:HB3	1:B:399:ILE:HG22	1.81	0.63
1:A:334:ASP:HB3	1:A:758:VAL:HG22	1.88	0.56
1:B:517:GLU:OE2	1:B:543:TYR:OH	2.24	0.55
1:C:818:LEU:CD2	1:C:820:LEU:HD11	2.38	0.54
1:A:665:HIS:CD2	1:A:775:TRP:HB3	2.45	0.51
1:B:678:ILE:HB	1:B:707:TYR:HB2	1.92	0.50
1:D:331:LEU:CD2	1:D:361:ALA:HB3	2.40	0.50
1:D:620:ASP:OD2	1:D:620:ASP:N	2.44	0.50
1:D:331:LEU:HD23	1:D:361:ALA:HB3	1.94	0.50
1:D:265:TRP:HB3	1:D:296:ARG:HG2	1.94	0.49
1:B:184:LYS:HB2	1:B:216:TYR:CZ	2.47	0.49
1:D:245:ILE:HD12	1:D:251:MET:HB2	1.94	0.49
1:A:594:ARG:NH1	1:A:804:TYR:OH	2.46	0.48
1:B:604:ALA:HB2	1:B:794:GLU:HG3	1.94	0.48
1:A:320:LEU:CD2	1:A:578:VAL:HG12	2.44	0.47
1:A:612:HIS:HD1	1:A:627:TYR:HH	1.61	0.47
1:D:34:PHE:CE1	1:D:179:LEU:HD22	2.49	0.47
1:A:678:ILE:HB	1:A:707:TYR:HB2	1.97	0.46
1:C:820:LEU:HD12	1:C:820:LEU:N	2.31	0.46
1:A:274:LEU:HD11	1:A:290:GLU:HB3	1.97	0.46
1:B:265:TRP:HB3	1:B:296:ARG:HG2	1.98	0.46
1:D:604:ALA:HB2	1:D:794:GLU:HG3	1.98	0.46
1:B:119:TYR:CD2	1:B:155:VAL:HG22	2.52	0.45
1:C:211:ARG:NH1	4:C:1001:HOH:O	2.44	0.45
1:A:334:ASP:CB	1:A:758:VAL:HG22	2.47	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:157:ILE:HD13	1:B:172:GLY:HA2	1.99	0.45
1:A:578:VAL:HG23	1:A:601:LEU:HB2	1.99	0.44
1:C:736:THR:HG22	1:C:771:ARG:CB	2.48	0.44
1:D:114:TYR:CZ	1:D:169:THR:HG21	2.53	0.44
1:C:678:ILE:HB	1:C:707:TYR:HB2	1.99	0.44
1:C:267:VAL:O	1:C:370:GLY:HA3	2.18	0.43
1:D:759:ILE:HD13	1:D:845:VAL:HG11	2.00	0.43
1:C:334:ASP:HB2	1:C:758:VAL:HG22	2.00	0.43
1:A:767:ASP:HB3	1:A:773:ILE:HD11	2.00	0.43
1:D:550:GLY:O	1:D:572:MET:HA	2.19	0.43
1:C:382:ASP:OD1	1:C:408:GLN:NE2	2.50	0.43
1:C:759:ILE:HD11	1:C:820:LEU:HD13	2.01	0.43
1:B:296:ARG:NH1	4:B:1002:HOH:O	2.51	0.43
1:B:338:ARG:NH2	1:B:604:ALA:O	2.52	0.42
1:A:632:ASN:O	1:A:646:LYS:NZ	2.52	0.42
1:A:612:HIS:ND1	1:A:627:TYR:OH	2.50	0.42
1:A:791:LEU:HD12	1:A:819:VAL:HG12	2.02	0.42
1:C:320:LEU:HD22	1:C:578:VAL:HG12	2.01	0.42
1:B:354:SER:HA	1:B:355:HIS:HA	1.84	0.42
1:B:74:GLU:HB3	1:B:76:THR:HG22	2.02	0.41
1:A:90:TRP:CZ2	1:A:156:ARG:HD2	2.56	0.41
1:D:296:ARG:NH2	4:D:1003:HOH:O	2.53	0.41
1:B:125:LYS:NZ	4:B:1004:HOH:O	2.54	0.41
1:C:759:ILE:HD12	1:C:761:VAL:HG23	2.02	0.41
1:D:665:HIS:CD2	1:D:775:TRP:HB3	2.56	0.41
1:D:331:LEU:HD22	1:D:359:ALA:CB	2.50	0.40
1:A:108:ILE:CD1	1:A:153:ILE:HD13	2.51	0.40
1:D:592:PRO:O	1:D:801:VAL:HG13	2.22	0.40

There are no symmetry-related clashes.

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	809/846 (96%)	775 (96%)	34 (4%)	0	100	100
1	B	808/846 (96%)	772 (96%)	36 (4%)	0	100	100
1	C	807/846 (95%)	759 (94%)	47 (6%)	1 (0%)	51	73
1	D	807/846 (95%)	779 (96%)	28 (4%)	0	100	100
All	All	3231/3384 (96%)	3085 (96%)	145 (4%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	233	ASN

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	689/738 (93%)	681 (99%)	8 (1%)	71	88
1	B	687/738 (93%)	683 (99%)	4 (1%)	86	95
1	C	676/738 (92%)	671 (99%)	5 (1%)	84	94
1	D	696/738 (94%)	682 (98%)	14 (2%)	55	79
All	All	2748/2952 (93%)	2717 (99%)	31 (1%)	73	89

All (31) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	67	TRP
1	A	230	ASP
1	A	331	LEU
1	A	404	CYS
1	A	429	PHE
1	A	460	PHE
1	A	586	THR
1	A	802	LEU
1	B	67	TRP
1	B	233	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	296	ARG
1	B	715	ASP
1	C	67	TRP
1	C	99	ASP
1	C	323	ASP
1	C	397	ASP
1	C	460	PHE
1	D	63	LEU
1	D	67	TRP
1	D	233	ASN
1	D	276	LYS
1	D	296	ARG
1	D	323	ASP
1	D	354	SER
1	D	468	THR
1	D	520	TYR
1	D	522	ASP
1	D	565	GLU
1	D	620	ASP
1	D	683	ASP
1	D	802	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	374	GLN
1	C	116	ASN
1	C	430	GLN

5.3.3 RNA ⓘ

There are no RNA molecules 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

9 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	$\# Z > 2$
2	EDO	D	901	-	3,3,3	0.42	0	2,2,2	0.44	0
2	EDO	A	902	-	3,3,3	0.52	0	2,2,2	0.10	0
2	EDO	C	902	-	3,3,3	0.44	0	2,2,2	0.29	0
3	SO4	C	901	-	4,4,4	0.34	0	6,6,6	0.14	0
2	EDO	A	903	-	3,3,3	0.47	0	2,2,2	0.28	0
2	EDO	A	904	-	3,3,3	0.58	0	2,2,2	0.13	0
2	EDO	D	902	-	3,3,3	0.52	0	2,2,2	0.25	0
3	SO4	B	901	-	4,4,4	0.38	0	6,6,6	0.14	0
2	EDO	A	901	-	3,3,3	0.52	0	2,2,2	0.13	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
2	EDO	D	901	-	-	1/1/1/1	-
2	EDO	A	902	-	-	1/1/1/1	-
2	EDO	C	902	-	-	0/1/1/1	-
2	EDO	A	903	-	-	0/1/1/1	-
2	EDO	A	904	-	-	1/1/1/1	-
2	EDO	D	902	-	-	0/1/1/1	-
2	EDO	A	901	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	904	EDO	O1-C1-C2-O2
2	D	901	EDO	O1-C1-C2-O2
2	A	902	EDO	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

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	811/846 (95%)	-0.06	7 (0%) 84 86	15, 29, 46, 58	0
1	B	812/846 (95%)	-0.09	5 (0%) 89 90	13, 29, 49, 73	0
1	C	811/846 (95%)	0.05	9 (1%) 80 82	15, 33, 57, 75	0
1	D	811/846 (95%)	-0.12	1 (0%) 95 96	13, 28, 42, 55	0
All	All	3245/3384 (95%)	-0.05	22 (0%) 87 89	13, 29, 50, 75	0

All (22) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	458	GLY	3.4
1	C	641	GLY	2.9
1	C	729	THR	2.8
1	B	452	SER	2.6
1	C	643	PHE	2.6
1	A	389	LEU	2.5
1	A	640	ASP	2.5
1	C	602	ASP	2.4
1	C	457	GLY	2.4
1	A	452	SER	2.4
1	B	726	GLY	2.3
1	C	458	GLY	2.3
1	B	647	ASP	2.2
1	C	621	ALA	2.2
1	B	639	LYS	2.2
1	D	661	ASP	2.1
1	A	32	SER	2.1
1	C	648	PRO	2.0
1	C	698	PHE	2.0
1	A	637	VAL	2.0
1	A	711	VAL	2.0
1	A	641	GLY	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 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, 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
2	EDO	A	904	4/4	0.63	0.24	45,47,49,49	0
2	EDO	A	902	4/4	0.81	0.26	47,48,49,50	0
2	EDO	A	903	4/4	0.82	0.23	46,46,48,50	0
2	EDO	A	901	4/4	0.85	0.22	43,43,44,44	0
2	EDO	C	902	4/4	0.87	0.19	30,30,30,31	0
2	EDO	D	902	4/4	0.89	0.16	36,37,37,38	0
2	EDO	D	901	4/4	0.93	0.18	30,30,31,31	0
3	SO4	C	901	5/5	0.94	0.14	64,64,65,65	0
3	SO4	B	901	5/5	0.95	0.16	46,47,48,49	0

6.5 Other polymers [i](#)

There are no such residues in this entry.