



# Full wwPDB X-ray Structure Validation Report ⓘ

May 26, 2020 – 01:11 pm BST

PDB ID : 2ZXQ  
Title : Crystal structure of endo-alpha-N-acetylgalactosaminidase from Bifidobacterium longum (EngBF)  
Authors : Suzuki, R.; Katayama, T.; Ashida, H.; Yamamoto, K.; Kitaoka, M.; Fushinobu, S.  
Deposited on : 2009-01-05  
Resolution : 2.00 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 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

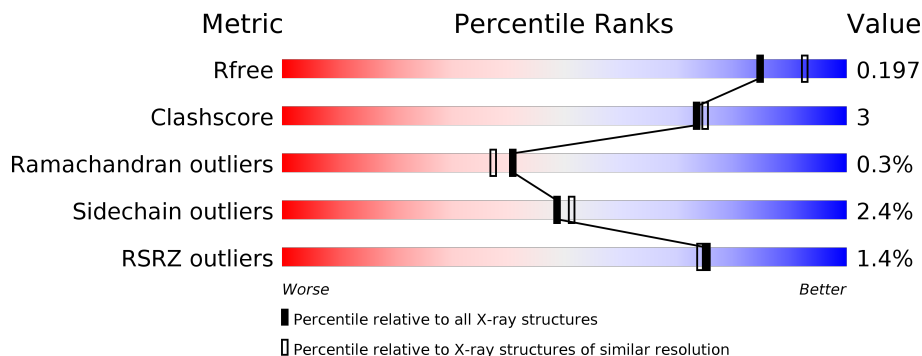
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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  $\geq 3$ , 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	1376	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	MPD	A	1695	X	-	-	-
3	MPD	A	1697	X	-	X	-

## 2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 10501 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 Endo-alpha-N-acetylgalactosaminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1178	9070	5643	1553	1850	24	0	0	0

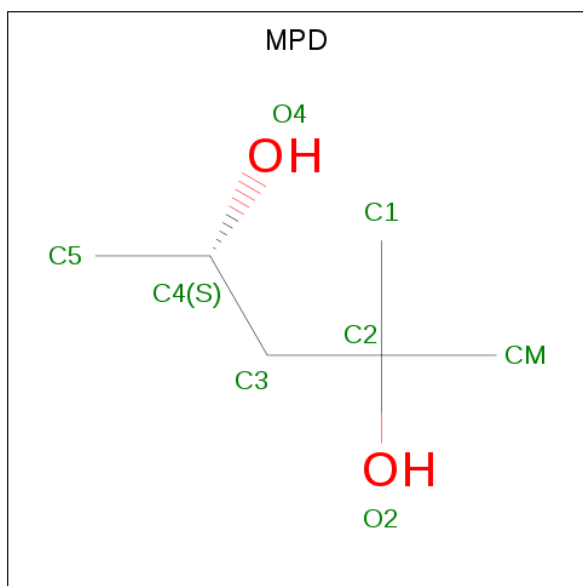
There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	319	MET	-	EXPRESSION TAG	UNP Q3T552
A	320	GLY	-	EXPRESSION TAG	UNP Q3T552
A	321	SER	-	EXPRESSION TAG	UNP Q3T552
A	322	SER	-	EXPRESSION TAG	UNP Q3T552
A	323	HIS	-	EXPRESSION TAG	UNP Q3T552
A	324	HIS	-	EXPRESSION TAG	UNP Q3T552
A	325	HIS	-	EXPRESSION TAG	UNP Q3T552
A	326	HIS	-	EXPRESSION TAG	UNP Q3T552
A	327	HIS	-	EXPRESSION TAG	UNP Q3T552
A	328	HIS	-	EXPRESSION TAG	UNP Q3T552
A	329	SER	-	EXPRESSION TAG	UNP Q3T552
A	330	SER	-	EXPRESSION TAG	UNP Q3T552
A	331	GLY	-	EXPRESSION TAG	UNP Q3T552
A	332	LEU	-	EXPRESSION TAG	UNP Q3T552
A	333	VAL	-	EXPRESSION TAG	UNP Q3T552
A	334	PRO	-	EXPRESSION TAG	UNP Q3T552
A	335	ARG	-	EXPRESSION TAG	UNP Q3T552
A	336	GLY	-	EXPRESSION TAG	UNP Q3T552
A	337	SER	-	EXPRESSION TAG	UNP Q3T552
A	338	HIS	-	EXPRESSION TAG	UNP Q3T552
A	339	MET	-	EXPRESSION TAG	UNP Q3T552
A	1123	SER	GLY	SEE REMARK 999	UNP Q3T552

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	4	Total Mn 4 4	0	0

- Molecule 3 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula:  $C_6H_{14}O_2$ ).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 8 6 2	0	0
3	A	1	Total C O 8 6 2	0	0
3	A	1	Total C O 8 6 2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1403	Total O 1403 1403	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	192.27Å 192.27Å 123.02Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	36.49 – 2.00 36.48 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.7 (36.49-2.00) 98.7 (36.48-2.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.16 (at 2.00Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.172 , 0.196 0.173 , 0.197	Depositor DCC
$R_{free}$ test set	8598 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	26.0	Xtrriage
Anisotropy	0.020	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 53.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	0.014 for h,-h-k,-l	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	10501	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.69% 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: MPD, MN

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	1/9250 (0.0%)	0.69	6/12541 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	390	GLU	CB-CG	5.85	1.63	1.52

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	694	ARG	NE-CZ-NH2	-14.71	112.95	120.30
1	A	694	ARG	NE-CZ-NH1	11.77	126.18	120.30
1	A	694	ARG	CG-CD-NE	-6.84	97.44	111.80
1	A	1211	ASN	C-N-CA	-6.28	109.11	122.30
1	A	1253	VAL	CB-CA-C	-6.04	99.92	111.40
1	A	694	ARG	CD-NE-CZ	5.52	131.33	123.60

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	9070	0	8655	53	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	4	0	0	0	0
3	A	24	0	42	17	0
4	A	1403	0	0	12	0
All	All	10501	0	8697	58	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.

All (58) 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:878:ASN:HD22	3:A:1697:MPD:HM2	0.98	1.08
1:A:878:ASN:ND2	3:A:1697:MPD:HM2	1.81	0.95
1:A:878:ASN:HD22	3:A:1697:MPD:CM	1.78	0.95
1:A:971:THR:HG21	3:A:1697:MPD:HM1	1.49	0.94
3:A:1697:MPD:H11	4:A:2431:HOH:O	1.69	0.92
1:A:580:LYS:HE2	4:A:1942:HOH:O	1.74	0.86
1:A:971:THR:CG2	3:A:1697:MPD:HM1	2.13	0.78
1:A:874:GLY:HA2	3:A:1697:MPD:C1	2.17	0.75
1:A:874:GLY:HA2	3:A:1697:MPD:H13	1.70	0.73
1:A:1274:GLY:H	3:A:1695:MPD:HM1	1.54	0.71
3:A:1695:MPD:H11	4:A:1869:HOH:O	1.93	0.68
1:A:1277:ASN:OD1	3:A:1695:MPD:H12	1.95	0.66
1:A:694:ARG:NH2	4:A:2221:HOH:O	2.07	0.66
1:A:580:LYS:HE3	4:A:2136:HOH:O	2.00	0.61
1:A:1171:GLY:HA3	1:A:1239:SER:HB3	1.83	0.61
1:A:1238:LYS:HG3	1:A:1239:SER:N	2.15	0.60
1:A:1208:ASN:OD1	1:A:1211:ASN:O	2.20	0.59
1:A:344:THR:OG1	1:A:353:GLN:NE2	2.36	0.58
3:A:1697:MPD:C1	4:A:2431:HOH:O	2.36	0.58
1:A:1473:LYS:O	1:A:1474:ARG:HB2	2.03	0.58
3:A:1697:MPD:H11	3:A:1697:MPD:H53	1.84	0.58
1:A:370:LYS:HE2	4:A:2206:HOH:O	2.06	0.55
1:A:874:GLY:HA2	3:A:1697:MPD:H12	1.89	0.55
1:A:1189:THR:HB	1:A:1345:GLN:HB3	1.90	0.53
1:A:1274:GLY:H	3:A:1695:MPD:CM	2.21	0.53
1:A:395:ASP:HA	1:A:411:LYS:HD2	1.89	0.53
1:A:570:TRP:CD2	1:A:592:PRO:HB3	2.43	0.53
1:A:1149:ARG:HH11	1:A:1149:ARG:HG3	1.74	0.53
1:A:1154:LYS:HB2	1:A:1157:GLN:HG3	1.91	0.53
1:A:1315:LYS:HG3	4:A:2276:HOH:O	2.08	0.53

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:557:THR:O	1:A:557:THR:CG2	2.57	0.52
1:A:1352:GLY:HA2	1:A:1512:ILE:HG13	1.91	0.51
1:A:1473:LYS:O	1:A:1474:ARG:CB	2.59	0.51
1:A:468:GLN:HG2	1:A:566:ALA:CB	2.42	0.50
1:A:1467:ARG:O	1:A:1471:ASP:HB2	2.14	0.48
1:A:634:VAL:HA	1:A:914:LYS:HG3	1.96	0.48
1:A:1145:ASP:OD1	1:A:1236:SER:HB3	2.13	0.47
1:A:870:TYR:CE1	1:A:997:ASP:HA	2.50	0.47
1:A:1462:LYS:O	1:A:1466:ILE:HG12	2.14	0.47
1:A:1239:SER:OG	1:A:1241:ASP:HB2	2.15	0.47
1:A:507:VAL:HG22	1:A:510:SER:HB2	1.96	0.47
1:A:861:HIS:CD2	1:A:861:HIS:H	2.33	0.46
3:A:1695:MPD:HM2	4:A:2257:HOH:O	2.15	0.46
1:A:861:HIS:HE1	4:A:73:HOH:O	1.99	0.46
1:A:1211:ASN:HD22	1:A:1211:ASN:HA	1.62	0.45
1:A:1135:ASN:HD22	1:A:1135:ASN:C	2.20	0.45
1:A:1357:VAL:O	1:A:1399:GLU:HA	2.17	0.44
1:A:835:HIS:HE1	4:A:75:HOH:O	1.99	0.44
1:A:599:ALA:HB1	1:A:602:GLU:HG2	2.00	0.44
1:A:1154:LYS:CB	1:A:1157:GLN:HG3	2.48	0.43
1:A:1302:GLU:HB2	1:A:1325:TRP:CZ3	2.54	0.42
1:A:468:GLN:NE2	1:A:567:SER:H	2.17	0.42
1:A:1149:ARG:NH1	1:A:1149:ARG:HG3	2.34	0.42
1:A:1093:GLN:OE1	4:A:1832:HOH:O	2.21	0.42
1:A:1516:LEU:O	1:A:1520:MET:HB2	2.21	0.41
1:A:878:ASN:HB2	3:A:1697:MPD:H13	2.03	0.40
1:A:358:PHE:CG	1:A:359:PRO:HD2	2.56	0.40
1:A:1139:ARG:HD2	1:A:1244:HIS:CG	2.56	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	1174/1376 (85%)	1130 (96%)	41 (4%)	3 (0%)	41 37

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	790	VAL
1	A	1474	ARG
1	A	1473	LYS

### 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	967/1124 (86%)	944 (98%)	23 (2%)	49 51

All (23) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	343	GLU
1	A	360	SER
1	A	390	GLU
1	A	447	GLU
1	A	507	VAL
1	A	557	THR
1	A	580	LYS
1	A	632	GLU
1	A	793	ASN
1	A	1025	GLN
1	A	1135	ASN
1	A	1184	LEU
1	A	1211	ASN
1	A	1240	THR
1	A	1253	VAL
1	A	1264	LYS
1	A	1280	GLN
1	A	1322	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	1382	SER
1	A	1402	LEU
1	A	1452	GLN
1	A	1473	LYS
1	A	1484	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (19) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	353	GLN
1	A	433	GLN
1	A	468	GLN
1	A	498	ASN
1	A	506	ASN
1	A	793	ASN
1	A	834	GLN
1	A	835	HIS
1	A	861	HIS
1	A	872	GLN
1	A	878	ASN
1	A	894	GLN
1	A	936	ASN
1	A	951	GLN
1	A	978	GLN
1	A	1047	ASN
1	A	1135	ASN
1	A	1211	ASN
1	A	1375	GLN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry

Of 7 ligands modelled in this entry, 4 are monoatomic - leaving 3 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	MPD	A	1695	-	7,7,7	0.93	0	9,10,10	1.88	3 (33%)
3	MPD	A	1697	-	7,7,7	0.74	0	9,10,10	1.32	1 (11%)
3	MPD	A	1696	-	7,7,7	0.31	0	9,10,10	0.67	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
3	MPD	A	1695	-	1/1/2/2	3/5/5/5	-
3	MPD	A	1697	-	1/1/2/2	0/5/5/5	-
3	MPD	A	1696	-	-	2/5/5/5	-

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1695	MPD	C1-C2-C3	3.44	125.99	109.96
3	A	1697	MPD	O2-C2-CM	-3.27	97.57	108.08
3	A	1695	MPD	CM-C2-C1	-2.99	104.35	110.57
3	A	1695	MPD	O2-C2-C3	-2.52	100.34	109.80

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	1695	MPD	C4
3	A	1697	MPD	C4

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1695	MPD	C1-C2-C3-C4
3	A	1695	MPD	O2-C2-C3-C4
3	A	1696	MPD	O2-C2-C3-C4
3	A	1695	MPD	CM-C2-C3-C4
3	A	1696	MPD	CM-C2-C3-C4

There are no ring outliers.

2 monomers are involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1695	MPD	5	0
3	A	1697	MPD	12	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	1178/1376 (85%)	-0.55	17 (1%) 75 74	17, 27, 47, 78	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1472	SER	5.5
1	A	1523	ALA	3.6
1	A	1475	ALA	3.3
1	A	1240	THR	3.3
1	A	1092	LYS	3.2
1	A	1474	ARG	2.9
1	A	1468	GLY	2.8
1	A	559	ASP	2.7
1	A	1482	TRP	2.6
1	A	1022	ASP	2.6
1	A	1241	ASP	2.6
1	A	1238	LYS	2.5
1	A	1524	PRO	2.4
1	A	1485	TYR	2.4
1	A	1095	GLN	2.3
1	A	1212	GLY	2.2
1	A	1473	LYS	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, 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	MPD	A	1697	8/8	0.79	0.34	49,52,60,60	0
3	MPD	A	1696	8/8	0.86	0.21	68,69,70,70	0
3	MPD	A	1695	8/8	0.87	0.21	23,29,48,50	0
2	MN	A	2	1/1	0.99	0.04	42,42,42,42	0
2	MN	A	4	1/1	0.99	0.03	37,37,37,37	0
2	MN	A	3	1/1	1.00	0.04	33,33,33,33	0
2	MN	A	1	1/1	1.00	0.02	30,30,30,30	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.