



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 22, 2024 – 02:39 PM EDT

PDB ID : 6EZT  
Title : Crystal structure of GH20 Exo beta-N-Acetylglucosaminidase D437A inactive mutant from *Vibrio harveyi*  
Authors : Porfetye, A.T.; Meekrathok, P.; Burger, M.; Vetter, I.R.; Suginta, W.  
Deposited on : 2017-11-16  
Resolution : 2.60 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 : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 1.20.1  
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

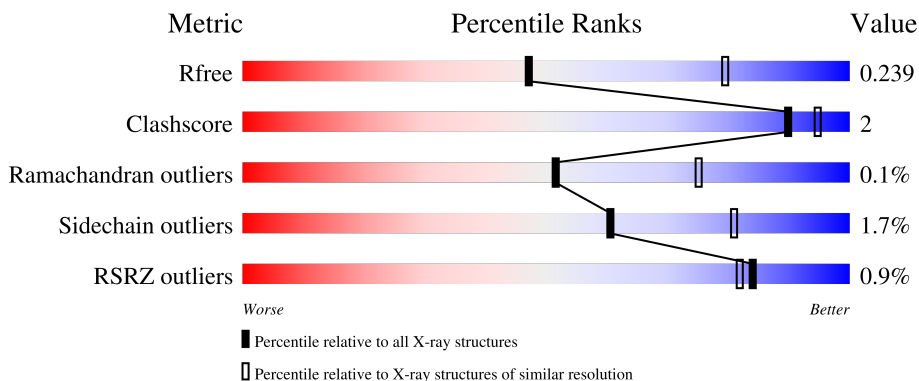
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

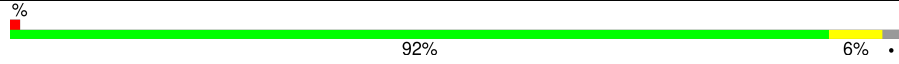
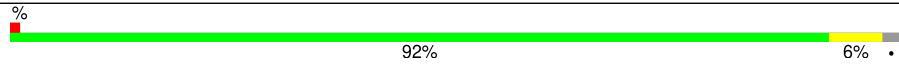
The reported resolution of this entry is 2.60 Å.

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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	649	 92% 6%
1	B	649	 92% 6%

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 10764 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 Beta-N-acetylglucosaminidase Nag2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	639	5145	3275	879	969	22	0	1	0
1	B	639	5155	3281	881	971	22	0	2	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	437	ALA	ASP	engineered mutation	UNP D9ISE0
A	643	ARG	-	expression tag	UNP D9ISE0
A	644	SER	-	expression tag	UNP D9ISE0
A	645	ARG	-	expression tag	UNP D9ISE0
A	646	SER	-	expression tag	UNP D9ISE0
A	647	HIS	-	expression tag	UNP D9ISE0
A	648	HIS	-	expression tag	UNP D9ISE0
A	649	HIS	-	expression tag	UNP D9ISE0
A	650	HIS	-	expression tag	UNP D9ISE0
A	651	HIS	-	expression tag	UNP D9ISE0
A	652	HIS	-	expression tag	UNP D9ISE0
B	437	ALA	ASP	engineered mutation	UNP D9ISE0
B	643	ARG	-	expression tag	UNP D9ISE0
B	644	SER	-	expression tag	UNP D9ISE0
B	645	ARG	-	expression tag	UNP D9ISE0
B	646	SER	-	expression tag	UNP D9ISE0
B	647	HIS	-	expression tag	UNP D9ISE0
B	648	HIS	-	expression tag	UNP D9ISE0
B	649	HIS	-	expression tag	UNP D9ISE0
B	650	HIS	-	expression tag	UNP D9ISE0
B	651	HIS	-	expression tag	UNP D9ISE0
B	652	HIS	-	expression tag	UNP D9ISE0

- Molecule 2 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: C<sub>8</sub>H<sub>18</sub>O<sub>5</sub>).



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

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula:  $C_4H_{10}O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			7	4	3		
3	B	1	Total	C	O	0	0
			7	4	3		

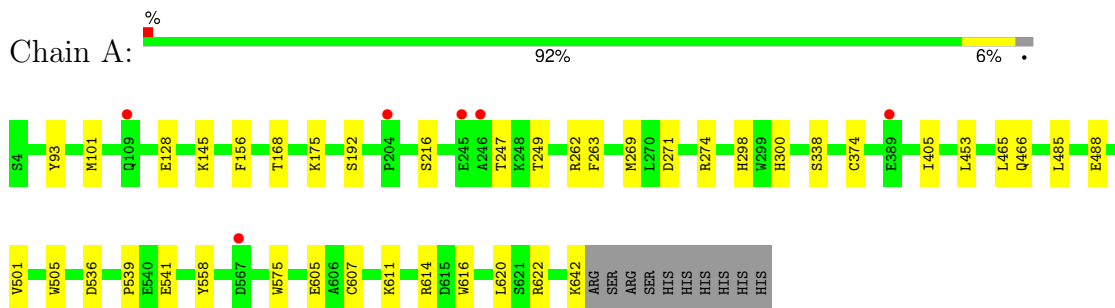
- Molecule 4 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	A	215	Total 215	O 215	0	0
4	B	209	Total 209	O 209	0	0

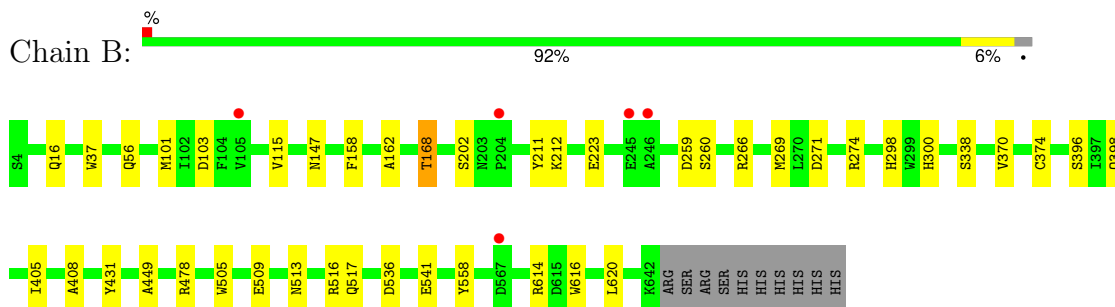
### 3 Residue-property plots [i](#)

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: Beta-N-acetylglucosaminidase Nag2



- Molecule 1: Beta-N-acetylglucosaminidase Nag2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	89.42Å 129.24Å 98.36Å 90.00° 112.20° 90.00°	Depositor
Resolution (Å)	91.09 – 2.60 49.65 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.9 (91.09-2.60) 99.9 (49.65-2.60)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.25 (at 2.61Å)	Xtrriage
Refinement program	REFMAC 5.8.0103	Depositor
R, $R_{free}$	0.195 , 0.236 0.200 , 0.239	Depositor DCC
$R_{free}$ test set	3180 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	53.1	Xtrriage
Anisotropy	0.661	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 35.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	10764	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	64.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.06% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: PEG, PG4

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	0.38	0/5285	0.63	0/7187
1	B	0.39	0/5295	0.64	0/7200
All	All	0.39	0/10580	0.63	0/14387

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-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 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	5145	0	4985	18	0
1	B	5155	0	4994	18	0
2	A	13	0	18	0	0
2	B	13	0	18	0	0
3	A	7	0	10	0	0
3	B	7	0	10	1	0
4	A	215	0	0	1	0
4	B	209	0	0	0	0
All	All	10764	0	10035	36	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 2.



All (36) 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:509:GLU:OE2	1:B:516:ARG:NH2	2.36	0.59
1:B:558:TYR:O	1:B:614:ARG:NH2	2.40	0.54
1:A:466:GLN:NE2	4:A:802:HOH:O	2.41	0.52
1:B:271:ASP:HA	1:B:300:HIS:HB3	1.91	0.51
1:B:266:ARG:HH12	3:B:701:PEG:H22	1.75	0.50
1:A:101:MET:HE1	1:A:338:SER:HA	1.93	0.50
1:A:269:MET:HG3	1:A:298:HIS:CG	2.46	0.50
1:A:536:ASP:HB3	1:A:541:GLU:HG3	1.94	0.50
1:B:211:TYR:CE1	1:B:259:ASP:HB3	2.46	0.50
1:B:269:MET:HG3	1:B:298:HIS:CG	2.47	0.49
1:A:616:TRP:CZ2	1:A:620:LEU:HD11	2.47	0.49
1:A:93:TYR:CE2	1:A:539:PRO:HB2	2.48	0.48
1:B:212:LYS:HB2	1:B:223:GLU:HB2	1.96	0.48
1:B:536:ASP:HB3	1:B:541:GLU:HG3	1.95	0.48
1:A:271:ASP:HA	1:A:300:HIS:HB3	1.96	0.47
1:B:370:VAL:HG12	1:B:405:ILE:HD12	1.97	0.46
1:A:262:ARG:HG2	1:A:263:PHE:CD2	2.52	0.45
1:B:101:MET:HE1	1:B:338:SER:HA	1.99	0.45
1:A:247:THR:HG22	1:A:249:THR:HG23	1.98	0.44
1:A:269:MET:HG3	1:A:298:HIS:CD2	2.52	0.44
1:A:156:PHE:HB2	1:A:216:SER:O	2.17	0.44
1:B:37:TRP:HA	1:B:103:ASP:O	2.19	0.43
1:A:485:LEU:HD22	1:A:501:VAL:HB	2.01	0.43
1:B:513:ASN:O	1:B:517:GLN:HG2	2.19	0.43
1:A:145:LYS:HB2	1:A:622:ARG:CZ	2.49	0.42
1:A:453:LEU:HD23	1:A:465:LEU:HD22	2.02	0.42
1:A:605:GLU:HG2	1:A:614:ARG:HD3	2.02	0.42
1:B:616:TRP:CZ2	1:B:620:LEU:HD11	2.54	0.42
1:B:147:ASN:ND2	1:B:260:SER:O	2.53	0.41
1:B:158:PHE:HB2	1:B:162:ALA:HB2	2.02	0.41
1:B:168:THR:CG2	1:B:202:SER:HB2	2.50	0.41
1:A:558:TYR:O	1:A:614:ARG:NH2	2.53	0.41
1:A:175:LYS:NZ	1:A:192:SER:OG	2.54	0.41
1:A:611:LYS:HA	1:A:614:ARG:HG3	2.03	0.41
1:B:101:MET:HA	1:B:115:VAL:HG23	2.03	0.41
1:B:408:ALA:HB1	1:B:449:ALA:HB1	2.04	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	638/649 (98%)	622 (98%)	15 (2%)	1 (0%)	47	71
1	B	639/649 (98%)	623 (98%)	16 (2%)	0	100	100
All	All	1277/1298 (98%)	1245 (98%)	31 (2%)	1 (0%)	51	75

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	488	GLU

### 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	556/565 (98%)	547 (98%)	9 (2%)	62	82
1	B	557/565 (99%)	546 (98%)	11 (2%)	55	78
All	All	1113/1130 (98%)	1093 (98%)	20 (2%)	60	80

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	128	GLU
1	A	168	THR
1	A	274	ARG
1	A	374	CYS
1	A	405	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	505	TRP
1	A	575	TRP
1	A	607	CYS
1	A	642	LYS
1	B	16	GLN
1	B	56[A]	GLN
1	B	56[B]	GLN
1	B	168	THR
1	B	274	ARG
1	B	374	CYS
1	B	396	SER
1	B	398	GLN
1	B	431	TYR
1	B	478	ARG
1	B	505	TRP

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](#)

4 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	$\# Z  > 2$
3	PEG	A	702	-	6,6,6	0.52	0	5,5,5	0.28	0
2	PG4	B	702	-	12,12,12	0.50	0	11,11,11	0.27	0
3	PEG	B	701	-	6,6,6	0.60	0	5,5,5	0.45	0
2	PG4	A	701	-	12,12,12	0.52	0	11,11,11	0.29	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. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	PEG	A	702	-	-	3/4/4/4	-
2	PG4	B	702	-	-	5/10/10/10	-
3	PEG	B	701	-	-	2/4/4/4	-
2	PG4	A	701	-	-	5/10/10/10	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (15) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	702	PEG	O1-C1-C2-O2
2	A	701	PG4	O2-C3-C4-O3
3	B	701	PEG	O2-C3-C4-O4
2	B	702	PG4	O1-C1-C2-O2
3	B	701	PEG	O1-C1-C2-O2
2	A	701	PG4	O4-C7-C8-O5
2	A	701	PG4	O1-C1-C2-O2
2	A	701	PG4	C8-C7-O4-C6
3	A	702	PEG	C1-C2-O2-C3
2	B	702	PG4	C5-C6-O4-C7
2	B	702	PG4	O2-C3-C4-O3
2	A	701	PG4	C3-C4-O3-C5
3	A	702	PEG	C4-C3-O2-C2
2	B	702	PG4	C6-C5-O3-C4
2	B	702	PG4	O3-C5-C6-O4

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	701	PEG	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	639/649 (98%)	-0.20	6 (0%) 84 82	44, 62, 87, 127	0
1	B	639/649 (98%)	-0.20	5 (0%) 86 84	45, 63, 88, 122	0
All	All	1278/1298 (98%)	-0.20	11 (0%) 84 82	44, 63, 88, 127	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	246	ALA	6.1
1	B	246	ALA	4.5
1	B	567	ASP	3.3
1	A	245	GLU	2.9
1	B	105	VAL	2.7
1	B	245	GLU	2.6
1	B	204	PRO	2.5
1	A	567	ASP	2.5
1	A	389	GLU	2.1
1	A	109	GLN	2.1
1	A	204	PRO	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, 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
3	PEG	B	701	7/7	0.67	0.38	72,73,74,74	0
2	PG4	B	702	13/13	0.81	0.23	95,98,105,105	0
3	PEG	A	702	7/7	0.82	0.30	95,97,98,98	0
2	PG4	A	701	13/13	0.82	0.25	92,97,101,101	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.