



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 16, 2018 – 04:27 pm GMT

PDB ID : 1KFG
Title : The X-ray Crystal Structure of Cel9G from Clostridium cellulolyticum complexed with a Thio-Oligosaccharide Inhibitor
Authors : Mandelman, D.; Belaich, A.; Belaich, J.-P.; Driguez, H.; Haser, R.
Deposited on : 2001-11-20
Resolution : 1.90 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : **NOT EXECUTED**
EDS : **NOT EXECUTED**
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30686

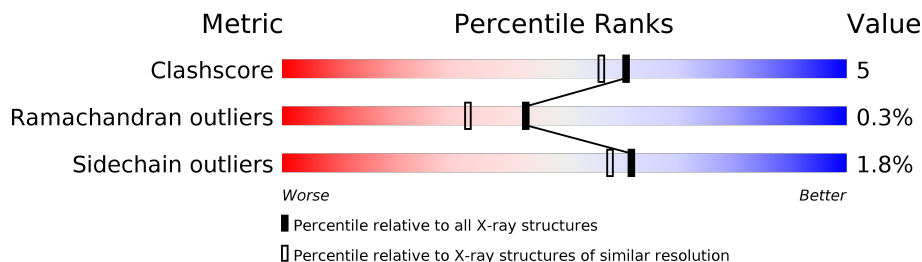
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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))
Clashscore	122078	6112 (1.90-1.90)
Ramachandran outliers	120005	6045 (1.90-1.90)
Sidechain outliers	119972	6045 (1.90-1.90)

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. 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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	614	
1	B	614	

2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 10341 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.

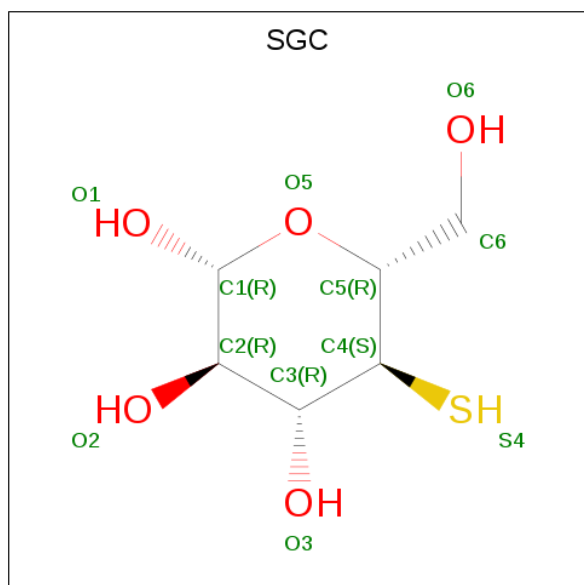
- Molecule 1 is a protein called ENDOGLUCANASE G.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	612	Total	C	N	O	S	0	0	0
			4781	3041	783	938	19			
1	B	613	Total	C	N	O	S	0	0	0
			4762	3028	780	935	19			

There are 4 discrepancies between the modelled and reference sequences:

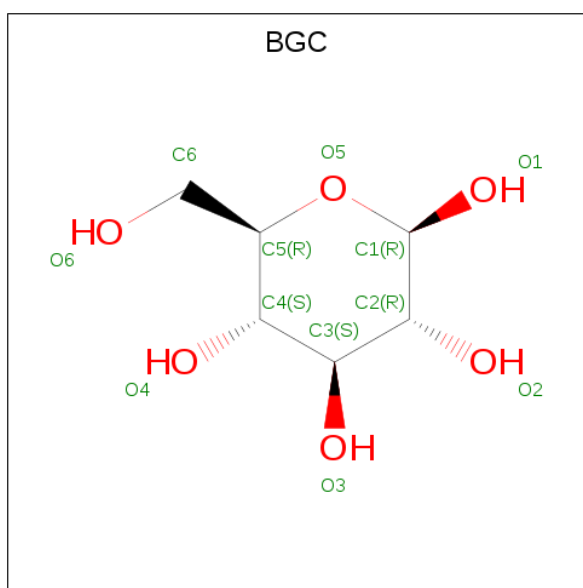
Chain	Residue	Modelled	Actual	Comment	Reference
A	574	THR	ARG	CONFLICT	UNP P37700
A	575	THR	ARG	CONFLICT	UNP P37700
B	574	THR	ARG	CONFLICT	UNP P37700
B	575	THR	ARG	CONFLICT	UNP P37700

- Molecule 2 is 4-DEOXY-4-THIO-BETA-D-GLUCOPYRANOSE (three-letter code: SGC) (formula: C₆H₁₂O₅S).



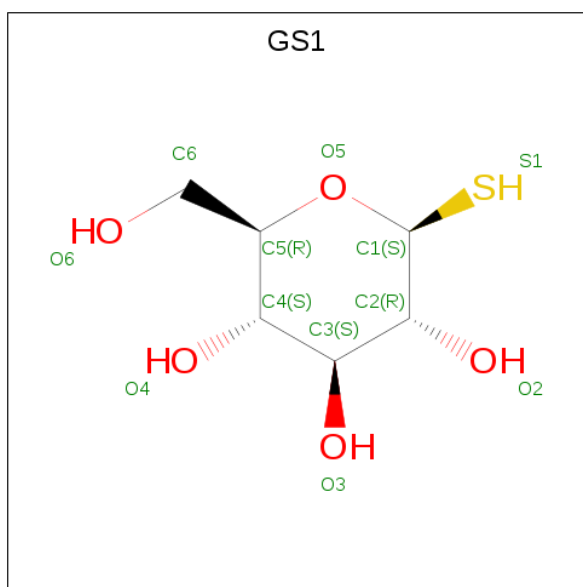
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	O	S	0	0
			11	6	4	1		
2	A	1	Total	C	O	S	0	0
			11	6	4	1		
2	B	1	Total	C	O	S	0	0
			11	6	4	1		
2	B	1	Total	C	O	S	0	0
			11	6	4	1		

- Molecule 3 is BETA-D-GLUCOSE (three-letter code: BGC) (formula: $C_6H_{12}O_6$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			11	6	5		
3	B	1	Total	C	O	0	0
			11	6	5		

- Molecule 4 is 1-THIO-BETA-D-GLUCOPYRANOSE (three-letter code: GS1) (formula: $C_6H_{12}O_5S$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	O	S	0	0
			12	6	5	1		
4	B	1	Total	C	O	S	0	0
			12	6	5	1		

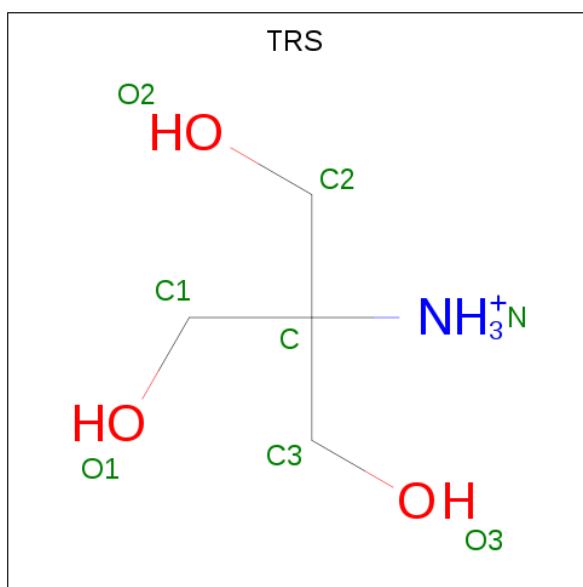
- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	2	Total	Ca	0	0
			2	2		
5	A	2	Total	Ca	0	0
			2	2		

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

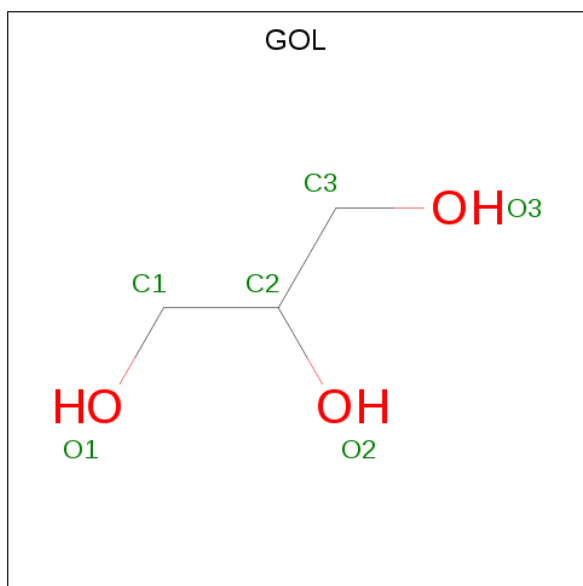
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	2	Total	Mg	0	0
			2	2		
6	A	3	Total	Mg	0	0
			3	3		

- Molecule 7 is 2-AMINO-2-HYDROXYMETHYL-PROPANE-1,3-DIOL (three-letter code: TRS) (formula: C₄H₁₂NO₃).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	A	1	Total	C	N	O	0	0
			8	4	1	3		

- Molecule 8 is GLYCEROL (three-letter code: GOL) (formula: $\text{C}_3\text{H}_8\text{O}_3$).



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

Continued on next page...

Continued from previous page...

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

- Molecule 9 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	B	1	Total	Ni	0	0
			1	1		

- Molecule 10 is water.

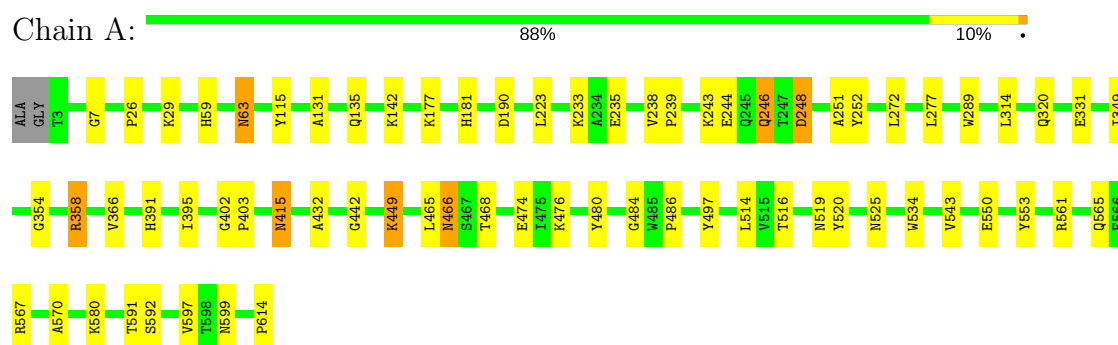
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	A	312	Total	O	0	0
			312	312		
10	B	354	Total	O	0	0
			354	354		

3 Residue-property plots [i](#)

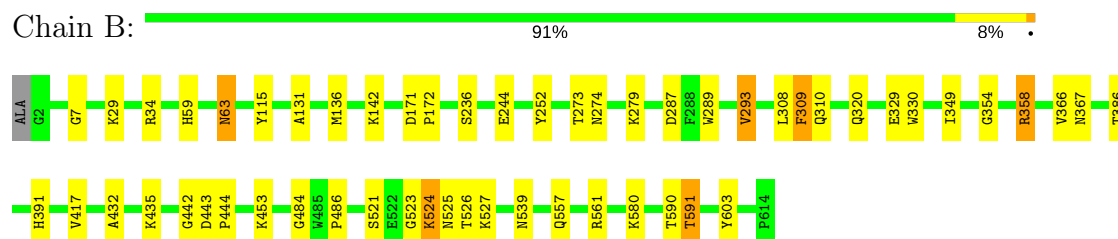
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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.

Note EDS was not executed.

• Molecule 1: ENDOGLUCANASE G



• Molecule 1: ENDOGLUCANASE G



4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	57.04Å 57.73Å 86.63Å 93.86° 100.83° 99.54°	Depositor
Resolution (Å)	42.42 – 1.90	Depositor
% Data completeness (in resolution range)	93.2 (42.42-1.90)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
Refinement program	CNS	Depositor
R, R_{free}	0.167 , 0.204	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	10341	wwPDB-VP
Average B, all atoms (Å ²)	17.0	wwPDB-VP

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: GS1, GOL, MG, BGC, NI, CA, TRS, SGC

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	0.42	0/4922	0.62	0/6710
1	B	0.43	0/4901	0.65	0/6682
All	All	0.43	0/9823	0.63	0/13392

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 [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	4781	0	4475	55	0
1	B	4762	0	4447	35	1
2	A	22	0	19	0	0
2	B	22	0	19	0	0
3	A	11	0	9	0	0
3	B	11	0	9	0	0
4	A	12	0	11	1	0
4	B	12	0	11	1	0
5	A	2	0	0	0	0
5	B	2	0	0	0	0
6	A	3	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	2	0	0	0	0
7	A	8	0	12	0	0
8	A	12	0	16	0	0
8	B	12	0	16	1	0
9	B	1	0	0	0	0
10	A	312	0	0	3	1
10	B	354	0	0	6	0
All	All	10341	0	9044	92	1

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 92 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:465:LEU:HD11	1:A:468:THR:HG23	1.43	0.97
1:A:497:TYR:HB3	1:A:543:VAL:HG22	1.65	0.76
1:B:236:SER:OG	8:B:5012:GOL:H32	1.87	0.73
1:A:314:LEU:HD13	1:A:402:GLY:HA3	1.71	0.70
4:A:618:GS1:H1	10:A:9052:HOH:O	1.91	0.70

All (1) 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:B:386:THR:O	10:A:9068:HOH:O[1_454]	2.17	0.03

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	610/614 (99%)	591 (97%)	17 (3%)	2 (0%)	43	33
1	B	611/614 (100%)	589 (96%)	20 (3%)	2 (0%)	43	33
All	All	1221/1228 (99%)	1180 (97%)	37 (3%)	4 (0%)	43	33

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	354	GLY
1	A	354	GLY
1	B	484	GLY
1	A	484	GLY

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	503/507 (99%)	494 (98%)	9 (2%)	62	58
1	B	498/507 (98%)	489 (98%)	9 (2%)	62	58
All	All	1001/1014 (99%)	983 (98%)	18 (2%)	62	58

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	486	PRO
1	B	63	ASN
1	B	391	HIS
1	A	449	LYS
1	A	466	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	544	ASN
1	B	59	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	557	GLN
1	A	605	ASN
1	B	63	ASN

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 [i](#)

Of 23 ligands modelled in this entry, 10 are monoatomic - leaving 13 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	$\# Z > 2$
8	GOL	A	5010	-	5,5,5	1.05	0	5,5,5	0.87	0
8	GOL	A	5011	-	5,5,5	1.07	0	5,5,5	0.82	0
7	TRS	A	5013	-	7,7,7	1.40	0	9,9,9	2.59	4 (44%)
2	SGC	A	615	3	10,11,12	0.74	0	12,15,17	0.39	0
3	BGC	A	616	2	11,11,12	0.78	0	15,15,17	0.58	0
2	SGC	A	617	3,4	10,11,12	0.78	0	12,15,17	0.38	0
4	GS1	A	618	2	11,12,12	0.66	0	15,17,17	0.29	0
8	GOL	B	5009	-	5,5,5	1.02	0	5,5,5	0.86	0
8	GOL	B	5012	-	5,5,5	0.91	0	5,5,5	0.92	0
2	SGC	B	615	3	10,11,12	0.79	0	12,15,17	0.28	0
3	BGC	B	616	2	11,11,12	0.72	0	15,15,17	0.50	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	SGC	B	617	3,4	10,11,12	0.98	1 (10%)	12,15,17	0.47	0
4	GS1	B	618	2	11,12,12	0.67	0	15,17,17	0.26	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
8	GOL	A	5010	-	-	0/4/4/4	0/0/0/0
8	GOL	A	5011	-	-	0/4/4/4	0/0/0/0
7	TRS	A	5013	-	-	0/9/9/9	0/0/0/0
2	SGC	A	615	3	-	0/2/19/22	0/1/1/1
3	BGC	A	616	2	-	0/2/19/22	0/1/1/1
2	SGC	A	617	3,4	-	0/2/19/22	0/1/1/1
4	GS1	A	618	2	-	0/2/22/22	0/1/1/1
8	GOL	B	5009	-	-	0/4/4/4	0/0/0/0
8	GOL	B	5012	-	-	0/4/4/4	0/0/0/0
2	SGC	B	615	3	-	0/2/19/22	0/1/1/1
3	BGC	B	616	2	-	0/2/19/22	0/1/1/1
2	SGC	B	617	3,4	-	0/2/19/22	0/1/1/1
4	GS1	B	618	2	-	0/2/22/22	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	617	SGC	C5-C4	2.21	1.55	1.53

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	5013	TRS	C2-C-C1	-3.84	100.19	111.06
7	A	5013	TRS	O1-C1-C	-3.06	101.62	110.47
7	A	5013	TRS	C3-C-N	3.33	114.80	107.73
7	A	5013	TRS	C2-C-N	3.77	115.75	107.73

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	618	GS1	1	0
8	B	5012	GOL	1	0
4	B	618	GS1	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

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.