



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 9, 2020 – 04:22 AM BST

PDB ID : 4GVF  
Title : Crystal structure of Salmonella typhimurium family 3 glycoside hydrolase (NagZ) bound to GlcNAc  
Authors : Bacik, J.P.; Mark, B.L.  
Deposited on : 2012-08-30  
Resolution : 1.35 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 : 2.13.1  
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

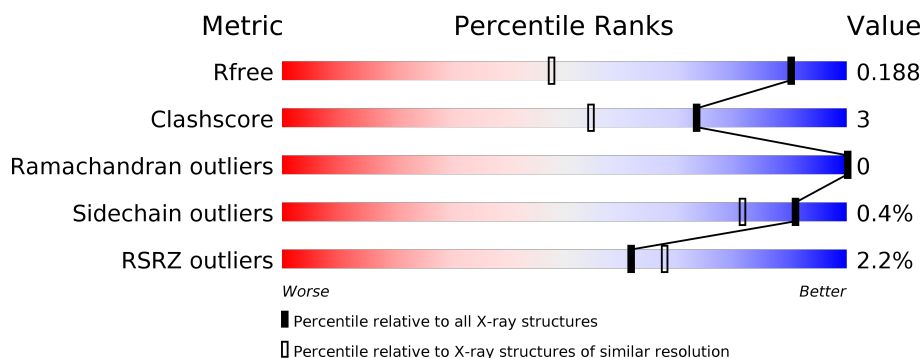
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.35 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	1509 (1.38-1.34)
Clashscore	141614	1551 (1.38-1.34)
Ramachandran outliers	138981	1530 (1.38-1.34)
Sidechain outliers	138945	1530 (1.38-1.34)
RSRZ outliers	127900	1487 (1.38-1.34)

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	349	<div> <div>3%</div> <div> <div></div> <div>91%</div> <div>7%</div> <div>.</div> </div> </div>
1	B	349	<div> <div>%</div> <div> <div></div> <div>88%</div> <div>6%</div> <div>6%</div> </div> </div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 6264 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.

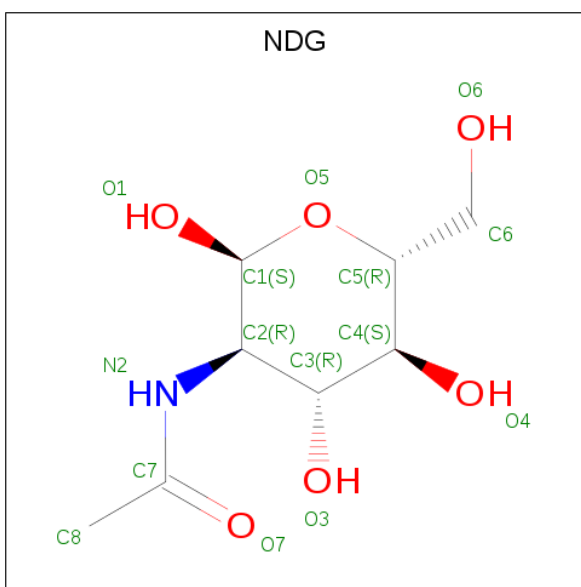
- Molecule 1 is a protein called Beta-hexosaminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	340	Total	C	N	O	S	0	8	0
			2644	1650	477	497	20			
1	B	328	Total	C	N	O	S	0	7	0
			2560	1601	463	479	17			

There are 16 discrepancies between the modelled and reference sequences:

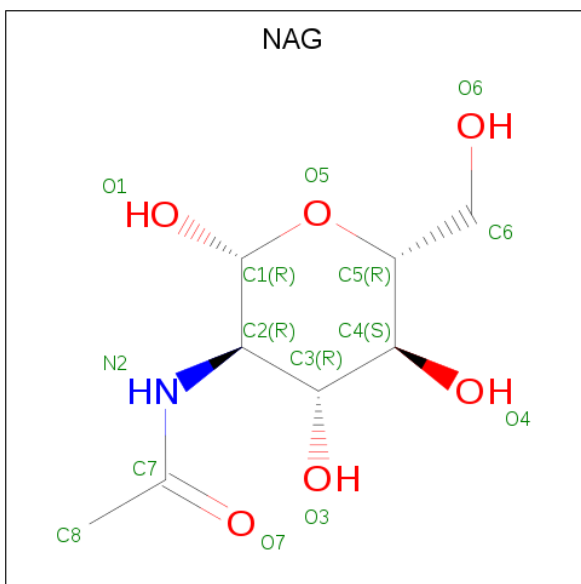
Chain	Residue	Modelled	Actual	Comment	Reference
A	342	GLY	-	expression tag	UNP Q8ZQ06
A	343	SER	-	expression tag	UNP Q8ZQ06
A	344	HIS	-	expression tag	UNP Q8ZQ06
A	345	HIS	-	expression tag	UNP Q8ZQ06
A	346	HIS	-	expression tag	UNP Q8ZQ06
A	347	HIS	-	expression tag	UNP Q8ZQ06
A	348	HIS	-	expression tag	UNP Q8ZQ06
A	349	HIS	-	expression tag	UNP Q8ZQ06
B	342	GLY	-	expression tag	UNP Q8ZQ06
B	343	SER	-	expression tag	UNP Q8ZQ06
B	344	HIS	-	expression tag	UNP Q8ZQ06
B	345	HIS	-	expression tag	UNP Q8ZQ06
B	346	HIS	-	expression tag	UNP Q8ZQ06
B	347	HIS	-	expression tag	UNP Q8ZQ06
B	348	HIS	-	expression tag	UNP Q8ZQ06
B	349	HIS	-	expression tag	UNP Q8ZQ06

- Molecule 2 is 2-acetamido-2-deoxy-alpha-D-glucopyranose (three-letter code: NDG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	1
			15	8	1	6		
2	B	1	Total	C	N	O	0	1
			15	8	1	6		

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



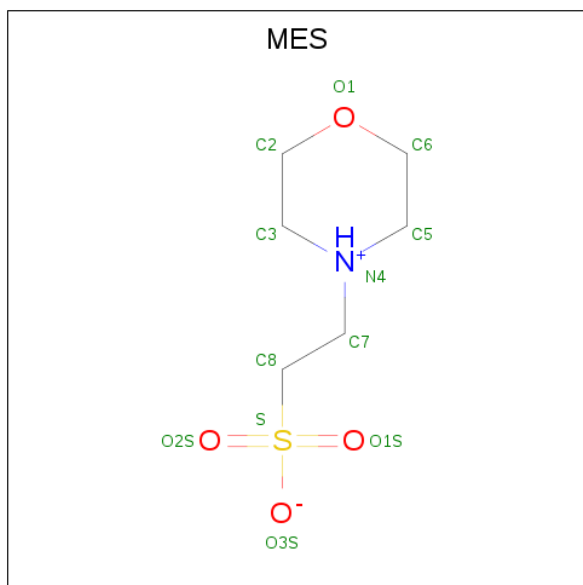
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	1
			15	8	1	6		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	B	1	Total	C	N	O	0	1
			15	8	1	6		

- Molecule 4 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: C<sub>6</sub>H<sub>13</sub>NO<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	B	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

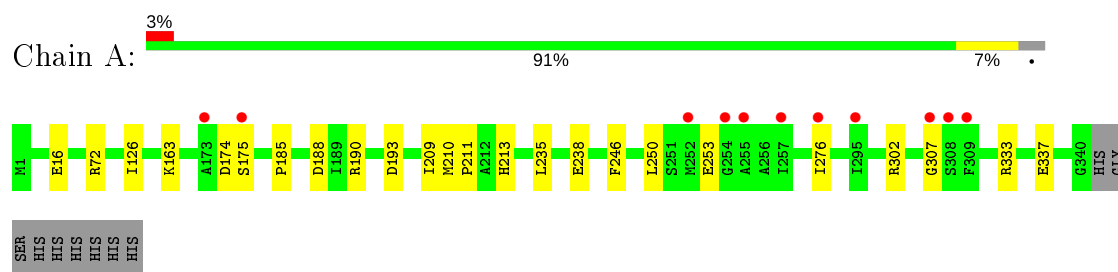
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	509	Total	O	0	0
			509	509		
5	B	479	Total	O	0	0
			479	479		

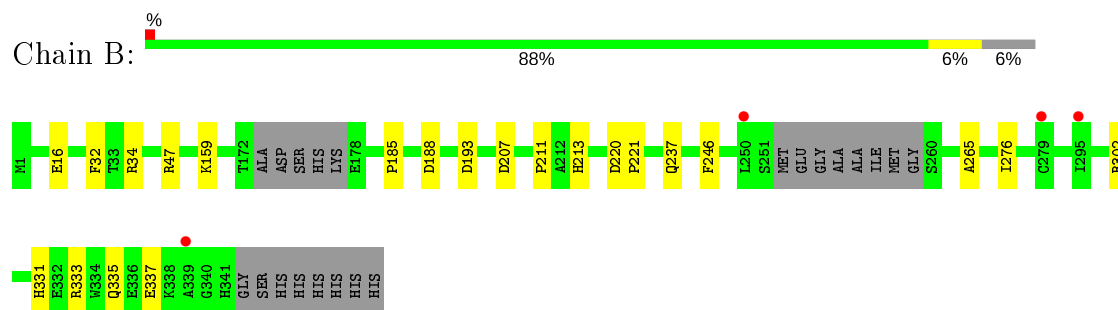
### 3 Residue-property plots [i](#)

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-hexosaminidase



- Molecule 1: Beta-hexosaminidase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	49.54Å 66.23Å 94.97Å 90.00° 99.32° 90.00°	Depositor
Resolution (Å)	48.89 – 1.35 48.89 – 1.35	Depositor EDS
% Data completeness (in resolution range)	96.9 (48.89-1.35) 96.8 (48.89-1.35)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.27 (at 1.35Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.4_486)	Depositor
R, $R_{free}$	0.167 , 0.190 0.164 , 0.188	Depositor DCC
$R_{free}$ test set	3951 reflections (3.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	7.9	Xtriage
Anisotropy	0.814	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 40.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	6264	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	13.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 52.51 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 4.8120e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: NAG, NDG, MES

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	0.30	0/2718	0.51	0/3674
1	B	0.30	0/2628	0.51	0/3550
All	All	0.30	0/5346	0.51	0/7224

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	2644	0	2578	18	0
1	B	2560	0	2513	16	0
2	A	15	0	12	0	0
2	B	15	0	12	0	0
3	A	15	0	15	0	0
3	B	15	0	15	0	0
4	B	12	0	12	0	0
5	A	509	0	0	6	0
5	B	479	0	0	7	0
All	All	6264	0	5157	34	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 (34) 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:302:ARG:HD2	5:A:1004:HOH:O	1.64	0.94
1:B:302:ARG:HD2	5:B:974:HOH:O	1.73	0.88
1:B:213:HIS:HE1	5:B:975:HOH:O	1.73	0.72
1:A:250:LEU:HD11	1:A:276:ILE:HG21	1.75	0.67
1:B:331:HIS:O	1:B:335:GLN:HG2	1.98	0.63
1:B:265:ALA:HB1	1:B:276:ILE:HD12	1.81	0.61
1:A:213:HIS:HE1	1:A:253:GLU:OE1	1.86	0.58
1:A:190:ARG:NH2	1:A:238:GLU:OE1	2.37	0.58
1:B:159:LYS:HD3	5:B:803:HOH:O	2.06	0.55
1:A:209:ILE:HD12	1:A:235:LEU:HD11	1.90	0.53
1:A:213:HIS:HD2	5:A:724:HOH:O	1.93	0.52
1:A:307:GLY:HA3	5:A:718:HOH:O	2.10	0.50
1:B:265:ALA:HB1	1:B:276:ILE:CD1	2.40	0.49
1:B:32:PHE:CD1	1:B:34[B]:ARG:NH1	2.80	0.49
1:B:213:HIS:HD2	5:B:977:HOH:O	1.95	0.49
1:A:16:GLU:HG2	5:A:944:HOH:O	2.13	0.49
1:A:163:LYS:HB3	1:A:210[B]:MET:HB3	1.95	0.48
1:A:185:PRO:HG2	1:A:188:ASP:OD2	2.14	0.48
1:A:72[B]:ARG:NH1	5:A:1003:HOH:O	2.47	0.48
1:B:47:ARG:HD2	5:B:893:HOH:O	2.15	0.47
1:A:126:ILE:HG12	5:A:640:HOH:O	2.13	0.47
1:A:190:ARG:HH22	1:A:238:GLU:CD	2.18	0.46
1:A:250:LEU:HD11	1:A:276:ILE:CG2	2.44	0.46
1:A:333:ARG:NH1	1:A:337:GLU:HB2	2.32	0.45
1:B:333:ARG:O	1:B:337:GLU:HG2	2.16	0.45
1:A:174:ASP:O	1:A:175:SER:HB3	2.18	0.44
1:B:16:GLU:HG2	5:B:778:HOH:O	2.18	0.44
1:B:211:PRO:HD2	1:B:246:PHE:O	2.17	0.44
1:B:159:LYS:HD2	1:B:207:ASP:CG	2.39	0.43
1:A:211:PRO:HD2	1:A:246:PHE:O	2.19	0.43
1:B:185:PRO:HG2	1:B:188:ASP:OD2	2.19	0.43
1:A:213:HIS:CE1	1:A:253:GLU:OE1	2.71	0.41
1:B:237:GLN:HG3	5:B:919:HOH:O	2.20	0.41
1:B:220:ASP:HA	1:B:221:PRO:HD3	1.93	0.41

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	346/349 (99%)	333 (96%)	13 (4%)	0	100	100
1	B	329/349 (94%)	317 (96%)	12 (4%)	0	100	100
All	All	675/698 (97%)	650 (96%)	25 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	272/281 (97%)	271 (100%)	1 (0%)	91	81
1	B	267/281 (95%)	266 (100%)	1 (0%)	91	81
All	All	539/562 (96%)	537 (100%)	2 (0%)	91	81

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	193	ASP
1	B	193	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	213	HIS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	213	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](#)

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](#)

5 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	NDG	A	401[A]	-	15,15,15	1.66	4 (26%)	21,21,21	1.18	2 (9%)
3	NAG	B	402[B]	-	15,15,15	1.57	3 (20%)	21,21,21	1.16	2 (9%)
2	NDG	B	401[A]	-	15,15,15	1.57	3 (20%)	21,21,21	1.14	2 (9%)
4	MES	B	403	-	12,12,12	2.06	1 (8%)	14,16,16	2.29	3 (21%)
3	NAG	A	402[B]	-	15,15,15	1.65	4 (26%)	21,21,21	1.22	2 (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
2	NDG	A	401[A]	-	-	0/6/26/26	0/1/1/1
3	NAG	B	402[B]	-	-	0/6/26/26	0/1/1/1
2	NDG	B	401[A]	-	-	0/6/26/26	0/1/1/1
4	MES	B	403	-	-	5/6/14/14	0/1/1/1
3	NAG	A	402[B]	-	-	0/6/26/26	0/1/1/1

All (15) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	403	MES	C8-S	-6.88	1.67	1.77
2	A	401[A]	NDG	C4-C3	-3.12	1.44	1.52
3	A	402[B]	NAG	C4-C3	-3.12	1.44	1.52
3	B	402[B]	NAG	C4-C3	-2.94	1.44	1.52
2	B	401[A]	NDG	C4-C3	-2.94	1.44	1.52
2	A	401[A]	NDG	C1-C2	-2.90	1.49	1.52
3	A	402[B]	NAG	C1-C2	-2.90	1.49	1.52
3	B	402[B]	NAG	C3-C2	-2.73	1.48	1.53
2	B	401[A]	NDG	C3-C2	-2.73	1.48	1.53
2	A	401[A]	NDG	C3-C2	-2.59	1.48	1.53
3	A	402[B]	NAG	C3-C2	-2.59	1.48	1.53
3	B	402[B]	NAG	C1-C2	-2.45	1.50	1.52
2	B	401[A]	NDG	C1-C2	-2.45	1.50	1.52
2	A	401[A]	NDG	C7-N2	2.08	1.41	1.34
3	A	402[B]	NAG	C7-N2	2.08	1.41	1.34

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	403	MES	O1S-S-C8	6.47	114.70	106.92
3	B	402[B]	NAG	C1-C2-N2	-3.27	106.94	110.73
2	B	401[A]	NDG	C1-C2-N2	-3.27	106.94	110.73
4	B	403	MES	C5-N4-C3	3.06	115.72	108.83
2	A	401[A]	NDG	C1-C2-N2	-3.01	107.24	110.73
3	A	402[B]	NAG	C1-C2-N2	-3.01	107.24	110.73
2	A	401[A]	NDG	O5-C1-C2	2.82	112.35	109.52
3	A	402[B]	NAG	O5-C1-C2	2.82	112.35	109.52
4	B	403	MES	C7-N4-C3	2.70	118.13	111.23
3	B	402[B]	NAG	O5-C1-C2	2.60	112.13	109.52
2	B	401[A]	NDG	O5-C1-C2	2.60	112.13	109.52

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	403	MES	C8-C7-N4-C3
4	B	403	MES	C7-C8-S-O3S
4	B	403	MES	C8-C7-N4-C5
4	B	403	MES	C7-C8-S-O1S
4	B	403	MES	C7-C8-S-O2S

There are no ring outliers.

No monomer is involved in short contacts.

## 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	340/349 (97%)	0.02	11 (3%) 47 53	4, 9, 23, 34	0
1	B	328/349 (93%)	0.01	4 (1%) 79 83	4, 9, 24, 34	0
All	All	668/698 (95%)	0.01	15 (2%) 62 68	4, 9, 24, 34	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	309	PHE	3.9
1	A	255	ALA	3.9
1	A	175	SER	3.7
1	A	173	ALA	3.4
1	B	339	ALA	3.3
1	A	254	GLY	3.3
1	A	308	SER	3.2
1	A	276	ILE	3.2
1	A	252	MET	3.1
1	A	257	ILE	3.1
1	A	307	GLY	2.9
1	A	295	ILE	2.6
1	B	250	LEU	2.4
1	B	279	CYS	2.3
1	B	295	ILE	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

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(Å <sup>2</sup> )	Q<0.9
4	MES	B	403	12/12	0.96	0.12	11,19,23,26	3
3	NAG	B	402[B]	15/15	0.98	0.07	4,6,7,14	15
2	NDG	B	401[A]	15/15	0.98	0.07	4,6,7,8	15
2	NDG	A	401[A]	15/15	0.98	0.06	5,6,9,9	15
3	NAG	A	402[B]	15/15	0.98	0.06	5,6,9,14	15

## 6.5 Other polymers

There are no such residues in this entry.