wwPDB X-ray Structure Validation Summary Report (i)

Sep 12, 2023 – 03:17 PM EDT

PDB ID	:	4OAA
Title	:	Crystal structure of E. coli lactose permease G46W,G262W bound to sugar
Authors	:	Kumar, H.; Kasho, V.; Smirnova, I.; Finer-Moore, J.; Kaback, H.R.; Stroud, R.M.
Deposited on	:	2014-01-03
Resolution	:	3.50 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:

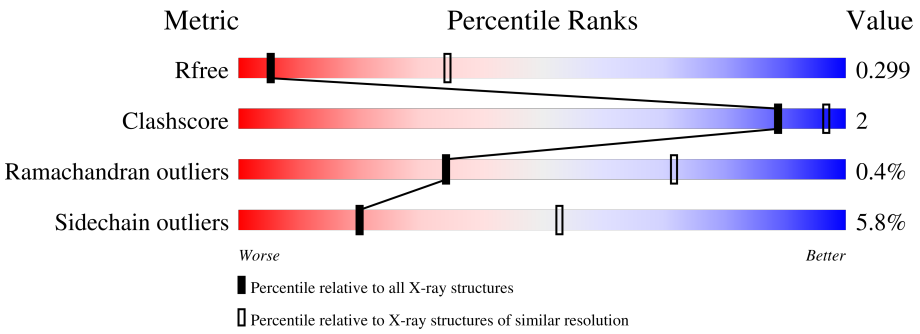
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.50 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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)

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%

Mol	Chain	Length	Quality of chain
1	A	417	86% 8% 6%
1	B	417	83% 11% 6%
2	E	2	100%
2	F	2	50% 50%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 12564 atoms, of which 6307 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 Lactose/galactose transporter.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	391	6230	2115	3127	468	500	20	0	0	0
1	B	391	6244	2118	3136	468	501	21	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	46	TRP	GLY	engineered mutation	UNP B1XBJ1
A	262	TRP	GLY	engineered mutation	UNP B1XBJ1
B	46	TRP	GLY	engineered mutation	UNP B1XBJ1
B	262	TRP	GLY	engineered mutation	UNP B1XBJ1

- Molecule 2 is an oligosaccharide called beta-D-galactopyranose-(1-1)-1-thio-beta-D-galactopyranose.

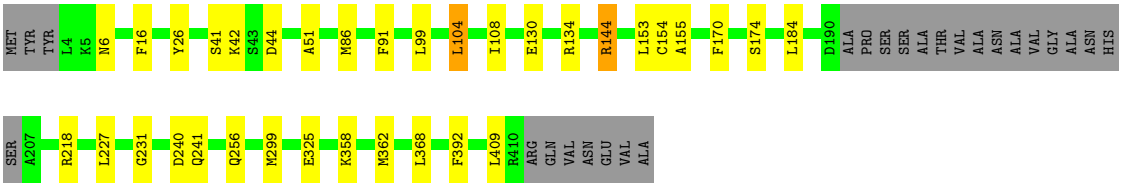
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	H	O	S			
2	E	2	45	12	22	10	1	0	0	0
2	F	2	45	12	22	10	1	0	0	0

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. 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. 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.

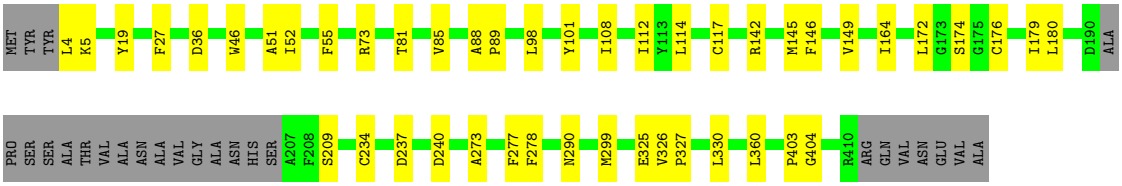
• Molecule 1: Lactose/galactose transporter

Chain A:



• Molecule 1: Lactose/galactose transporter

Chain B:



• Molecule 2: beta-D-galactopyranose-(1-1)-1-thio-beta-D-galactopyranose

Chain E:

• Molecule 2: beta-D-galactopyranose-(1-1)-1-thio-beta-D-galactopyranose

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	101.56Å 121.94Å 264.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.00 – 3.50 44.00 – 2.98	Depositor EDS
% Data completeness (in resolution range)	99.4 (44.00-3.50) 86.0 (44.00-2.98)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.08 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069), ELVES	Depositor
R, R_{free}	0.257 , 0.293 0.255 , 0.299	Depositor DCC
R_{free} test set	1491 reflections (5.14%)	wwPDB-VP
Wilson B-factor (Å^2)	51.8	Xtriage
Anisotropy	0.531	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 62.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	12564	wwPDB-VP
Average B, all atoms (Å^2)	170.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.96% 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: YIO, GAL

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	0.31	0/3199	0.50	0/4337
1	B	0.32	0/3204	0.51	0/4343
All	All	0.32	0/6403	0.51	0/8680

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	3103	3127	3128	10	0
1	B	3108	3136	3140	15	0
2	E	23	22	21	2	0
2	F	23	22	21	0	0
All	All	6257	6307	6310	24	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 2.

The worst 5 of 24 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:B:85:VAL:O	1:B:174:SER:OG	2.21	0.58
1:B:55:PHE:CD2	1:B:112:ILE:HG22	2.39	0.57
1:A:104:LEU:O	1:A:108:ILE:N	2.42	0.53
1:B:19:TYR:OH	1:B:117:CYS:O	2.17	0.53
1:A:154:CYS:SG	1:A:155:ALA:N	2.85	0.50

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	387/417 (93%)	361 (93%)	25 (6%)	1 (0%)	41	75
1	B	387/417 (93%)	358 (92%)	27 (7%)	2 (0%)	29	68
All	All	774/834 (93%)	719 (93%)	52 (7%)	3 (0%)	34	72

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	241	GLN
1	B	403	PRO
1	B	164	ILE

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	324/348 (93%)	304 (94%)	20 (6%)	18	51
1	B	326/348 (94%)	308 (94%)	18 (6%)	21	54
All	All	650/696 (93%)	612 (94%)	38 (6%)	20	53

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	142	ARG
1	B	278	PHE
1	B	172	LEU
1	B	237	ASP
1	B	360	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	YIO	E	1	2	11,12,12	0.40	0	15,17,17	1.11	2 (13%)
2	GAL	E	2	2	11,11,12	0.26	0	15,15,17	0.92	1 (6%)
2	YIO	F	1	2	11,12,12	0.37	0	15,17,17	0.98	1 (6%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	GAL	F	2	2	11,11,12	0.33	0	15,15,17	1.13	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
2	YIO	E	1	2	-	1/2/22/22	0/1/1/1
2	GAL	E	2	2	-	1/2/19/22	0/1/1/1
2	YIO	F	1	2	-	2/2/22/22	0/1/1/1
2	GAL	F	2	2	-	2/2/19/22	0/1/1/1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	2	GAL	C1-O5-C5	3.10	116.39	112.19
2	F	1	YIO	O5-C5-C6	2.57	112.83	106.44
2	E	1	YIO	O3-C3-C2	-2.15	105.38	110.35
2	E	1	YIO	O4-C4-C3	-2.14	105.40	110.35

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	F	2	GAL	O5-C5-C6-O6
2	F	1	YIO	O5-C5-C6-O6
2	E	2	GAL	O5-C5-C6-O6
2	E	1	YIO	O5-C5-C6-O6
2	F	2	GAL	C4-C5-C6-O6

There are no ring outliers.

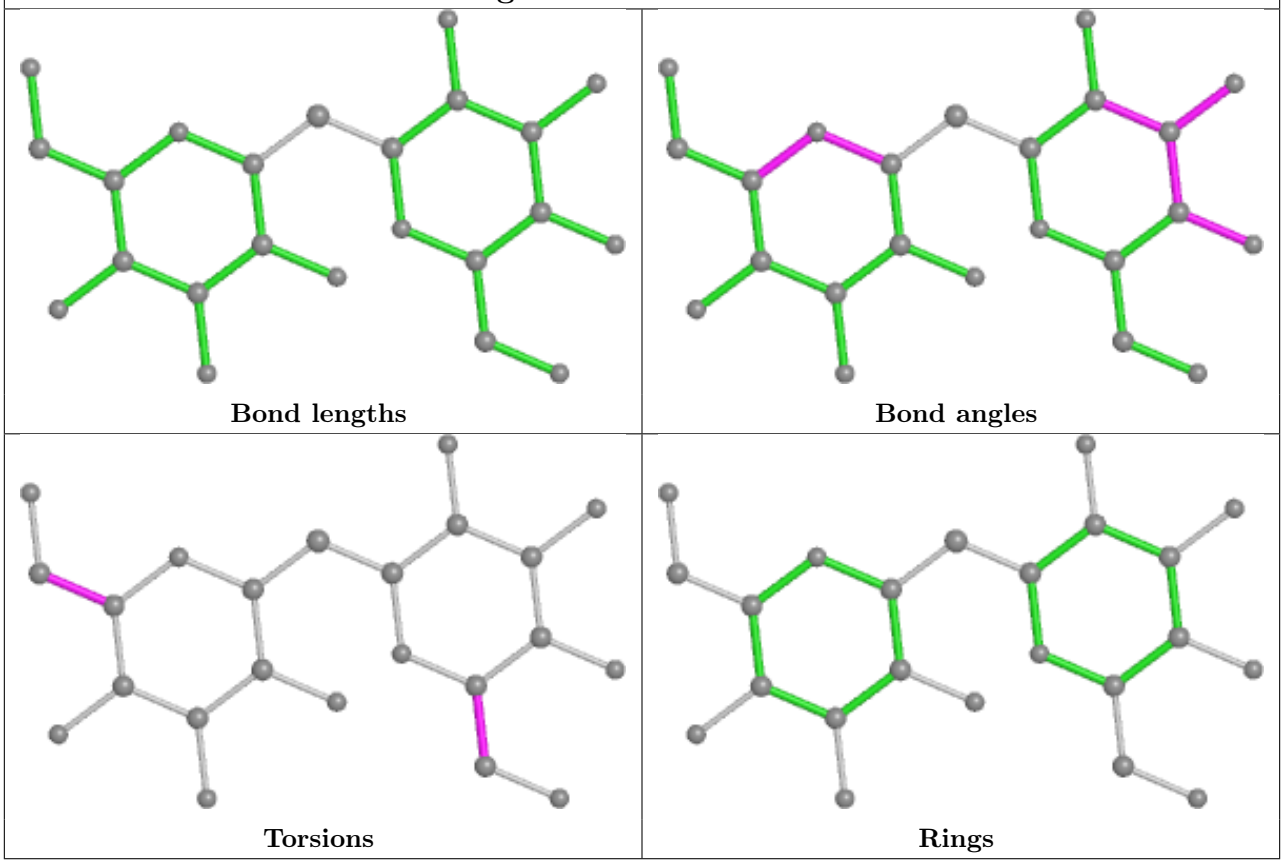
2 monomers are involved in 2 short contacts:

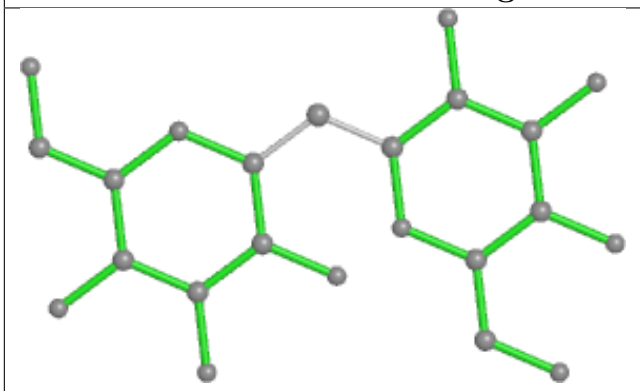
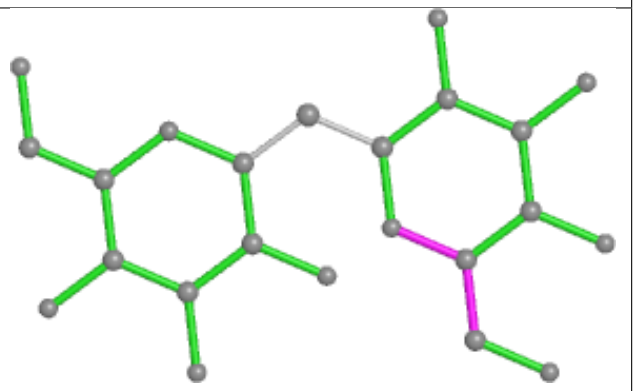
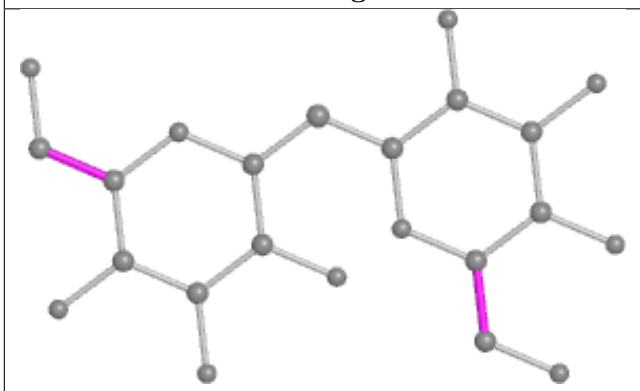
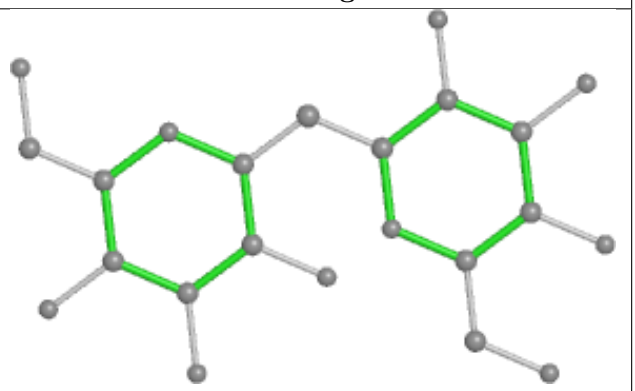
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	1	YIO	1	0
2	E	2	GAL	1	0

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 E



Oligosaccharide Chain F	
 Bond lengths	 Bond angles
 Torsions	 Rings

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)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

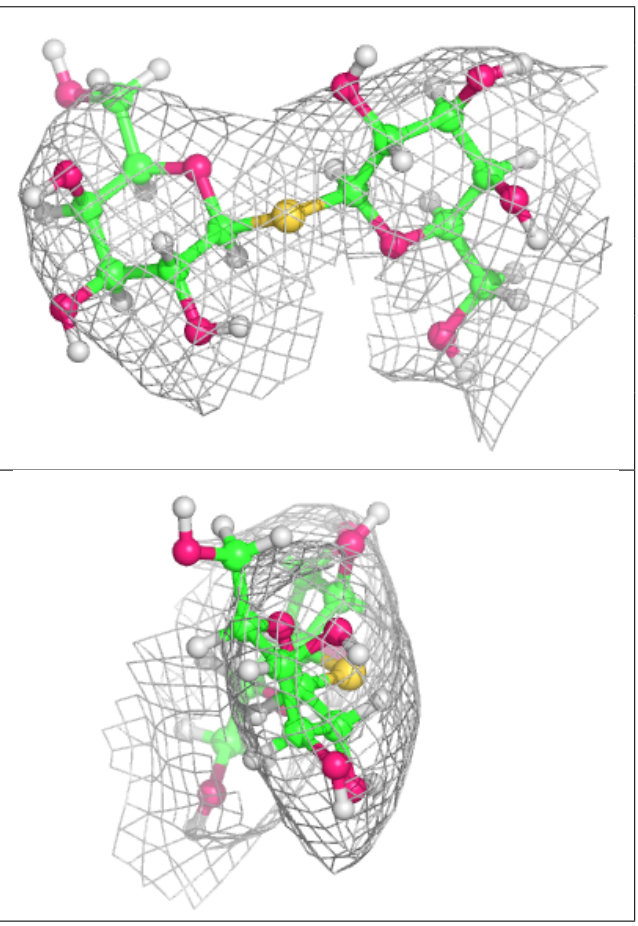
6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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.

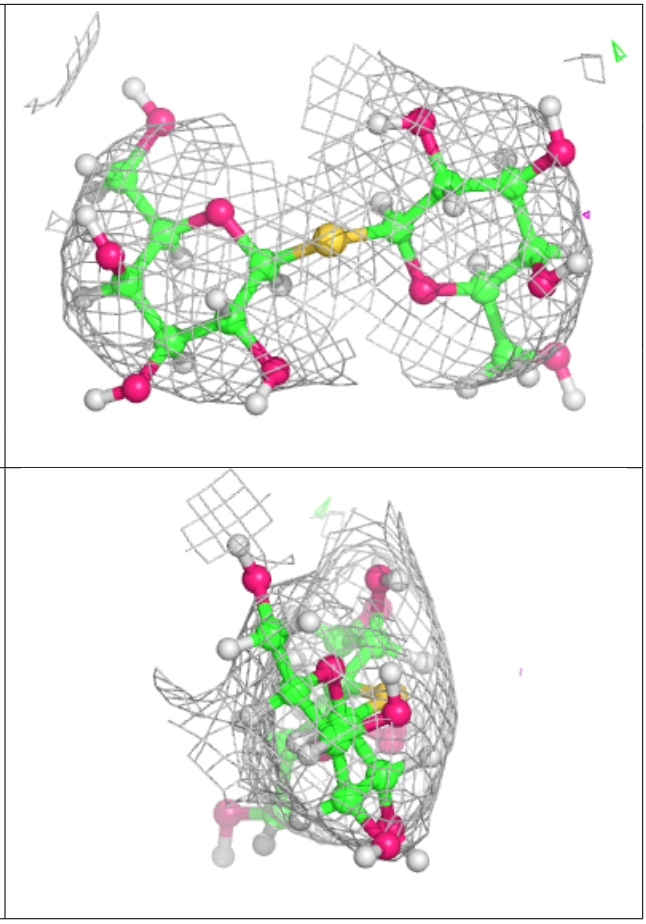
Electron density around Chain E:

$2mF_o$-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 F:

$2mF_o$-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)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.