



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 13, 2024 – 07:16 pm GMT

PDB ID : 6YP1
Title : HiCel7B unliganded
Authors : McGregor, N.G.S.; Davies, G.J.
Deposited on : 2020-04-15
Resolution : 1.20 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 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 : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.36
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) : 2.36

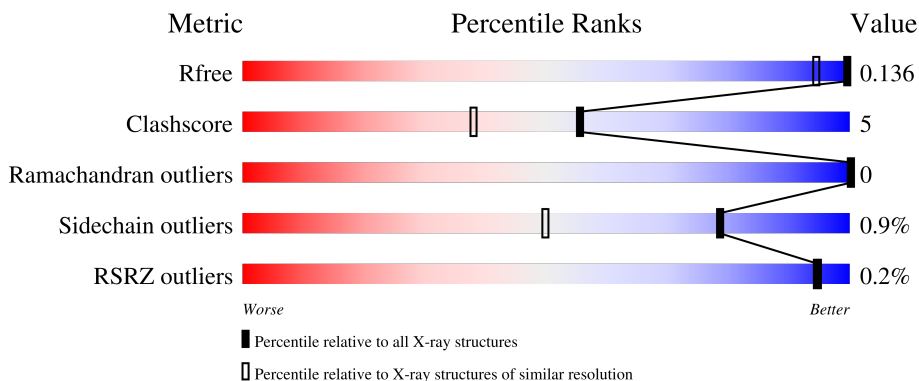
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION


The reported resolution of this entry is 1.20 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	1223 (1.22-1.18)
Clashscore	141614	1286 (1.22-1.18)
Ramachandran outliers	138981	1240 (1.22-1.18)
Sidechain outliers	138945	1239 (1.22-1.18)
RSRZ outliers	127900	1200 (1.22-1.18)

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 ≥ 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	AAA	402	 91% 8%

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
2	GOL	AAA	714	-	-	X	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	ACM	AAA	706	-	X	-	-
5	NAG	AAA	717	-	X	-	-

2 Entry composition [i](#)

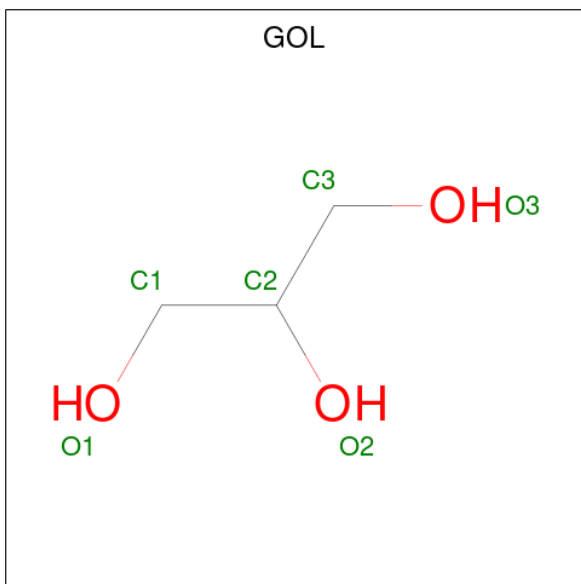
There are 6 unique types of molecules in this entry. The entry contains 6708 atoms, of which 3057 are hydrogens and 0 are deuteriums.

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 Endoglucanase 1.

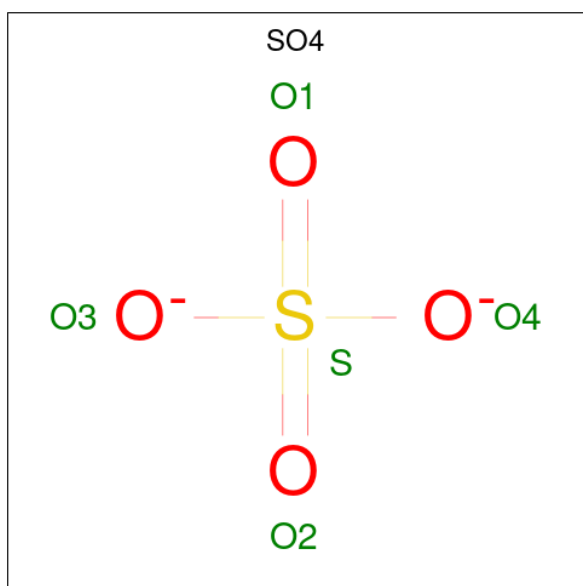
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	AAA	402	6116	1978	2976	528	602	32	160	12	0

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



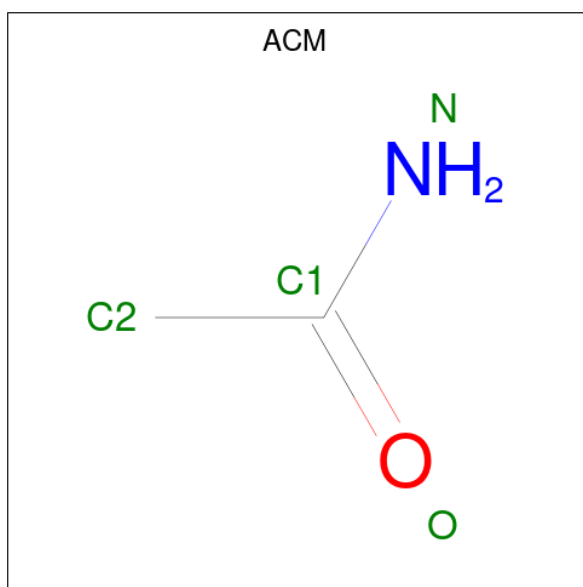
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
2	AAA	1	14	3	8	3	2	0
2	AAA	1	14	3	8	3	2	0
2	AAA	1	28	6	16	6	4	1
2	AAA	1	14	3	8	3	2	0
2	AAA	1	14	3	8	3	2	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



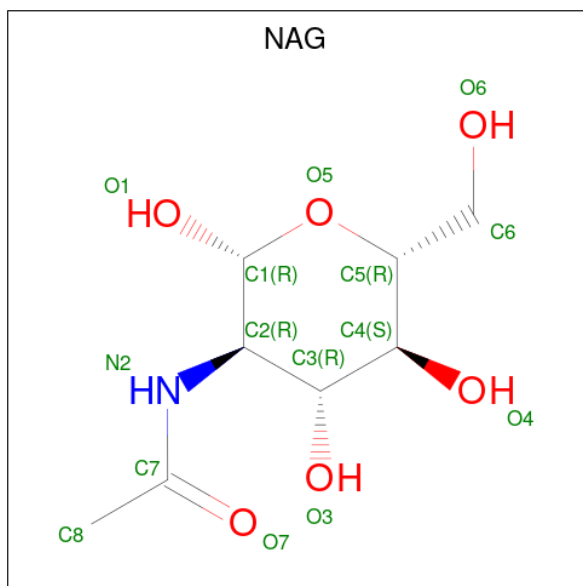
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	AAA	1	Total	O	S	0	1
			10	8	2		
3	AAA	1	Total	O	S	0	0
			5	4	1		
3	AAA	1	Total	O	S	0	0
			5	4	1		
3	AAA	1	Total	O	S	0	0
			5	4	1		
3	AAA	1	Total	O	S	0	0
			5	4	1		
3	AAA	1	Total	O	S	0	0
			5	4	1		
3	AAA	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is ACETAMIDE (three-letter code: ACM) (formula: C₂H₅NO).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	H	N	O		
4	AAA	1	9	2	5	1	1	2	0

- Molecule 5 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	H	N	O		
5	AAA	1	28	8	14	1	5	3	0
5	AAA	1	28	8	14	1	5	3	0

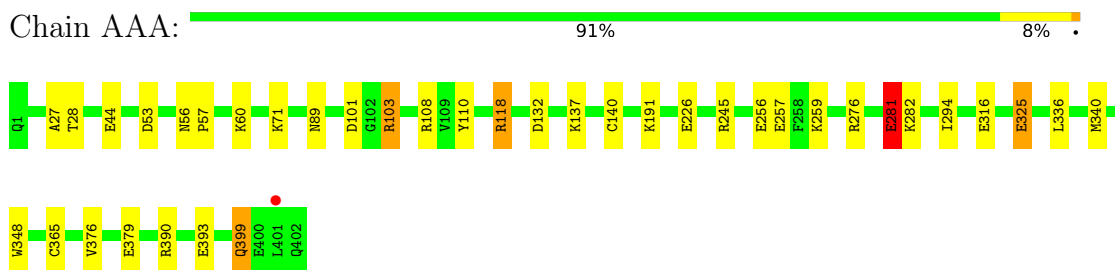
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	AAA	391	Total 393	O 393	0	2

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: Endoglucanase 1



4 Data and refinement statistics

Property	Value	Source
Space group	P 42 21 2	Depositor
Cell constants a, b, c, α , β , γ	92.31Å 92.31Å 100.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.20 – 1.20 46.16 – 1.20	Depositor EDS
% Data completeness (in resolution range)	99.7 (46.20-1.20) 99.7 (46.16-1.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.08 (at 1.20Å)	Xtrriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.117 , 0.135 0.118 , 0.136	Depositor DCC
R_{free} test set	6795 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	11.5	Xtrriage
Anisotropy	0.134	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.45 , 46.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.98	EDS
Total number of atoms	6708	wwPDB-VP
Average B, all atoms (Å ²)	14.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.68% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: NAG, GOL, ACM, SO4, PCA

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	AAA	1.04	14/3254 (0.4%)	1.11	14/4413 (0.3%)

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	AAA	0	3

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	AAA	393	GLU	CD-OE1	14.86	1.42	1.25
1	AAA	325	GLU	CD-OE2	-12.49	1.11	1.25
1	AAA	379	GLU	CD-OE2	10.63	1.37	1.25
1	AAA	256[A]	GLU	CD-OE1	9.83	1.36	1.25
1	AAA	256[B]	GLU	CD-OE1	9.83	1.36	1.25

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	AAA	103	ARG	NE-CZ-NH1	-25.11	107.75	120.30
1	AAA	103	ARG	NE-CZ-NH2	16.85	128.73	120.30
1	AAA	276	ARG	N-CA-CB	-11.07	90.67	110.60
1	AAA	118	ARG	NE-CZ-NH2	10.66	125.63	120.30
1	AAA	399	GLN	N-CA-CB	7.47	124.05	110.60

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	AAA	110	TYR	Sidechain
1	AAA	325	GLU	Sidechain
1	AAA	399	GLN	Mainchain

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	AAA	3140	2976	2955	23	2
2	AAA	36	48	48	8	0
3	AAA	50	0	0	2	0
4	AAA	4	5	5	0	0
5	AAA	28	28	26	3	1
6	AAA	393	0	0	11	1
All	All	3651	3057	3034	30	2

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 5.

The worst 5 of 30 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:AAA:140[B]:CYS:SG	1:AAA:365:CYS:HB3	2.04	0.97
1:AAA:28:THR:H	2:AAA:714:GOL:H2	1.32	0.93
1:AAA:140[B]:CYS:SG	1:AAA:365:CYS:CB	2.60	0.90
1:AAA:118:ARG:NH1	3:AAA:702[B]:SO4:O1	2.08	0.86
2:AAA:714:GOL:H32	6:AAA:961:HOH:O	1.79	0.83

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:AAA:257[A]:GLU:OE1	6:AAA:869:HOH:O[5_545]	2.00	0.20
1:AAA:132:ASP:OD2	5:AAA:717:NAG:H82[8_555]	1.60	0.00

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	AAA	412/402 (102%)	405 (98%)	7 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	AAA	334/334 (100%)	330 (99%)	4 (1%)	71	37

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	71	LYS
1	AAA	281[B]	GLU
1	AAA	281[C]	GLU
1	AAA	348	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](#)

1 non-standard protein/DNA/RNA residue is 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
1	PCA	AAA	1	1	7,8,9	0.77	0	9,10,12	0.75	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
1	PCA	AAA	1	1	-	0/0/11/13	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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
3	SO4	AAA	702[B]	-	4,4,4	0.51	0	6,6,6	0.39	0
4	ACM	AAA	706	-	3,3,3	1.23	1 (33%)	3,3,3	2.19	2 (66%)
5	NAG	AAA	717	1	14,14,15	3.60	9 (64%)	17,19,21	5.47	10 (58%)
3	SO4	AAA	713	-	4,4,4	0.57	0	6,6,6	0.42	0
3	SO4	AAA	708	-	4,4,4	0.68	0	6,6,6	0.78	0
3	SO4	AAA	703	-	4,4,4	0.52	0	6,6,6	0.79	0
3	SO4	AAA	710	-	4,4,4	0.94	0	6,6,6	0.31	0
3	SO4	AAA	704	-	4,4,4	0.40	0	6,6,6	0.27	0
2	GOL	AAA	705	-	5,5,5	0.62	0	5,5,5	0.34	0
2	GOL	AAA	714	-	5,5,5	0.33	0	5,5,5	1.44	0
2	GOL	AAA	711[A]	-	5,5,5	0.39	0	5,5,5	0.50	0
3	SO4	AAA	715	-	4,4,4	0.28	0	6,6,6	0.54	0
2	GOL	AAA	711[B]	-	5,5,5	0.20	0	5,5,5	0.88	0
2	GOL	AAA	712	-	5,5,5	0.33	0	5,5,5	0.55	0
3	SO4	AAA	716	-	4,4,4	0.38	0	6,6,6	0.26	0
5	NAG	AAA	707	1	14,14,15	1.49	2 (14%)	17,19,21	1.33	2 (11%)
2	GOL	AAA	701	-	5,5,5	1.59	2 (40%)	5,5,5	0.97	0
3	SO4	AAA	709	-	4,4,4	0.67	0	6,6,6	0.65	0
3	SO4	AAA	702[A]	-	4,4,4	0.33	0	6,6,6	0.65	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
5	NAG	AAA	717	1	-	4/6/23/26	0/1/1/1
2	GOL	AAA	705	-	-	0/4/4/4	-
2	GOL	AAA	712	-	-	2/4/4/4	-
2	GOL	AAA	714	-	-	0/4/4/4	-
5	NAG	AAA	707	1	-	0/6/23/26	0/1/1/1
2	GOL	AAA	701	-	-	1/4/4/4	-
2	GOL	AAA	711[A]	-	-	0/4/4/4	-
2	GOL	AAA	711[B]	-	-	0/4/4/4	-

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	AAA	717	NAG	O7-C7	7.88	1.41	1.23
5	AAA	717	NAG	O4-C4	5.16	1.55	1.43
5	AAA	717	NAG	C8-C7	-4.79	1.40	1.50
5	AAA	717	NAG	C7-N2	4.24	1.48	1.34
5	AAA	707	NAG	C1-C2	-3.59	1.47	1.52

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	AAA	717	NAG	O7-C7-N2	-11.94	100.01	121.95
5	AAA	717	NAG	C1-C2-N2	-10.24	92.99	110.49
5	AAA	717	NAG	O7-C7-C8	8.20	137.29	122.06
5	AAA	717	NAG	C2-N2-C7	7.89	134.14	122.90
5	AAA	717	NAG	C4-C3-C2	-5.94	102.31	111.02

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	AAA	712	GOL	C1-C2-C3-O3
5	AAA	717	NAG	C8-C7-N2-C2
5	AAA	717	NAG	O7-C7-N2-C2
2	AAA	701	GOL	C1-C2-C3-O3
2	AAA	712	GOL	O2-C2-C3-O3

There are no ring outliers.

8 monomers are involved in 14 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	AAA	702[B]	SO4	1	0
5	AAA	717	NAG	3	1
2	AAA	714	GOL	5	0
2	AAA	711[A]	GOL	1	0
2	AAA	711[B]	GOL	1	0
2	AAA	712	GOL	1	0
2	AAA	701	GOL	1	0
3	AAA	702[A]	SO4	1	0

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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, 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	AAA	401/402 (99%)	-0.52	1 (0%) 95 94	8, 11, 23, 51	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	401	LEU	3.5

6.2 Non-standard residues in protein, DNA, RNA chains [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(Å ²)	Q<0.9
1	PCA	AAA	1	8/9	0.97	0.07	14,15,16,17	0

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, 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(Å ²)	Q<0.9
5	NAG	AAA	717	14/15	0.79	0.25	20,30,44,64	3

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	GOL	AAA	712	6/6	0.85	0.15	29,38,47,48	2
3	SO4	AAA	716	5/5	0.86	0.13	35,36,47,53	5
3	SO4	AAA	708	5/5	0.88	0.14	19,28,41,46	5
3	SO4	AAA	702[B]	5/5	0.90	0.16	21,22,28,38	5
3	SO4	AAA	702[A]	5/5	0.90	0.16	26,30,40,44	5
3	SO4	AAA	713	5/5	0.91	0.17	25,26,36,39	5
2	GOL	AAA	705	6/6	0.92	0.11	11,15,18,18	14
2	GOL	AAA	711[A]	6/6	0.92	0.14	12,20,35,37	14
2	GOL	AAA	711[B]	6/6	0.92	0.14	14,18,34,34	14
3	SO4	AAA	703	5/5	0.92	0.14	16,21,32,38	5
3	SO4	AAA	709	5/5	0.93	0.10	26,34,54,82	5
2	GOL	AAA	714	6/6	0.94	0.16	16,28,60,63	2
2	GOL	AAA	701	6/6	0.95	0.07	14,19,26,27	2
3	SO4	AAA	715	5/5	0.95	0.31	33,42,49,52	5
4	ACM	AAA	706	4/4	0.96	0.08	13,14,25,26	9
3	SO4	AAA	704	5/5	0.96	0.17	31,39,59,65	0
5	NAG	AAA	707	14/15	0.97	0.07	12,14,21,22	3
3	SO4	AAA	710	5/5	0.97	0.13	22,25,34,35	5

6.5 Other polymers [i](#)

There are no such residues in this entry.