



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 23, 2024 – 01:33 PM EDT

PDB ID : 5M1I  
Title : Structure of GH36 alpha-galactosidase from *Thermotoga maritima* in a covalent complex with a cyclopropyl carbasugar.  
Authors : Pengelly, R.; Gloster, T.  
Deposited on : 2016-10-07  
Resolution : 1.55 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](mailto: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

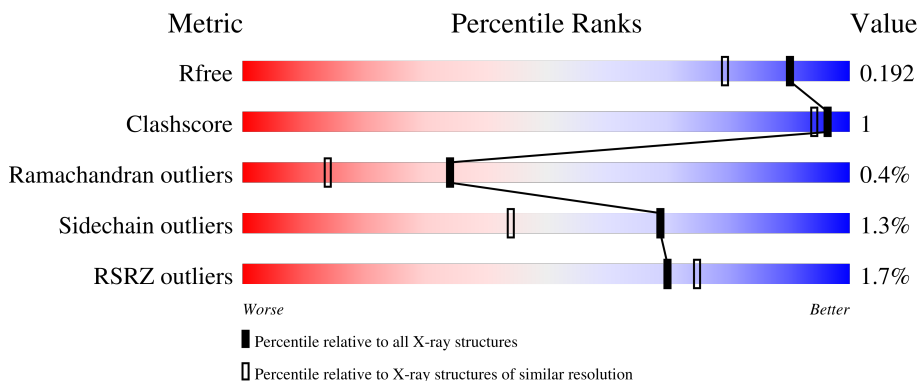
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.55 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	1483 (1.56-1.56)
Clashscore	141614	1529 (1.56-1.56)
Ramachandran outliers	138981	1498 (1.56-1.56)
Sidechain outliers	138945	1495 (1.56-1.56)
RSRZ outliers	127900	1465 (1.56-1.56)

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  $\geq 3$ , 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	575	

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 4931 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 Alpha-galactosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	539	4478	2883	749	827	19	0	16	0

There are 23 discrepancies between the modelled and reference sequences:

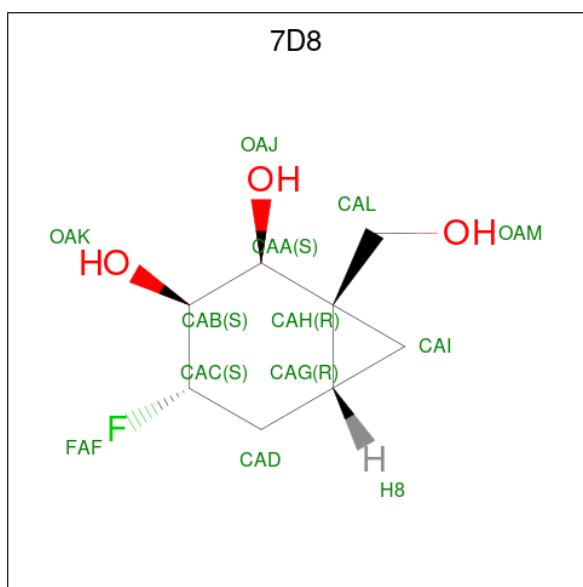
Chain	Residue	Modelled	Actual	Comment	Reference
A	-22	MET	-	initiating methionine	UNP O33835
A	-21	GLY	-	expression tag	UNP O33835
A	-20	SER	-	expression tag	UNP O33835
A	-19	SER	-	expression tag	UNP O33835
A	-18	HIS	-	expression tag	UNP O33835
A	-17	HIS	-	expression tag	UNP O33835
A	-16	HIS	-	expression tag	UNP O33835
A	-15	HIS	-	expression tag	UNP O33835
A	-14	HIS	-	expression tag	UNP O33835
A	-13	HIS	-	expression tag	UNP O33835
A	-12	SER	-	expression tag	UNP O33835
A	-11	SER	-	expression tag	UNP O33835
A	-10	GLY	-	expression tag	UNP O33835
A	-9	LEU	-	expression tag	UNP O33835
A	-8	VAL	-	expression tag	UNP O33835
A	-7	PRO	-	expression tag	UNP O33835
A	-6	ARG	-	expression tag	UNP O33835
A	-5	GLY	-	expression tag	UNP O33835
A	-4	SER	-	expression tag	UNP O33835
A	-3	HIS	-	expression tag	UNP O33835
A	-2	MET	-	expression tag	UNP O33835
A	-1	ALA	-	expression tag	UNP O33835
A	0	SER	-	expression tag	UNP O33835

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0

- Molecule 3 is (1 {R},2 {S},3 {S},4 {S},6 {R})-4-fluoranyl-1-(hydroxymethyl)bicyclo[4.1.0]heptane-2,3-diol (three-letter code: 7D8) (formula: C<sub>8</sub>H<sub>13</sub>FO<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	
			Total	C	F			O
3	A	1	12	8	1	3	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



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

*Continued on next page...*

*Continued from previous page...*

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

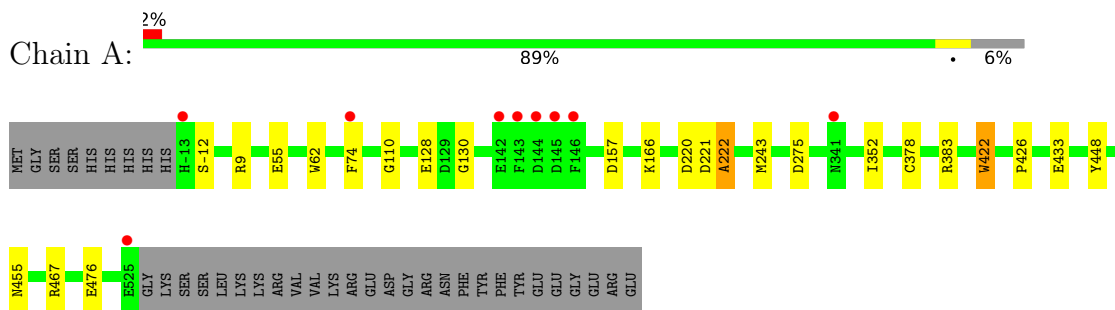
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	360	Total O 363 363	0	3

### 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: Alpha-galactosidase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	66.90Å 96.24Å 97.59Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	68.52 – 1.55 48.79 – 1.55	Depositor EDS
% Data completeness (in resolution range)	99.9 (68.52-1.55) 99.9 (48.79-1.55)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.19 (at 1.55Å)	Xtrriage
Refinement program	REFMAC 5.8.0155	Depositor
R, $R_{free}$	0.154 , 0.183 0.168 , 0.192	Depositor DCC
$R_{free}$ test set	4571 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	17.8	Xtrriage
Anisotropy	0.741	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 43.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	0.000 for -h,l,k	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	4931	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.62% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: EDO, 7D8, SO4

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	0.76	0/4600	0.84	2/6235 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	243	MET	CG-SD-CE	-8.99	85.81	100.20
1	A	275	ASP	CB-CG-OD1	6.92	124.52	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	383	ARG	Sidechain

### 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	4478	0	4372	9	0
2	A	30	0	0	0	0
3	A	12	0	0	0	0
4	A	48	0	72	2	0
5	A	363	0	0	0	0
All	All	4931	0	4444	9	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 (9) 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:352[A]:ILE:HG12	1:A:378[A]:CYS:SG	2.48	0.54
1:A:426:PRO:HD3	1:A:455:ASN:HB3	1.93	0.51
1:A:-12:SER:H	4:A:614:EDO:H12	1.79	0.48
1:A:221:ASP:O	1:A:222:ALA:HB2	2.15	0.46
1:A:448:TYR:OH	1:A:476:GLU:OE1	2.26	0.45
1:A:128:GLU:OE2	4:A:619:EDO:C2	2.67	0.42
1:A:433:GLU:OE1	1:A:467:ARG:HB2	2.18	0.42
1:A:74:PHE:HE1	1:A:130:GLY:HA2	1.85	0.41
1:A:110:GLY:HA2	1:A:157:ASP:O	2.21	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	553/575 (96%)	534 (97%)	17 (3%)	2 (0%)	34 14

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	222	ALA
1	A	422	TRP

### 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	485/505 (96%)	478 (99%)	7 (1%)	67 41

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

Mol	Chain	Res	Type
1	A	9	ARG
1	A	55	GLU
1	A	62	TRP
1	A	166[A]	LYS
1	A	166[B]	LYS
1	A	220	ASP
1	A	422	TRP

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

Mol	Chain	Res	Type
1	A	99	ASN

### 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](#)

19 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	SO4	A	604	-	4,4,4	0.52	0	6,6,6	1.47	1 (16%)
4	EDO	A	612	-	3,3,3	0.20	0	2,2,2	1.16	0
2	SO4	A	601	-	4,4,4	0.54	0	6,6,6	1.09	1 (16%)
4	EDO	A	608	-	3,3,3	1.30	0	2,2,2	0.07	0
4	EDO	A	616	-	3,3,3	0.18	0	2,2,2	0.38	0
4	EDO	A	613	-	3,3,3	0.46	0	2,2,2	0.44	0
4	EDO	A	614	-	3,3,3	0.43	0	2,2,2	0.31	0
4	EDO	A	618	-	3,3,3	0.43	0	2,2,2	0.52	0
2	SO4	A	603	-	4,4,4	0.48	0	6,6,6	1.66	1 (16%)
4	EDO	A	609	-	3,3,3	0.45	0	2,2,2	0.39	0
4	EDO	A	615	-	3,3,3	0.51	0	2,2,2	0.35	0
2	SO4	A	602	-	4,4,4	0.63	0	6,6,6	0.92	0
3	7D8	A	607	1	13,13,13	1.54	1 (7%)	11,21,21	2.40	2 (18%)
4	EDO	A	611	-	3,3,3	0.26	0	2,2,2	0.67	0
2	SO4	A	605	-	4,4,4	0.60	0	6,6,6	0.44	0
4	EDO	A	617	-	3,3,3	0.52	0	2,2,2	0.44	0
2	SO4	A	606	-	4,4,4	0.40	0	6,6,6	0.29	0
4	EDO	A	619	-	3,3,3	0.48	0	2,2,2	0.21	0
4	EDO	A	610	-	3,3,3	0.35	0	2,2,2	0.57	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	EDO	A	612	-	-	0/1/1/1	-
4	EDO	A	609	-	-	0/1/1/1	-
4	EDO	A	618	-	-	1/1/1/1	-
4	EDO	A	617	-	-	1/1/1/1	-
4	EDO	A	608	-	-	0/1/1/1	-
4	EDO	A	615	-	-	1/1/1/1	-
4	EDO	A	619	-	-	0/1/1/1	-
4	EDO	A	616	-	-	1/1/1/1	-
4	EDO	A	613	-	-	1/1/1/1	-
4	EDO	A	610	-	-	0/1/1/1	-
4	EDO	A	614	-	-	1/1/1/1	-
4	EDO	A	611	-	-	1/1/1/1	-
3	7D8	A	607	1	-	3/3/29/29	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	607	7D8	CAH-CAA	-4.16	1.46	1.55

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	607	7D8	CAD-CAC-CAB	6.61	118.83	111.98
2	A	604	SO4	O4-S-O3	2.89	121.41	109.06
2	A	603	SO4	O4-S-O3	2.85	121.22	109.06
3	A	607	7D8	OAJ-CAA-CAB	-2.58	104.31	110.03
2	A	601	SO4	O4-S-O2	-2.13	98.22	109.31

There are no chirality outliers.

All (10) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	607	7D8	CAG-CAH-CAL-OAM
4	A	614	EDO	O1-C1-C2-O2
4	A	618	EDO	O1-C1-C2-O2
3	A	607	7D8	CAA-CAH-CAL-OAM
4	A	616	EDO	O1-C1-C2-O2
4	A	611	EDO	O1-C1-C2-O2
4	A	613	EDO	O1-C1-C2-O2
4	A	615	EDO	O1-C1-C2-O2
4	A	617	EDO	O1-C1-C2-O2
3	A	607	7D8	CAI-CAH-CAL-OAM

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	614	EDO	1	0
4	A	619	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, 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	539/575 (93%)	-0.21	9 (1%) 70 75	14, 20, 38, 52	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	341	ASN	3.3
1	A	144	ASP	3.2
1	A	525	GLU	2.6
1	A	145	ASP	2.5
1	A	74	PHE	2.5
1	A	142	GLU	2.3
1	A	143	PHE	2.2
1	A	-13	HIS	2.2
1	A	146	PHE	2.1

### 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, 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	B-factors( $\text{\AA}^2$ )	Q<0.9
4	EDO	A	615	4/4	0.69	0.16	32,40,44,50	0
4	EDO	A	614	4/4	0.82	0.16	36,39,40,47	0
4	EDO	A	613	4/4	0.84	0.14	33,38,41,45	0
4	EDO	A	618	4/4	0.84	0.14	46,50,51,57	0
4	EDO	A	616	4/4	0.87	0.15	37,41,44,45	0
4	EDO	A	611	4/4	0.88	0.12	36,37,45,46	0
4	EDO	A	612	4/4	0.92	0.15	21,33,41,44	0
4	EDO	A	619	4/4	0.92	0.22	33,41,46,51	0
2	SO4	A	604	5/5	0.93	0.20	35,39,52,52	0
2	SO4	A	606	5/5	0.93	0.36	42,63,72,74	0
4	EDO	A	617	4/4	0.94	0.08	40,43,45,47	0
4	EDO	A	608	4/4	0.96	0.08	18,20,20,22	0
2	SO4	A	603	5/5	0.96	0.09	35,40,41,46	0
3	7D8	A	607	12/12	0.97	0.07	15,16,16,20	0
2	SO4	A	605	5/5	0.97	0.18	33,33,36,41	5
4	EDO	A	610	4/4	0.97	0.06	19,19,22,22	0
2	SO4	A	602	5/5	0.99	0.14	27,28,31,42	0
4	EDO	A	609	4/4	0.99	0.06	19,21,23,23	0
2	SO4	A	601	5/5	0.99	0.05	23,24,25,30	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.