



Full wwPDB X-ray Structure Validation Report ⓘ

May 28, 2020 – 03:04 am BST

PDB ID : 5XXL
Title : Crystal structure of GH3 beta-glucosidase from *Bacteroides thetaiotaomicron*
Authors : Nakajima, M.; Ishiguro, R.; Tanaka, N.; Abe, K.; Maeda, T.; Miyanaga, A.;
Takahashi, Y.; Sugimoto, N.; Nakai, H.; Taguchi, H.
Deposited on : 2017-07-04
Resolution : 1.60 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

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

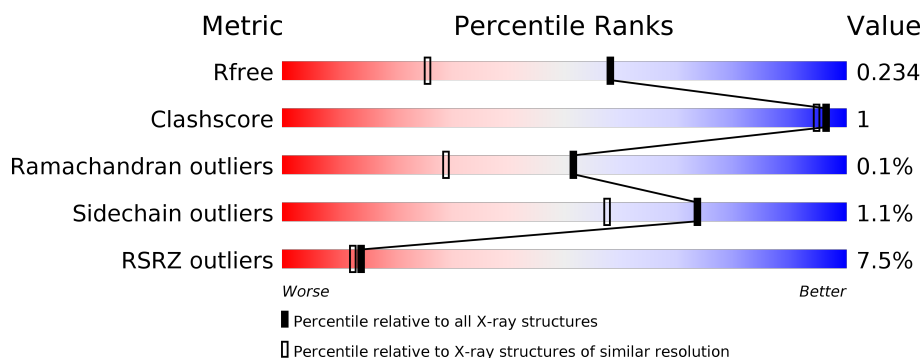
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.60 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	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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 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	760	<div> <div>4%</div> <div>95%</div> </div>
1	B	760	<div> <div>11%</div> <div>93%</div> </div>

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 12861 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 Periplasmic beta-glucosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	745	Total	C	N	O	S	0	10	0
			5847	3699	990	1120	38			
1	B	731	Total	C	N	O	S	0	8	0
			5745	3639	973	1097	36			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	MET	-	expression tag	UNP Q8A1U1
A	772	LEU	-	expression tag	UNP Q8A1U1
A	773	GLU	-	expression tag	UNP Q8A1U1
A	774	HIS	-	expression tag	UNP Q8A1U1
A	775	HIS	-	expression tag	UNP Q8A1U1
A	776	HIS	-	expression tag	UNP Q8A1U1
A	777	HIS	-	expression tag	UNP Q8A1U1
A	778	HIS	-	expression tag	UNP Q8A1U1
A	779	HIS	-	expression tag	UNP Q8A1U1
B	20	MET	-	expression tag	UNP Q8A1U1
B	772	LEU	-	expression tag	UNP Q8A1U1
B	773	GLU	-	expression tag	UNP Q8A1U1
B	774	HIS	-	expression tag	UNP Q8A1U1
B	775	HIS	-	expression tag	UNP Q8A1U1
B	776	HIS	-	expression tag	UNP Q8A1U1
B	777	HIS	-	expression tag	UNP Q8A1U1
B	778	HIS	-	expression tag	UNP Q8A1U1
B	779	HIS	-	expression tag	UNP Q8A1U1

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Mg	0	0
			1	1		

Continued on next page...

Continued from previous page...

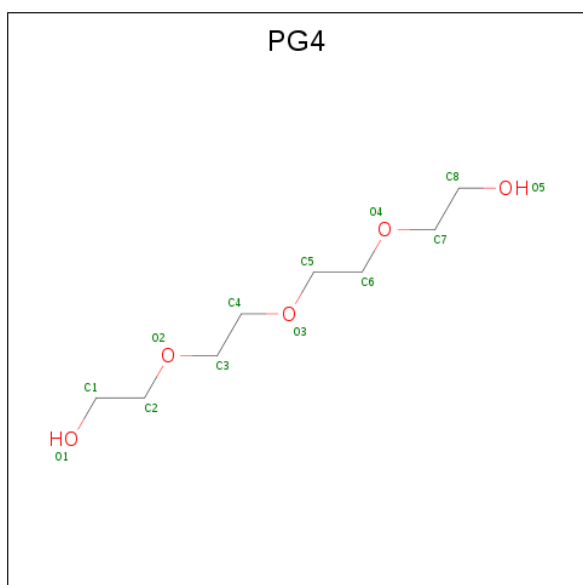
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Mg	0	0
			1	1		

- 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		

- Molecule 4 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).



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

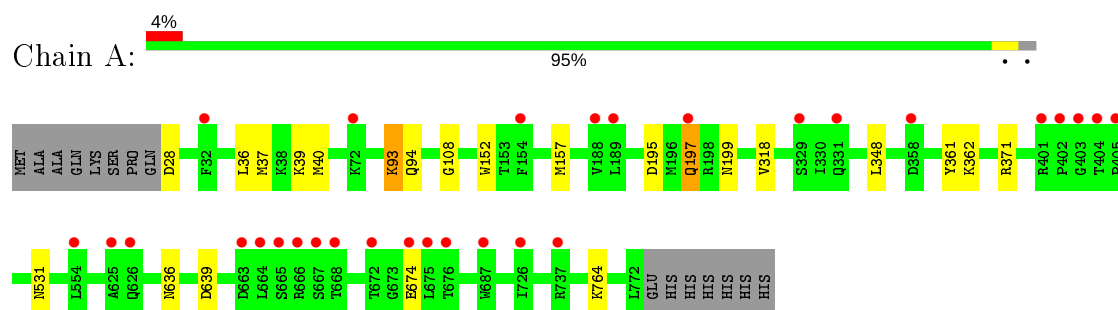
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	691	Total	O	0	0
			691	691		
5	B	556	Total	O	0	0
			556	556		

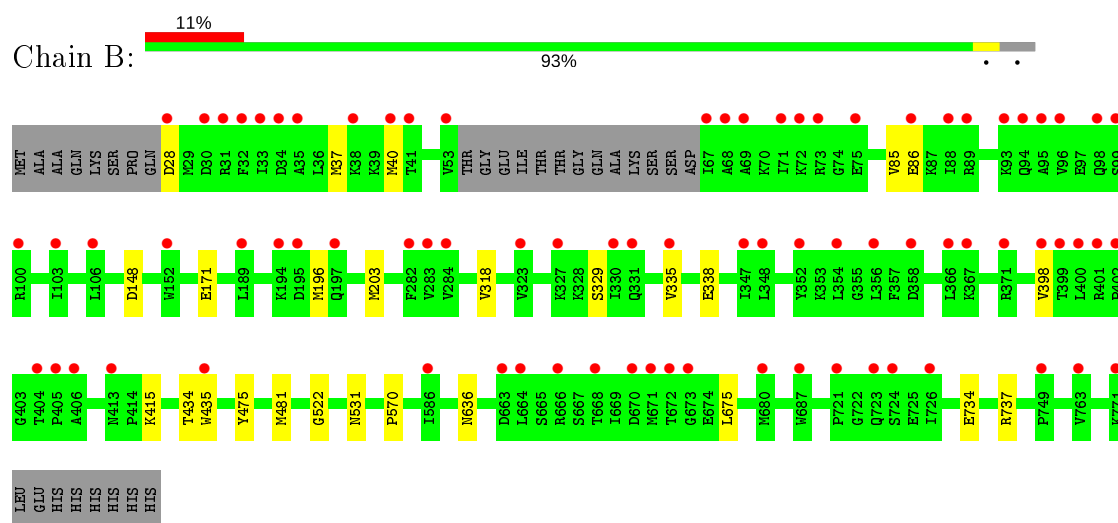
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: Periplasmic beta-glucosidase



- Molecule 1: Periplasmic beta-glucosidase



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	81.94Å 167.88Å 224.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.94 – 1.60 29.94 – 1.60	Depositor EDS
% Data completeness (in resolution range)	99.9 (29.94-1.60) 99.9 (29.94-1.60)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.94 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.8.0123	Depositor
R, R_{free}	0.204 , 0.227 0.213 , 0.234	Depositor DCC
R_{free} test set	10325 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å ²)	15.7	Xtriage
Anisotropy	0.084	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 41.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	12861	wwPDB-VP
Average B, all atoms (Å ²)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.74% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² 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: PG4, PEG, MG

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	0.43	0/5961	0.72	1/8065 (0.0%)
1	B	0.43	0/5860	0.74	0/7929
All	All	0.43	0/11821	0.73	1/15994 (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	371	ARG	NE-CZ-NH1	5.15	122.87	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	5847	0	5804	12	0
1	B	5745	0	5689	12	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	7	0	10	0	0
4	A	13	0	18	1	0
5	A	691	0	0	2	0
5	B	556	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	12861	0	11521	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:639:ASP:HB3	1:B:475:TYR:HE2	1.72	0.53
1:B:434:THR:HG23	1:B:435[A]:TRP:CD1	2.45	0.52
1:A:37:MET:HG2	1:A:40[B]:MET:HE3	1.93	0.51
1:B:435[A]:TRP:CZ2	1:B:522:GLY:HA2	2.48	0.48
1:B:85:VAL:HB	1:B:148:ASP:HA	1.97	0.46
1:B:86:GLU:CD	1:B:86:GLU:H	2.18	0.46
1:A:39:LYS:NZ	4:A:803:PG4:H61	2.31	0.45
1:B:481[A]:MET:HE3	1:B:481[A]:MET:HB3	1.49	0.45
1:B:196:MET:HE1	1:B:203[B]:MET:HB2	1.99	0.44
1:B:734:GLU:O	1:B:737:ARG:HD3	2.18	0.43
1:B:37:MET:HG2	1:B:40:MET:HE3	2.00	0.43
1:A:36:LEU:HG	1:A:40[B]:MET:CE	2.48	0.43
1:A:195:ASP:OD1	1:A:197:GLN:HB2	2.19	0.43
1:A:36:LEU:HG	1:A:40[B]:MET:HE2	2.00	0.43
1:A:199:ASN:HA	1:A:362:LYS:HE2	2.01	0.43
1:B:329:SER:HB3	1:B:335:VAL:HG12	2.01	0.42
1:A:108:GLY:HA2	1:A:152:TRP:O	2.20	0.42
1:A:39:LYS:NZ	5:A:903:HOH:O	2.42	0.42
1:B:171[B]:GLU:OE2	1:B:435[B]:TRP:NE1	2.53	0.42
1:A:40[B]:MET:HE3	1:A:348:LEU:HD11	2.02	0.41
1:B:398:VAL:CG2	1:B:570:PRO:HB3	2.51	0.41
1:A:28:ASP:N	5:A:916:HOH:O	2.54	0.41
1:A:93:LYS:HG2	1:A:361:TYR:CE1	2.56	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	753/760 (99%)	735 (98%)	17 (2%)	1 (0%)	51	29
1	B	735/760 (97%)	717 (98%)	17 (2%)	1 (0%)	51	29
All	All	1488/1520 (98%)	1452 (98%)	34 (2%)	2 (0%)	51	29

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	318	VAL
1	B	318	VAL

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	628/631 (100%)	618 (98%)	10 (2%)	62	41
1	B	615/631 (98%)	609 (99%)	6 (1%)	76	61
All	All	1243/1262 (98%)	1227 (99%)	16 (1%)	73	50

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	93	LYS
1	A	94[A]	GLN
1	A	94[B]	GLN
1	A	157	MET
1	A	197	GLN
1	A	531[A]	ASN
1	A	531[B]	ASN
1	A	636	ASN
1	A	674	GLU
1	A	764	LYS
1	B	28	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	338	GLU
1	B	415	LYS
1	B	531	ASN
1	B	636	ASN
1	B	675	LEU

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	490	ASN

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 ⓘ

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	$\# Z > 2$
3	PEG	A	802	-	6,6,6	0.45	0	5,5,5	0.25	0
4	PG4	A	803	-	12,12,12	0.63	0	11,11,11	0.97	1 (9%)

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
3	PEG	A	802	-	-	3/4/4/4	-
4	PG4	A	803	-	-	4/10/10/10	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	803	PG4	O4-C6-C5	2.50	121.66	110.39

There are no chirality outliers.

All (7) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	803	PG4	O3-C5-C6-O4
3	A	802	PEG	O1-C1-C2-O2
4	A	803	PG4	O4-C7-C8-O5
3	A	802	PEG	C4-C3-O2-C2
4	A	803	PG4	C6-C5-O3-C4
4	A	803	PG4	C3-C4-O3-C5
3	A	802	PEG	O2-C3-C4-O4

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	803	PG4	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 ⓘ

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	745/760 (98%)	0.58	30 (4%) 38 35	8, 14, 28, 48	0
1	B	731/760 (96%)	0.82	80 (10%) 5 4	9, 17, 34, 54	0
All	All	1476/1520 (97%)	0.70	110 (7%) 14 12	8, 16, 31, 54	0

All (110) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	404	THR	6.2
1	B	400	LEU	5.6
1	B	68	ALA	5.5
1	A	672	THR	4.7
1	B	401	ARG	4.7
1	B	28	ASP	4.4
1	B	69	ALA	4.4
1	B	67	ILE	4.4
1	B	100	ARG	4.4
1	B	335	VAL	4.3
1	B	367	LYS	4.3
1	B	31	ARG	4.2
1	B	672	THR	4.2
1	A	403	GLY	4.1
1	B	404	THR	4.0
1	B	435[A]	TRP	3.9
1	A	405	PRO	3.9
1	B	726	ILE	3.8
1	A	402	PRO	3.7
1	B	406	ALA	3.7
1	B	34	ASP	3.6
1	B	197	GLN	3.6
1	B	356	LEU	3.6
1	B	405	PRO	3.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	626	GLN	3.5
1	A	663	ASP	3.5
1	B	106	LEU	3.3
1	A	668	THR	3.2
1	A	189	LEU	3.2
1	B	96	VAL	3.2
1	B	75	GLU	3.2
1	B	93	LYS	3.1
1	A	331	GLN	3.1
1	B	38	LYS	3.0
1	A	687	TRP	3.0
1	B	347	ILE	3.0
1	B	32	PHE	3.0
1	B	666	ARG	2.9
1	B	95	ALA	2.9
1	B	771	LYS	2.9
1	B	668	THR	2.9
1	B	664	LEU	2.9
1	A	665	SER	2.8
1	B	98	GLN	2.8
1	B	330	ILE	2.8
1	A	401	ARG	2.8
1	B	354	LEU	2.8
1	B	30	ASP	2.7
1	B	103	ILE	2.7
1	A	197	GLN	2.7
1	B	331	GLN	2.7
1	B	366	LEU	2.7
1	A	675	LEU	2.6
1	B	398	VAL	2.6
1	B	94	GLN	2.6
1	B	35	ALA	2.6
1	B	371	ARG	2.5
1	B	195	ASP	2.5
1	B	413	ASN	2.5
1	B	680	MET	2.5
1	B	687	TRP	2.5
1	B	152	TRP	2.5
1	B	348	LEU	2.5
1	B	673	GLY	2.5
1	B	402	PRO	2.5
1	B	41	THR	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	194	LYS	2.5
1	B	323	VAL	2.4
1	B	670	ASP	2.4
1	A	625	ALA	2.4
1	B	33	ILE	2.4
1	A	667	SER	2.4
1	B	86	GLU	2.4
1	B	358	ASP	2.3
1	B	72	LYS	2.3
1	B	88	ILE	2.3
1	A	329	SER	2.3
1	A	726	ILE	2.3
1	A	676	THR	2.3
1	B	724	SER	2.3
1	B	749	PRO	2.2
1	A	358	ASP	2.2
1	B	89	ARG	2.2
1	A	188	VAL	2.2
1	B	284	VAL	2.2
1	B	723	GLN	2.2
1	B	40	MET	2.2
1	A	666	ARG	2.2
1	A	737	ARG	2.2
1	B	283	VAL	2.2
1	B	99	SER	2.2
1	B	763	VAL	2.1
1	B	663	ASP	2.1
1	B	352	TYR	2.1
1	A	154	PHE	2.1
1	A	674	GLU	2.1
1	B	327	LYS	2.1
1	A	554	LEU	2.1
1	B	189	LEU	2.1
1	A	72	LYS	2.1
1	B	399	THR	2.1
1	A	32	PHE	2.1
1	B	282	PHE	2.1
1	B	53	VAL	2.1
1	B	586	ILE	2.1
1	B	721	PRO	2.0
1	B	671	MET	2.0
1	B	71	ILE	2.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	73	ARG	2.0
1	A	664	LEU	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
4	PG4	A	803	13/13	0.79	0.14	26,28,30,32	0
3	PEG	A	802	7/7	0.83	0.18	39,40,43,44	0
2	MG	A	801	1/1	0.91	0.07	20,20,20,20	0
2	MG	B	801	1/1	0.97	0.09	15,15,15,15	0

6.5 Other polymers [i](#)

There are no such residues in this entry.