



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 21, 2020 – 10:47 PM BST

PDB ID : 6B6L
Title : The crystal structure of glycosyl hydrolase family 2 (GH2) member from *Bacteroides cellulosilyticus* DSM 14838
Authors : Tan, K.; Joachimiak, G.; Nocek, B.; Enddres, M.; Babnigg, G.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on : 2017-10-02
Resolution : 2.00 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 ⓘ 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 : 2.13.1
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) : 2.13.1

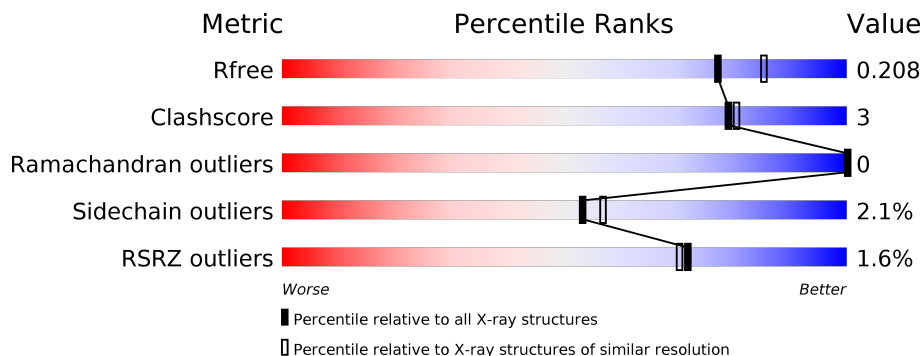
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	776	 2% 86% 8% 6%
1	B	776	 2% 86% 7% 6%

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 12209 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.

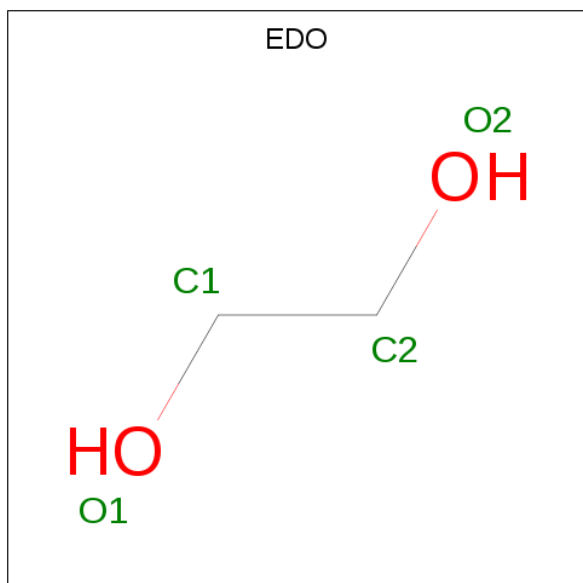
- Molecule 1 is a protein called Glycosyl hydrolase family 2, sugar binding domain protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	727	5725	3624	987	1095	9	10	0	3	0
1	B	726	5736	3637	981	1101	9	8	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

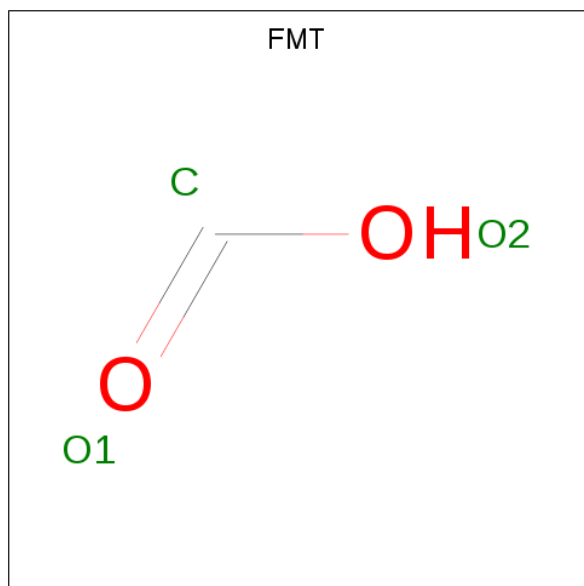
Chain	Residue	Modelled	Actual	Comment	Reference
A	22	SER	-	expression tag	UNP E2N721
A	23	ASN	ASP	conflict	UNP E2N721
A	24	ALA	SER	conflict	UNP E2N721
B	22	SER	-	expression tag	UNP E2N721
B	23	ASN	ASP	conflict	UNP E2N721
B	24	ALA	SER	conflict	UNP E2N721

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



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

- Molecule 3 is FORMIC ACID (three-letter code: FMT) (formula: CH₂O₂).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 3 1 2	0	0
3	A	1	Total C O 3 1 2	0	0
3	B	1	Total C O 3 1 2	0	0

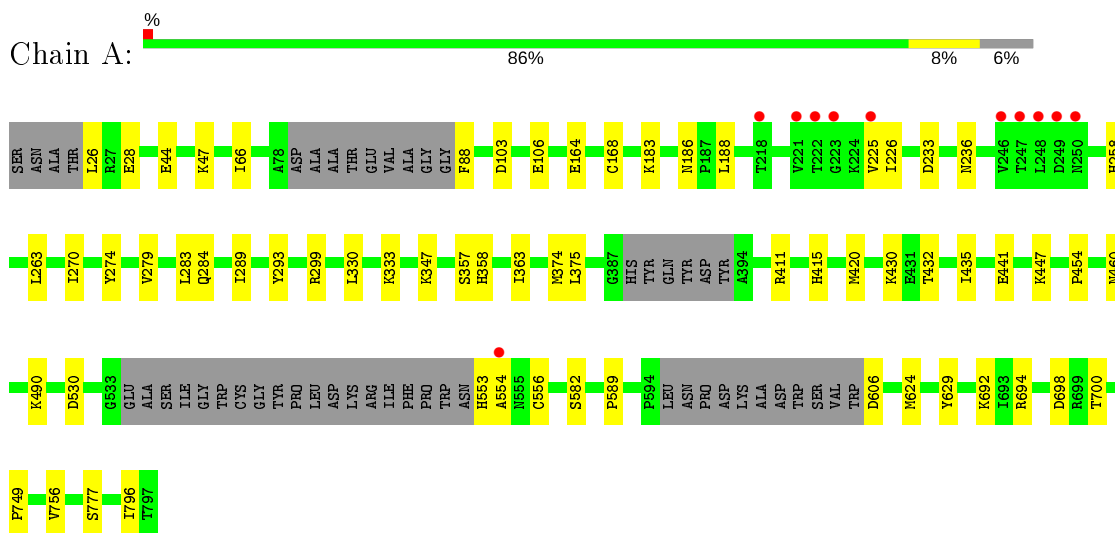
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	409	Total O 409 409	0	0
4	B	318	Total O 318 318	0	0

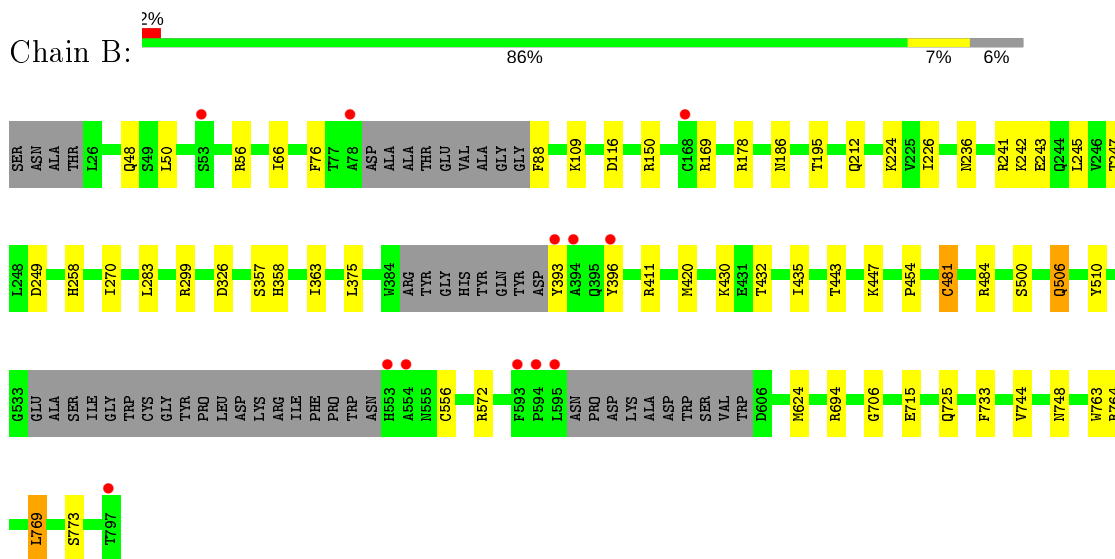
3 Residue-property plots [i](#)

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 hydrolase family 2, sugar binding domain protein



- Molecule 1: Glycosyl hydrolase family 2, sugar binding domain protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	57.53Å 227.13Å 66.90Å 90.00° 108.83° 90.00°	Depositor
Resolution (Å)	34.88 – 2.00 48.86 – 1.99	Depositor EDS
% Data completeness (in resolution range)	96.6 (34.88-2.00) 96.6 (48.86-1.99)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.79 (at 1.98Å)	Xtrriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.163 , 0.208 0.163 , 0.208	Depositor DCC
R_{free} test set	5319 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å ²)	24.2	Xtrriage
Anisotropy	0.014	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 47.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	12209	wwPDB-VP
Average B, all atoms (Å ²)	32.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.33% 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: CSD, FMT, 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	0.69	0/5828	0.74	2/7903 (0.0%)
1	B	0.63	0/5841	0.70	0/7919
All	All	0.66	0/11669	0.72	2/15822 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	698	ASP	CB-CG-OD1	6.27	123.94	118.30
1	A	447	LYS	CD-CE-NZ	-5.17	99.82	111.70

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	5725	0	5519	31	0
1	B	5736	0	5549	32	0
2	A	8	0	12	0	0
2	B	4	0	6	1	0
3	A	6	0	2	0	0
3	B	3	0	1	0	0
4	A	409	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	318	0	0	0	0
All	All	12209	0	11089	63	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 3.

The worst 5 of 63 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:B:744:VAL:HG12	1:B:769:LEU:HD22	1.68	0.75
1:A:420:MSE:HG2	1:A:454:PRO:HG2	1.69	0.73
1:A:28:GLU:HG2	1:A:183:LYS:HG2	1.73	0.70
1:A:363:ILE:HD11	1:A:411:ARG:HD2	1.76	0.67
1:B:481:CSD:OD1	1:B:484:ARG:HD3	1.96	0.66

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	718/776 (92%)	704 (98%)	14 (2%)	0	100	100
1	B	714/776 (92%)	698 (98%)	16 (2%)	0	100	100
All	All	1432/1552 (92%)	1402 (98%)	30 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	604/646 (94%)	592 (98%)	12 (2%)	55	58
1	B	610/646 (94%)	596 (98%)	14 (2%)	50	53
All	All	1214/1292 (94%)	1188 (98%)	26 (2%)	53	57

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	777	SER
1	B	169	ARG
1	B	556	CYS
1	B	50	LEU
1	B	56	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	189	HIS

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

4 non-standard protein/DNA/RNA residues 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	$\# Z > 2$
1	CSD	B	481	1	3,7,8	0.94	0	1,8,10	4.34	1 (100%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
1	CSD	A	110	1	3,7,8	0.75	0	1,8,10	0.23	0
1	CSD	B	110	1	3,7,8	1.04	0	1,8,10	0.16	0
1	CSD	A	481	1	3,7,8	1.01	0	1,8,10	1.20	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
1	CSD	B	481	1	-	2/2/6/8	-
1	CSD	A	110	1	-	0/2/6/8	-
1	CSD	B	110	1	-	0/2/6/8	-
1	CSD	A	481	1	-	1/2/6/8	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	481	CSD	OD1-SG-CB	-4.34	97.27	105.54

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	B	481	CSD	N-CA-CB-SG
1	B	481	CSD	CA-CB-SG-OD1
1	A	481	CSD	N-CA-CB-SG

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	481	CSD	1	0

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry

6 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	B	801	-	3,3,3	0.74	0	2,2,2	0.54	0
2	EDO	A	801	-	3,3,3	0.59	0	2,2,2	0.76	0
2	EDO	A	802	-	3,3,3	0.76	0	2,2,2	0.51	0
3	FMT	A	804	-	0,2,2	0.00	-	0,1,1	0.00	-
3	FMT	A	803	-	0,2,2	0.00	-	0,1,1	0.00	-
3	FMT	B	802	-	0,2,2	0.00	-	0,1,1	0.00	-

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	B	801	-	-	0/1/1/1	-
2	EDO	A	802	-	-	0/1/1/1	-
2	EDO	A	801	-	-	0/1/1/1	-

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 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	801	EDO	1	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	716/776 (92%)	-0.40	11 (1%) 73 72	12, 27, 56, 95	0
1	B	715/776 (92%)	-0.40	12 (1%) 70 68	18, 32, 58, 107	0
All	All	1431/1552 (92%)	-0.40	23 (1%) 72 70	12, 30, 58, 107	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	553	HIS	8.5
1	B	595	LEU	6.0
1	B	554	ALA	5.1
1	B	393	TYR	4.1
1	B	594	PRO	3.4

6.2 Non-standard residues in protein, DNA, RNA chains [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
1	CSD	B	481	8/9	0.92	0.16	25,29,55,70	0
1	CSD	A	481	8/9	0.96	0.09	23,26,49,63	0
1	CSD	B	110	8/9	0.97	0.09	35,41,54,55	0
1	CSD	A	110	8/9	0.97	0.06	36,43,65,67	0

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
3	FMT	A	804	3/3	0.85	0.12	49,49,53,54	0
2	EDO	A	801	4/4	0.91	0.14	35,36,39,40	0
2	EDO	B	801	4/4	0.97	0.09	20,21,22,29	0
3	FMT	A	803	3/3	0.97	0.08	30,30,32,33	0
3	FMT	B	802	3/3	0.97	0.11	25,25,37,43	0
2	EDO	A	802	4/4	0.99	0.10	16,18,22,29	0

6.5 Other polymers [i](#)

There are no such residues in this entry.