wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2024 – 03:11 AM EST

PDB ID	:	4IPN
Title	:	The complex structure of 6-phospho-beta-glucosidase BglA-2 with thiocellobiose-6P from Streptococcus pneumoniae
Authors	:	Yu, W.L.; Jiang, Y.L.; Andreas, P.; Cheng, W.; Bai, X.H.; Ren, Y.M.; Thompsonn, J.; Zhou, C.Z.; Chen, Y.X.
Deposited on	:	2013-01-10
Resolution	:	2.41 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 (i) 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 (i)) were used in the production of this report:

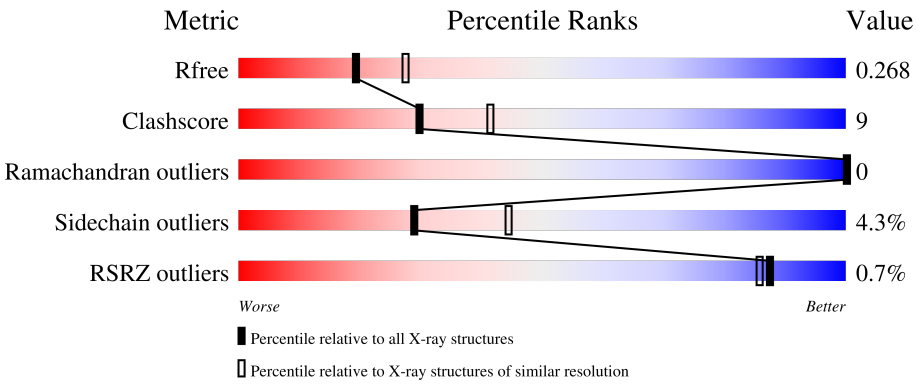
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.41 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	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)
RSRZ outliers	127900	4543 (2.44-2.40)

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 <=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	B	487	73% 19% • 6%
1	E	487	% 71% 22% • 5%
2	A	2	100%
2	C	2	50% 50%

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	SGC	A	1	-	-	-	X

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7778 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 6-phospho-beta-glucosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	457	3738	2409	604	712	13	0	0	0
1	E	463	3784	2436	612	723	13	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-15	HIS	-	expression tag	UNP Q97S37
B	-14	HIS	-	expression tag	UNP Q97S37
B	-13	HIS	-	expression tag	UNP Q97S37
B	-12	HIS	-	expression tag	UNP Q97S37
B	-11	HIS	-	expression tag	UNP Q97S37
B	-10	HIS	-	expression tag	UNP Q97S37
B	-9	SER	-	expression tag	UNP Q97S37
B	-8	SER	-	expression tag	UNP Q97S37
B	-7	GLY	-	expression tag	UNP Q97S37
B	-6	LEU	-	expression tag	UNP Q97S37
B	-5	VAL	-	expression tag	UNP Q97S37
B	-4	PRO	-	expression tag	UNP Q97S37
B	-3	ARG	-	expression tag	UNP Q97S37
B	-2	GLY	-	expression tag	UNP Q97S37
B	-1	SER	-	expression tag	UNP Q97S37
B	0	HIS	-	expression tag	UNP Q97S37
E	-15	HIS	-	expression tag	UNP Q97S37
E	-14	HIS	-	expression tag	UNP Q97S37
E	-13	HIS	-	expression tag	UNP Q97S37
E	-12	HIS	-	expression tag	UNP Q97S37
E	-11	HIS	-	expression tag	UNP Q97S37
E	-10	HIS	-	expression tag	UNP Q97S37
E	-9	SER	-	expression tag	UNP Q97S37
E	-8	SER	-	expression tag	UNP Q97S37
E	-7	GLY	-	expression tag	UNP Q97S37

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	-6	LEU	-	expression tag	UNP Q97S37
E	-5	VAL	-	expression tag	UNP Q97S37
E	-4	PRO	-	expression tag	UNP Q97S37
E	-3	ARG	-	expression tag	UNP Q97S37
E	-2	GLY	-	expression tag	UNP Q97S37
E	-1	SER	-	expression tag	UNP Q97S37
E	0	HIS	-	expression tag	UNP Q97S37

- Molecule 2 is an oligosaccharide called 6-O-phosphono-alpha-L-idopyranose-(1-4)-4-thio-beta-D-glucopyranose.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	O	P	S			
2	A	2	27	12	13	1	1	0	0	0
			Total	C	O	P	S			
2	C	2	27	12	13	1	1	0	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	B	100	100	100	0	0
			Total	O		
3	E	102	102	102	0	0

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: 6-phospho-beta-glucosidase

Chain B: 73% 19% • 6%

HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET T2 I3 D7 F8 L9 W10 A15 Q18 V19 E20 K28 V32 V35 K38 G42 K54 F60 E73 M74 G75 F76 N77 V78 W85 S86 R87 K103 E107

H125 Y136 W139 R143 M144 Y148 R158 Y159 K160 D161 K162 V163 W166 P179 F180 T181 I187 P188 K189 V211 T212 K213 E217 F222 G225 C226 M227 V228 L255 D258 P266 N267 Y268 R271 Y272 F273 N276 D277 I278 E288 L289

N292 V295 D296 F297 L298 S299 F300 S301 Y302 Y303 M304 S309 A310 L311 P312 THR GLN TYR ASN SER GLY GLU GLY ASN ILE ILE GLY G325 L326 V327 N328 P329 E335 W336 I340 D341 P342 Y356 F361 E364 K370 D371 E377 D385 D386 Y387 R388 G411

G416 D419 C420 L428 S429 K430 G433 Y436 R439 T446 L447 K452 F455 T456 K459 E463 F470 LYS

• Molecule 1: 6-phospho-beta-glucosidase

Chain E: % 71% 22% • 5%

HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET T2 I3 D7 N17 Q18 V19 E20 Y23 K28 V32 V35 L36 D51 K54 L55 D59 F60 Y61 H62 K63 Y64 K65 E66 S69 E73 F76 R87 I88 F89 P90

K91 G92 D93 E94 D106 D110 I117 H125 Y126 E127 A133 R142 R158 K162 N173 E177 L178 D186 I187 P188 L197 I201 L205 L210 R216 N219 F222 C226 M227 V228 L229 A230 M236 T237 P238 E248 N253 Y254

L255 F256 V259 P266 N267 Y268 F273 I278 V295 D296 S299 F300 S301 T307 L311 Q314 Y315 N316 S317 G318 GLU GLY ASN ILE ILE GLY G325 L326 E335 I347 I348 R351 F361 E364 K370 D371 E377 T382 V383 Q384 D385 R388

Q404 D405 W415 D419 S422 M423 S424 T425 A426 S429 K430 R431 Y432 G433 Y436 V437 D438 R439 N440 D441 S444 G445 T446 L447 N448 R449 Y450 K451 K452 K459 F470 LYS

• Molecule 2: 6-O-phosphono-alpha-L-idopyranose-(1-4)-4-thio-beta-D-glucopyranose

Chain A: 100%

- Molecule 2: 6-O-phosphono-alpha-L-idopyranose-(1-4)-4-thio-beta-D-glucopyranose

Chain C:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	184.03Å 66.65Å 133.35Å 90.00° 136.33° 90.00°	Depositor
Resolution (Å)	45.95 – 2.41 45.95 – 2.41	Depositor EDS
% Data completeness (in resolution range)	97.8 (45.95-2.41) 98.0 (45.95-2.41)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.42 (at 2.42Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.1_743)	Depositor
R, R_{free}	0.174 , 0.251 0.197 , 0.268	Depositor DCC
R_{free} test set	2147 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	37.2	Xtriage
Anisotropy	1.363	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 44.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.006 for -h-2*l,k,h+l 0.016 for h,-k,-h-l 0.018 for h+2*l,-k,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7778	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.16% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: RTG, 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	B	0.37	0/3837	0.56	0/5204
1	E	0.38	0/3884	0.56	0/5268
All	All	0.37	0/7721	0.56	0/10472

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	B	3738	0	3591	61	0
1	E	3784	0	3629	78	0
2	A	27	0	10	1	0
2	C	27	0	10	3	0
3	B	100	0	0	2	0
3	E	102	0	0	0	0
All	All	7778	0	7240	139	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 9.

The worst 5 of 139 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:E:177:GLU:C	1:E:178:LEU:HD23	1.87	0.93
1:E:382:THR:HG22	1:E:383:VAL:H	1.51	0.75
1:E:364:GLU:OE1	2:C:2:RTG:O2	2.05	0.74
1:E:125:HIS:O	1:E:127:GLU:HG2	1.91	0.70
1:E:382:THR:HG23	1:E:451:LYS:O	1.91	0.69

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	B	453/487 (93%)	443 (98%)	10 (2%)	0	100	100
1	E	459/487 (94%)	449 (98%)	10 (2%)	0	100	100
All	All	912/974 (94%)	892 (98%)	20 (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	B	402/427 (94%)	384 (96%)	18 (4%)	27	42
1	E	407/427 (95%)	390 (96%)	17 (4%)	30	46

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	809/854 (95%)	774 (96%)	35 (4%)	29 44

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	311	LEU
1	E	314	GLN
1	E	377	GLU
1	B	311	LEU
1	B	289	LEU

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)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

4 monosaccharides 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	SGC	A	1	2	11,12,12	2.53	3 (27%)	13,17,17	0.89	1 (7%)
2	RTG	A	2	2	15,15,16	2.36	3 (20%)	22,22,24	1.47	4 (18%)
2	SGC	C	1	2	11,12,12	2.87	4 (36%)	13,17,17	0.93	0
2	RTG	C	2	2	15,15,16	2.53	3 (20%)	22,22,24	1.08	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	SGC	A	1	2	-	2/2/22/22	0/1/1/1
2	RTG	A	2	2	-	3/6/23/26	0/1/1/1
2	SGC	C	1	2	-	2/2/22/22	0/1/1/1
2	RTG	C	2	2	-	2/6/23/26	0/1/1/1

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	2	RTG	C2-C3	-7.73	1.41	1.52
2	C	1	SGC	C3-C4	-6.90	1.47	1.53
2	A	2	RTG	C2-C3	-6.77	1.42	1.52
2	A	1	SGC	C3-C4	-5.72	1.48	1.53
2	C	1	SGC	C3-C2	-4.91	1.39	1.52

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	2	RTG	C3-C4-C5	3.87	117.14	110.24
2	C	2	RTG	O6-C6-C5	2.60	117.93	108.99
2	C	2	RTG	C3-C4-C5	2.51	114.72	110.24
2	A	2	RTG	O5-C5-C4	2.38	116.62	110.83
2	A	2	RTG	C1-O5-C5	2.33	115.35	112.19

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	2	RTG	C5-C6-O6-PBA
2	A	2	RTG	O5-C5-C6-O6
2	A	2	RTG	C4-C5-C6-O6
2	C	2	RTG	C6-O6-PBA-OAJ
2	C	2	RTG	C5-C6-O6-PBA

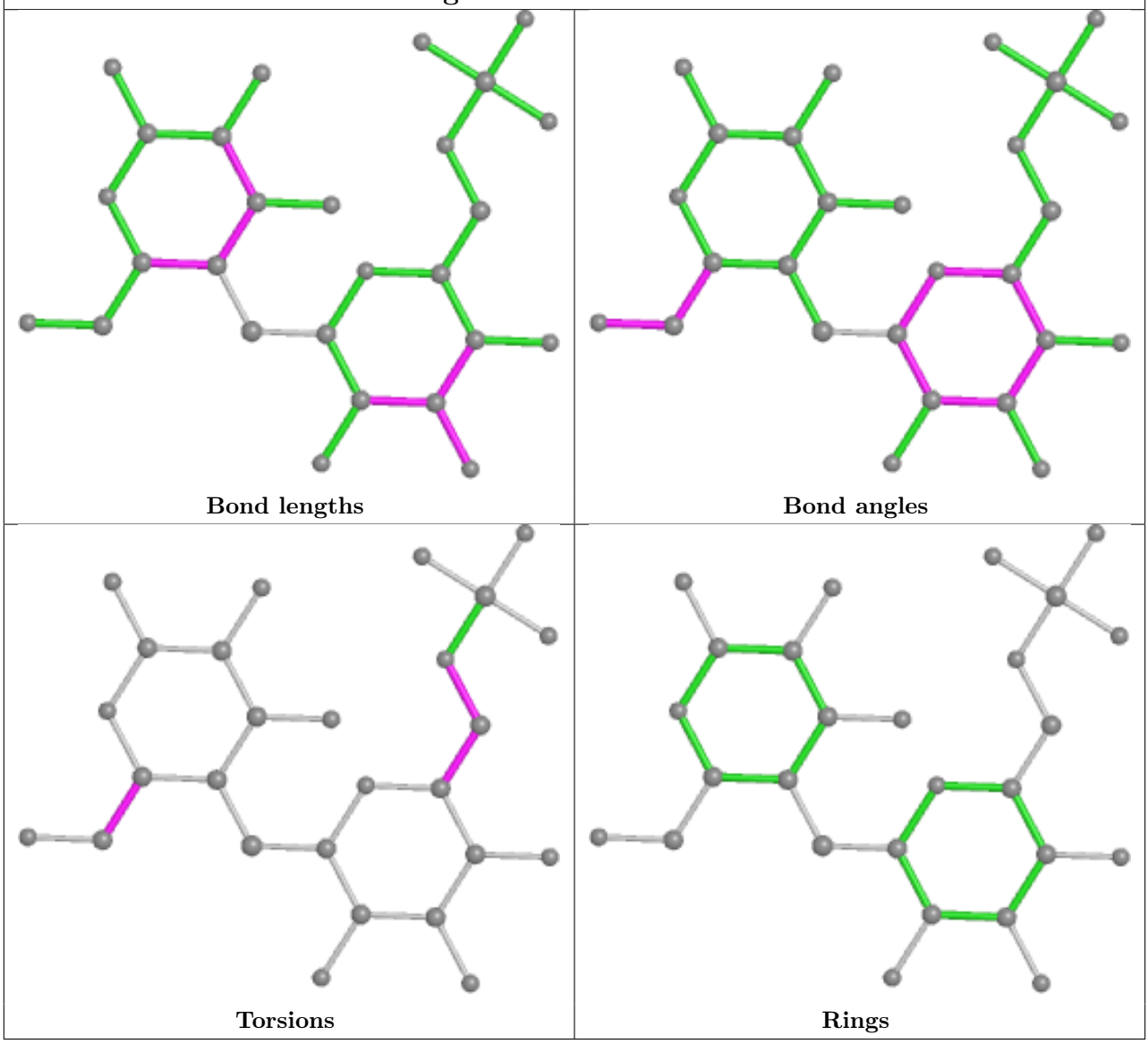
There are no ring outliers.

3 monomers are involved in 4 short contacts:

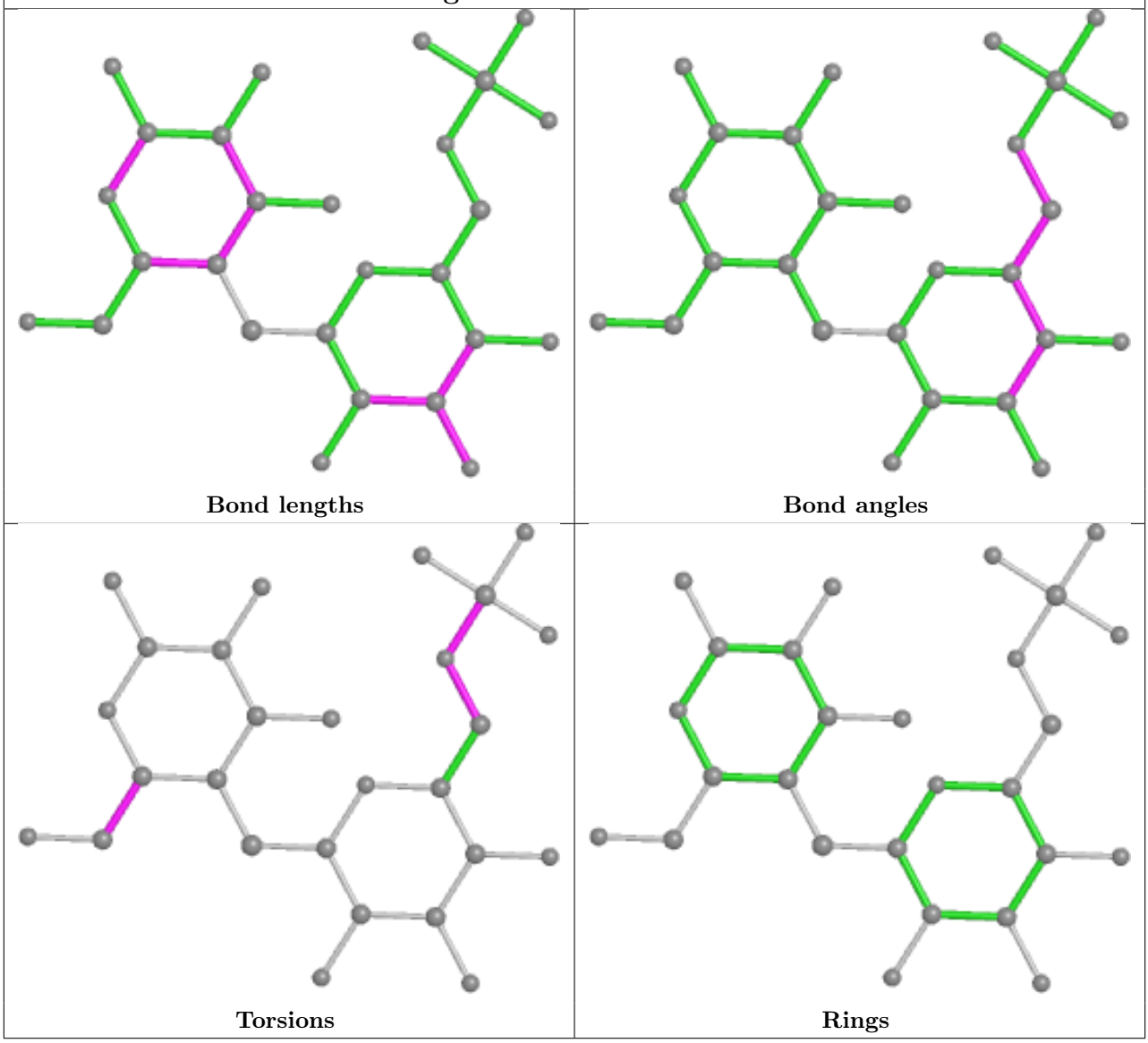
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1	SGC	1	0
2	C	2	RTG	3	0
2	A	2	RTG	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

Oligosaccharide Chain A



Oligosaccharide Chain C



5.6 Ligand geometry (i)

There are no ligands in this entry.

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, 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	B	457/487 (93%)	-0.17	2 (0%)	92	91	29, 47, 67, 84	0
1	E	463/487 (95%)	-0.05	4 (0%)	84	82	30, 50, 70, 97	0
All	All	920/974 (94%)	-0.11	6 (0%)	87	86	29, 48, 69, 97	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	335	GLU	2.8
1	B	377	GLU	2.5
1	E	316	ASN	2.4
1	E	317	SER	2.4
1	E	377	GLU	2.3

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)

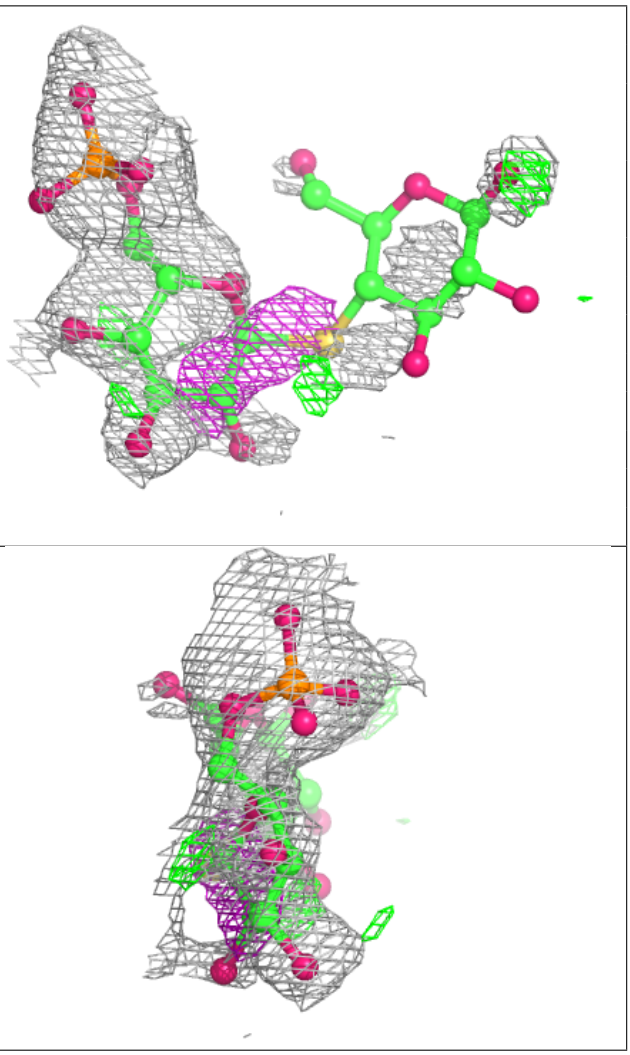
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	SGC	A	1	12/12	0.63	0.54	63,72,79,81	12
2	RTG	A	2	15/16	0.82	0.25	36,53,59,60	15
2	SGC	C	1	12/12	0.85	0.24	73,76,82,87	12
2	RTG	C	2	15/16	0.86	0.23	60,64,69,72	15

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

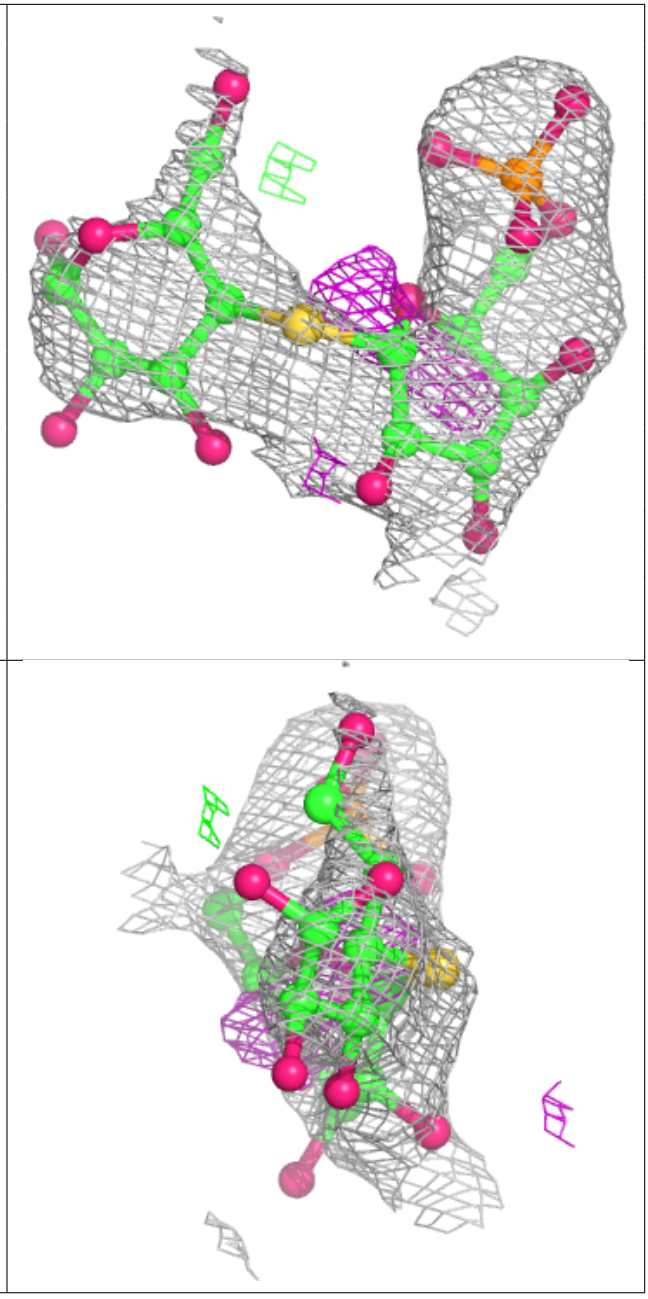
Electron density around Chain A:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



Electron density around Chain C:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.