wwPDB X-ray Structure Validation Summary Report (i)

Jun 24, 2024 – 12:36 PM EDT

PDB ID	:	8SBU
Title	:	Crystal structure of MBP fusion with HPPK from Methanocaldococcus jannaschii
Authors	:	Shaw, G.X.; Needle, D.; Stair, N.R.; Cherry, S.; Tropea, J.E.; Waugh, D.S.; Ji, X.
Deposited on	:	2023-04-04
Resolution	:	2.20 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 (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:

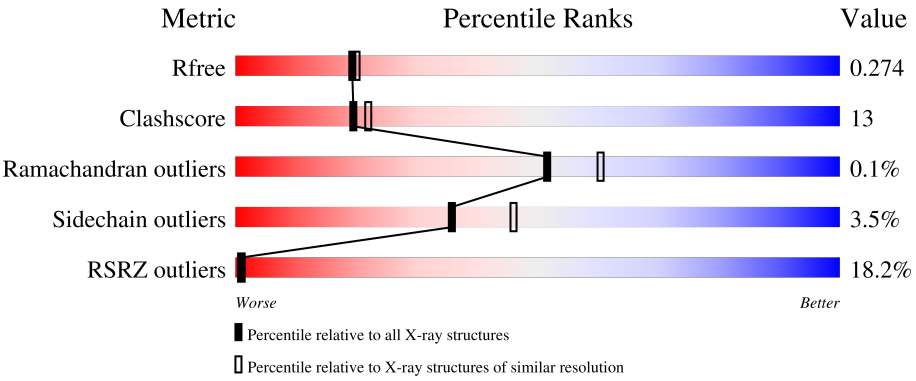
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	A	618	7% 74% 18% • 5%
1	B	618	28% 72% 20% • 6%
2	C	2	50% 50%
2	D	2	100%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9492 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 Maltose/maltodextrin-binding periplasmic protein,6-hydroxy methyl-7,8-dihydropterin pyrophosphokinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	585	Total	C	N	O	S	0	0	0
			4564	2947	745	858	14			
1	B	583	Total	C	N	O	S	0	0	0
			4543	2935	742	852	14			

There are 86 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	LYS	deletion	UNP P0AEX9
A	?	-	THR	deletion	UNP P0AEX9
A	?	-	GLY	deletion	UNP P0AEX9
A	?	-	ALA	deletion	UNP P0AEX9
A	?	-	ARG	deletion	UNP P0AEX9
A	?	-	ILE	deletion	UNP P0AEX9
A	?	-	LEU	deletion	UNP P0AEX9
A	?	-	ALA	deletion	UNP P0AEX9
A	?	-	LEU	deletion	UNP P0AEX9
A	?	-	SER	deletion	UNP P0AEX9
A	?	-	ALA	deletion	UNP P0AEX9
A	?	-	LEU	deletion	UNP P0AEX9
A	?	-	THR	deletion	UNP P0AEX9
A	?	-	THR	deletion	UNP P0AEX9
A	?	-	MET	deletion	UNP P0AEX9
A	?	-	MET	deletion	UNP P0AEX9
A	?	-	PHE	deletion	UNP P0AEX9
A	?	-	SER	deletion	UNP P0AEX9
A	?	-	ALA	deletion	UNP P0AEX9
A	?	-	SER	deletion	UNP P0AEX9
A	?	-	ALA	deletion	UNP P0AEX9
A	?	-	LEU	deletion	UNP P0AEX9
A	?	-	ALA	deletion	UNP P0AEX9
A	?	-	LYS	deletion	UNP P0AEX9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	ILE	deletion	UNP P0AEX9
A	4	HIS	-	expression tag	UNP P0AEX9
A	5	HIS	-	expression tag	UNP P0AEX9
A	6	HIS	-	expression tag	UNP P0AEX9
A	7	HIS	-	expression tag	UNP P0AEX9
A	8	HIS	-	expression tag	UNP P0AEX9
A	9	HIS	-	expression tag	UNP P0AEX9
A	89	ALA	ASP	engineered mutation	UNP P0AEX9
A	90	ALA	LYS	engineered mutation	UNP P0AEX9
A	179	ALA	GLU	engineered mutation	UNP P0AEX9
A	180	ALA	ASN	engineered mutation	UNP P0AEX9
A	246	ALA	LYS	engineered mutation	UNP P0AEX9
A	366	ALA	GLU	engineered mutation	UNP P0AEX9
A	369	ALA	LYS	engineered mutation	UNP P0AEX9
A	370	ALA	ASP	engineered mutation	UNP P0AEX9
A	374	ASN	ARG	engineered mutation	UNP P0AEX9
A	375	ALA	ILE	engineered mutation	UNP P0AEX9
A	376	ALA	THR	engineered mutation	UNP P0AEX9
A	377	ALA	LYS	engineered mutation	UNP P0AEX9
B	?	-	LYS	deletion	UNP P0AEX9
B	?	-	THR	deletion	UNP P0AEX9
B	?	-	GLY	deletion	UNP P0AEX9
B	?	-	ALA	deletion	UNP P0AEX9
B	?	-	ARG	deletion	UNP P0AEX9
B	?	-	ILE	deletion	UNP P0AEX9
B	?	-	LEU	deletion	UNP P0AEX9
B	?	-	ALA	deletion	UNP P0AEX9
B	?	-	LEU	deletion	UNP P0AEX9
B	?	-	SER	deletion	UNP P0AEX9
B	?	-	ALA	deletion	UNP P0AEX9
B	?	-	LEU	deletion	UNP P0AEX9
B	?	-	THR	deletion	UNP P0AEX9
B	?	-	THR	deletion	UNP P0AEX9
B	?	-	MET	deletion	UNP P0AEX9
B	?	-	MET	deletion	UNP P0AEX9
B	?	-	PHE	deletion	UNP P0AEX9
B	?	-	SER	deletion	UNP P0AEX9
B	?	-	ALA	deletion	UNP P0AEX9
B	?	-	SER	deletion	UNP P0AEX9
B	?	-	ALA	deletion	UNP P0AEX9
B	?	-	LEU	deletion	UNP P0AEX9
B	?	-	ALA	deletion	UNP P0AEX9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	LYS	deletion	UNP P0AEX9
B	?	-	ILE	deletion	UNP P0AEX9
B	4	HIS	-	expression tag	UNP P0AEX9
B	5	HIS	-	expression tag	UNP P0AEX9
B	6	HIS	-	expression tag	UNP P0AEX9
B	7	HIS	-	expression tag	UNP P0AEX9
B	8	HIS	-	expression tag	UNP P0AEX9
B	9	HIS	-	expression tag	UNP P0AEX9
B	89	ALA	ASP	engineered mutation	UNP P0AEX9
B	90	ALA	LYS	engineered mutation	UNP P0AEX9
B	179	ALA	GLU	engineered mutation	UNP P0AEX9
B	180	ALA	ASN	engineered mutation	UNP P0AEX9
B	246	ALA	LYS	engineered mutation	UNP P0AEX9
B	366	ALA	GLU	engineered mutation	UNP P0AEX9
B	369	ALA	LYS	engineered mutation	UNP P0AEX9
B	370	ALA	ASP	engineered mutation	UNP P0AEX9
B	374	ASN	ARG	engineered mutation	UNP P0AEX9
B	375	ALA	ILE	engineered mutation	UNP P0AEX9
B	376	ALA	THR	engineered mutation	UNP P0AEX9
B	377	ALA	LYS	engineered mutation	UNP P0AEX9

- Molecule 2 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	C	2	Total C O 23 12 11	0	0	0
2	D	2	Total C O 23 12 11	0	0	0

- Molecule 3 is water.

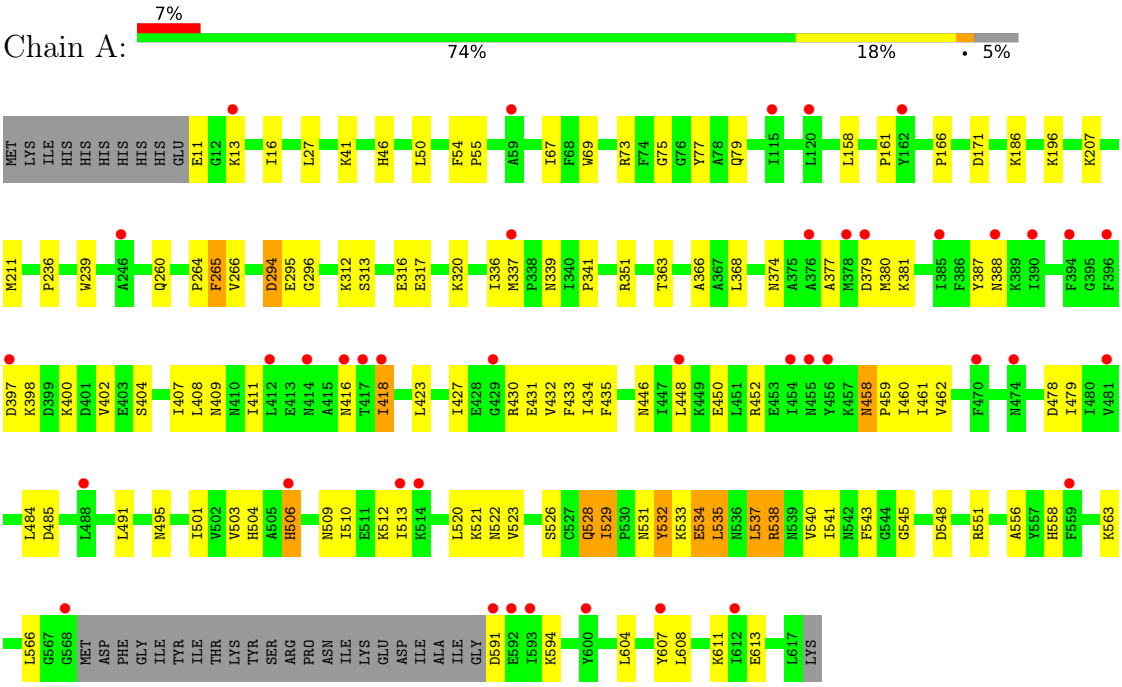
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	183	Total O 183 183	0	0
3	B	156	Total O 156 156	0	0

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.

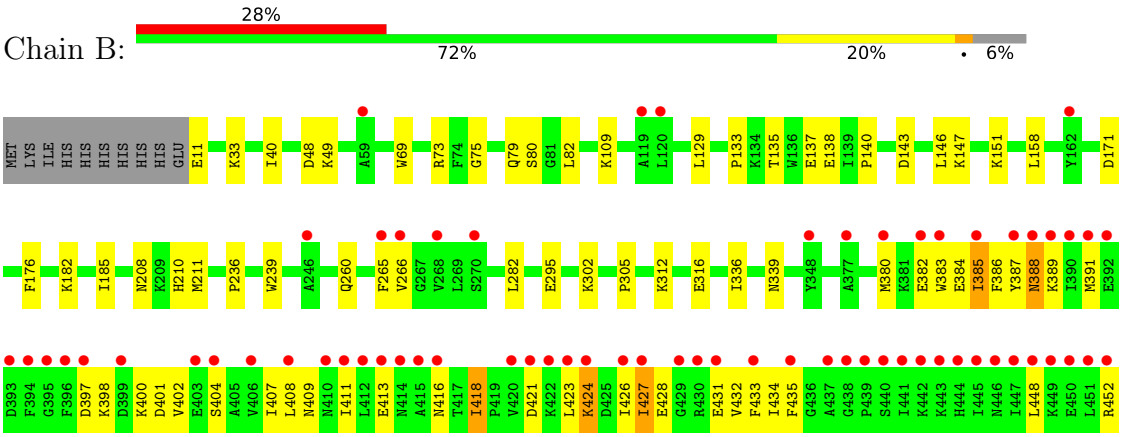
• Molecule 1: Maltose/maltodextrin-binding periplasmic protein,6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase

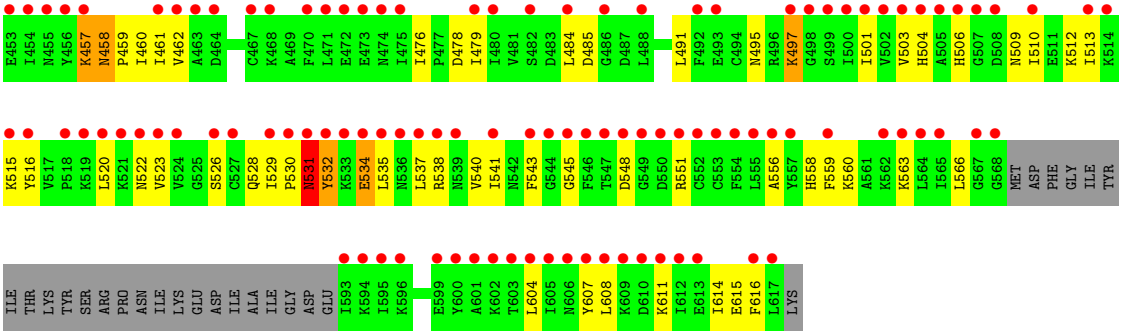
Chain A:



• Molecule 1: Maltose/maltodextrin-binding periplasmic protein,6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase

Chain B:



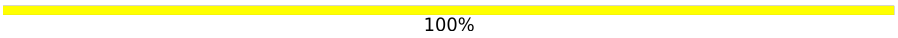


• Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain C:

• Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	58.17Å 69.83Å 73.12Å 67.44° 74.63° 78.13°	Depositor
Resolution (Å)	29.28 – 2.20 29.28 – 2.19	Depositor EDS
% Data completeness (in resolution range)	97.9 (29.28-2.20) 97.9 (29.28-2.19)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.43 (at 2.20Å)	Xtriage
Refinement program	PHENIX v1.16	Depositor
R, R_{free}	0.221 , 0.274 0.221 , 0.274	Depositor DCC
R_{free} test set	1002 reflections (1.96%)	wwPDB-VP
Wilson B-factor (Å^2)	41.3	Xtriage
Anisotropy	0.229	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 57.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	9492	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.51% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GLC

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	0.28	0/4663	0.46	0/6313
1	B	0.33	0/4642	0.49	0/6286
All	All	0.31	0/9305	0.47	0/12599

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	4564	0	4578	91	0
1	B	4543	0	4557	144	0
2	C	23	0	21	0	0
2	D	23	0	21	0	0
3	A	183	0	0	6	0
3	B	156	0	0	1	0
All	All	9492	0	9177	235	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 13.

The worst 5 of 235 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:402:VAL:HA	1:B:529:ILE:CD1	1.58	1.31
1:B:409:ASN:HB2	1:B:537:LEU:CD1	1.74	1.18
1:B:402:VAL:CA	1:B:529:ILE:HD11	1.73	1.17
1:B:402:VAL:HA	1:B:529:ILE:HD11	1.15	1.14
1:B:409:ASN:HB2	1:B:537:LEU:HD11	1.33	1.06

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	581/618 (94%)	559 (96%)	22 (4%)	0	100	100
1	B	579/618 (94%)	555 (96%)	23 (4%)	1 (0%)	47	55
All	All	1160/1236 (94%)	1114 (96%)	45 (4%)	1 (0%)	51	60

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	531	ASN

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	476/506 (94%)	462 (97%)	14 (3%)	42	54
1	B	473/506 (94%)	454 (96%)	19 (4%)	31	40
All	All	949/1012 (94%)	916 (96%)	33 (4%)	36	46

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	531	ASN
1	B	532	TYR
1	B	538	ARG
1	A	537	LEU
1	A	535	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	388	ASN
1	B	416	ASN
1	B	506	HIS
1	B	458	ASN
1	B	260	GLN

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	GLC	C	1	2	12,12,12	0.55	0	17,17,17	0.91	0
2	GLC	C	2	2	11,11,12	0.33	0	15,15,17	1.19	3 (20%)
2	GLC	D	1	2	12,12,12	0.53	0	17,17,17	0.84	1 (5%)
2	GLC	D	2	2	11,11,12	0.35	0	15,15,17	1.08	2 (13%)

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	GLC	C	1	2	-	2/2/22/22	0/1/1/1
2	GLC	C	2	2	-	0/2/19/22	0/1/1/1
2	GLC	D	1	2	-	1/2/22/22	0/1/1/1
2	GLC	D	2	2	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	1	GLC	C1-O5-C5	-2.30	109.32	113.66
2	C	2	GLC	C2-C3-C4	-2.22	107.05	110.89
2	C	2	GLC	C1-O5-C5	-2.17	109.25	112.19
2	D	2	GLC	C6-C5-C4	-2.11	108.05	113.00
2	C	2	GLC	C6-C5-C4	-2.11	108.06	113.00

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	1	GLC	C4-C5-C6-O6
2	C	1	GLC	O5-C5-C6-O6
2	D	1	GLC	C4-C5-C6-O6

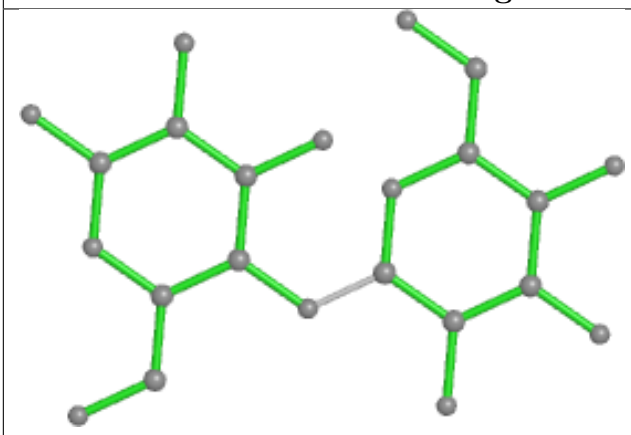
There are no ring outliers.

No monomer is involved in short contacts.

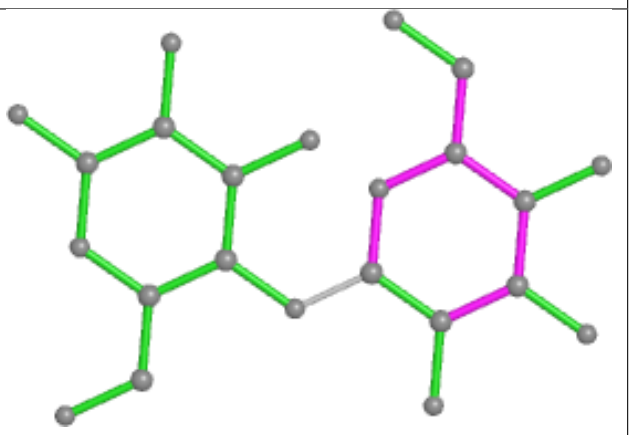
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths,

bond angles, torsion angles, and ring geometry for oligosaccharide.

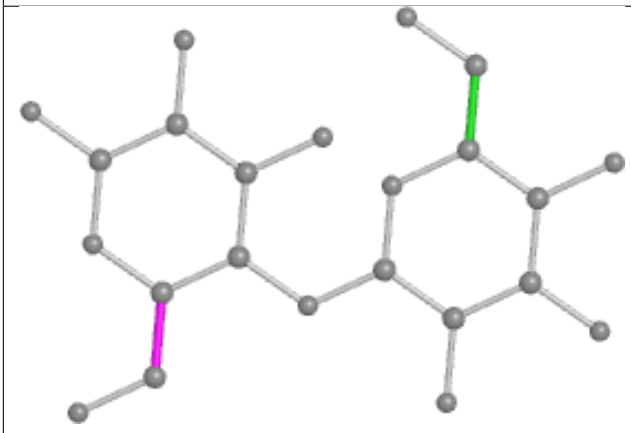
Oligosaccharide Chain C



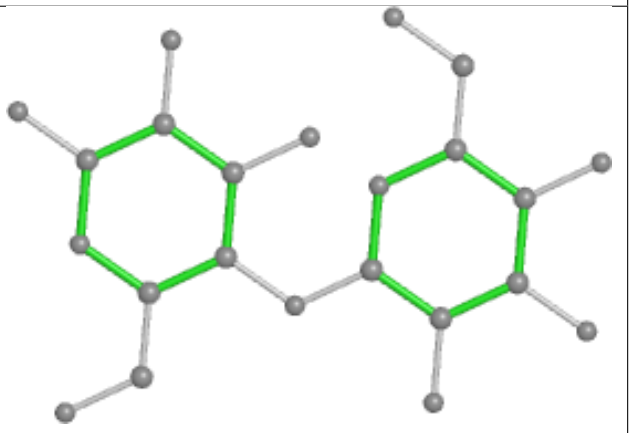
Bond lengths



Bond angles

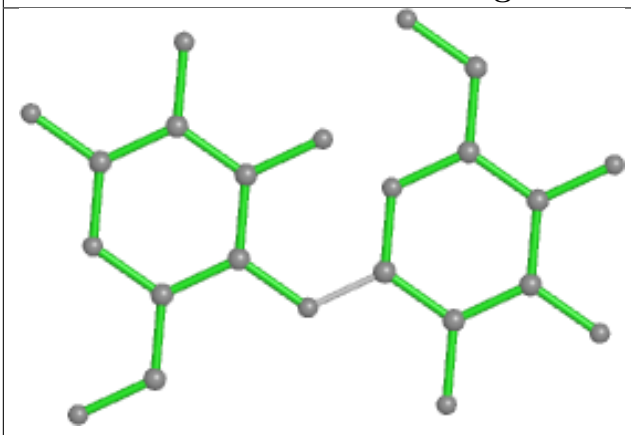


Torsions

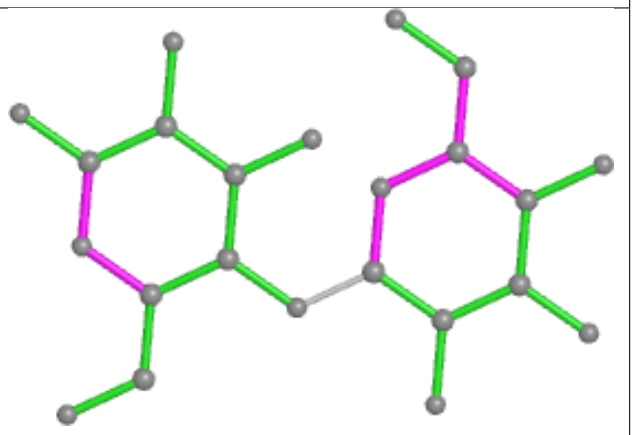


Rings

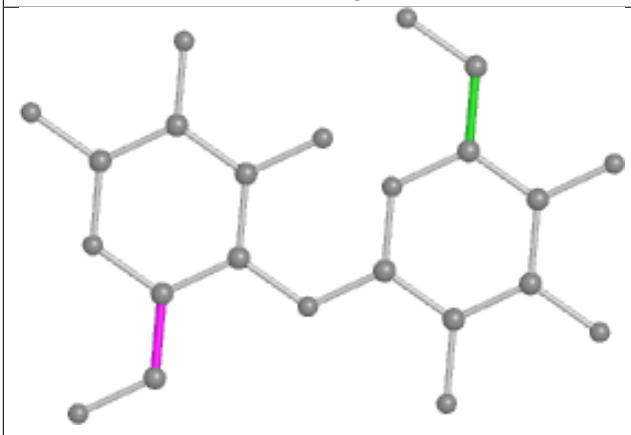
Oligosaccharide Chain D



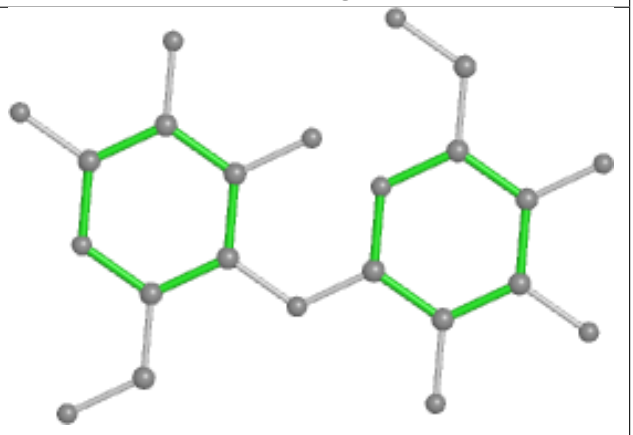
Bond lengths



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

6.1 Protein, DNA and RNA chains

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^{th} 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	A	585/618 (94%)	0.36	41 (7%)	16	15	28, 58, 103, 136	0
1	B	583/618 (94%)	1.36	172 (29%)	0	0	28, 59, 149, 201	0
All	All	1168/1236 (94%)	0.86	213 (18%)	1	1	28, 58, 141, 201	0

The worst 5 of 213 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	595	ILE	18.5
1	B	456	TYR	14.2
1	B	596	LYS	12.1
1	B	452	ARG	11.0
1	B	546	PHE	10.8

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

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

6.3 Carbohydrates

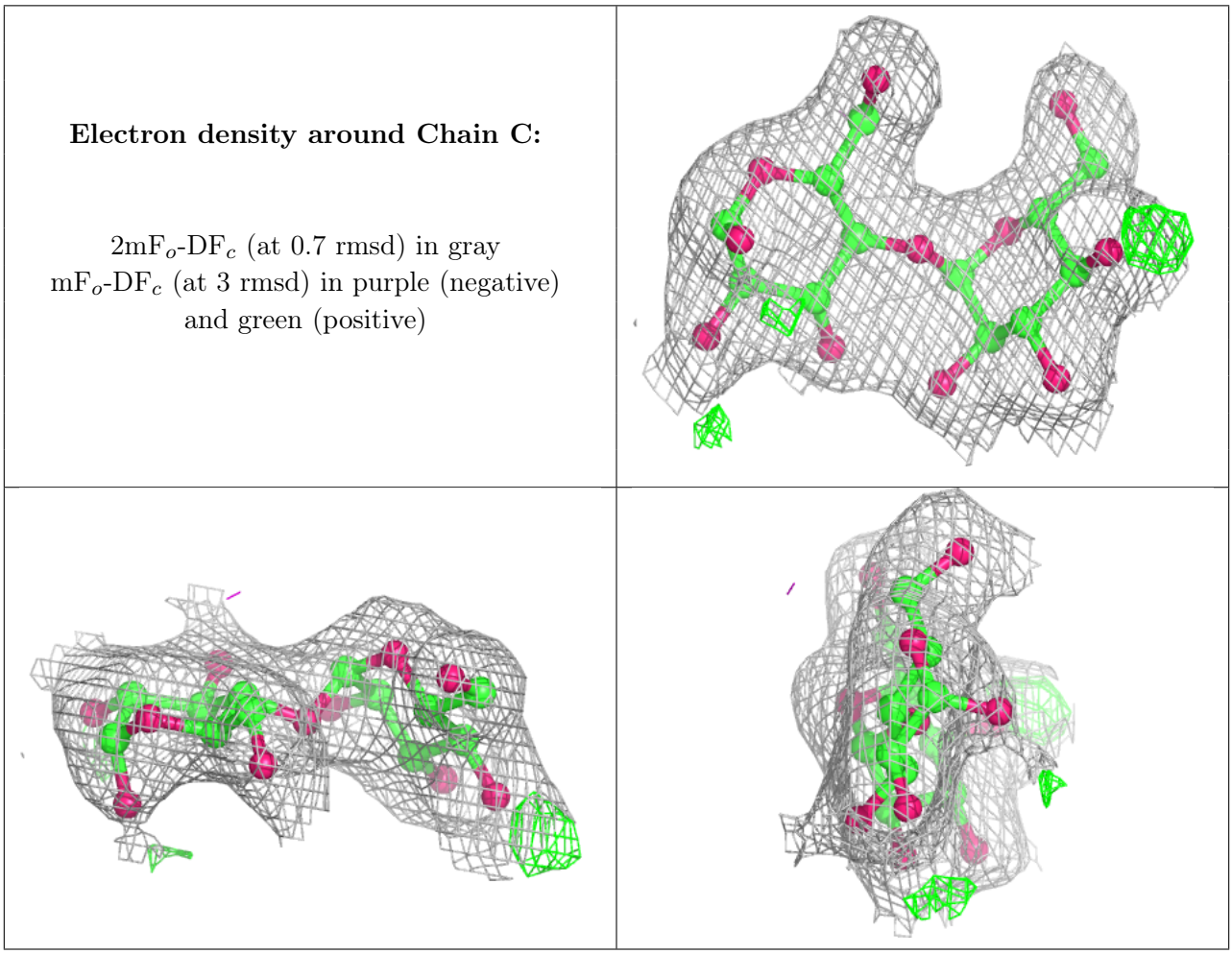
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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	GLC	D	1	12/12	0.88	0.26	32,40,46,53	0
2	GLC	C	1	12/12	0.94	0.26	28,34,39,51	0
2	GLC	C	2	11/12	0.96	0.25	27,33,36,40	0
2	GLC	D	2	11/12	0.96	0.24	32,36,41,44	0

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.

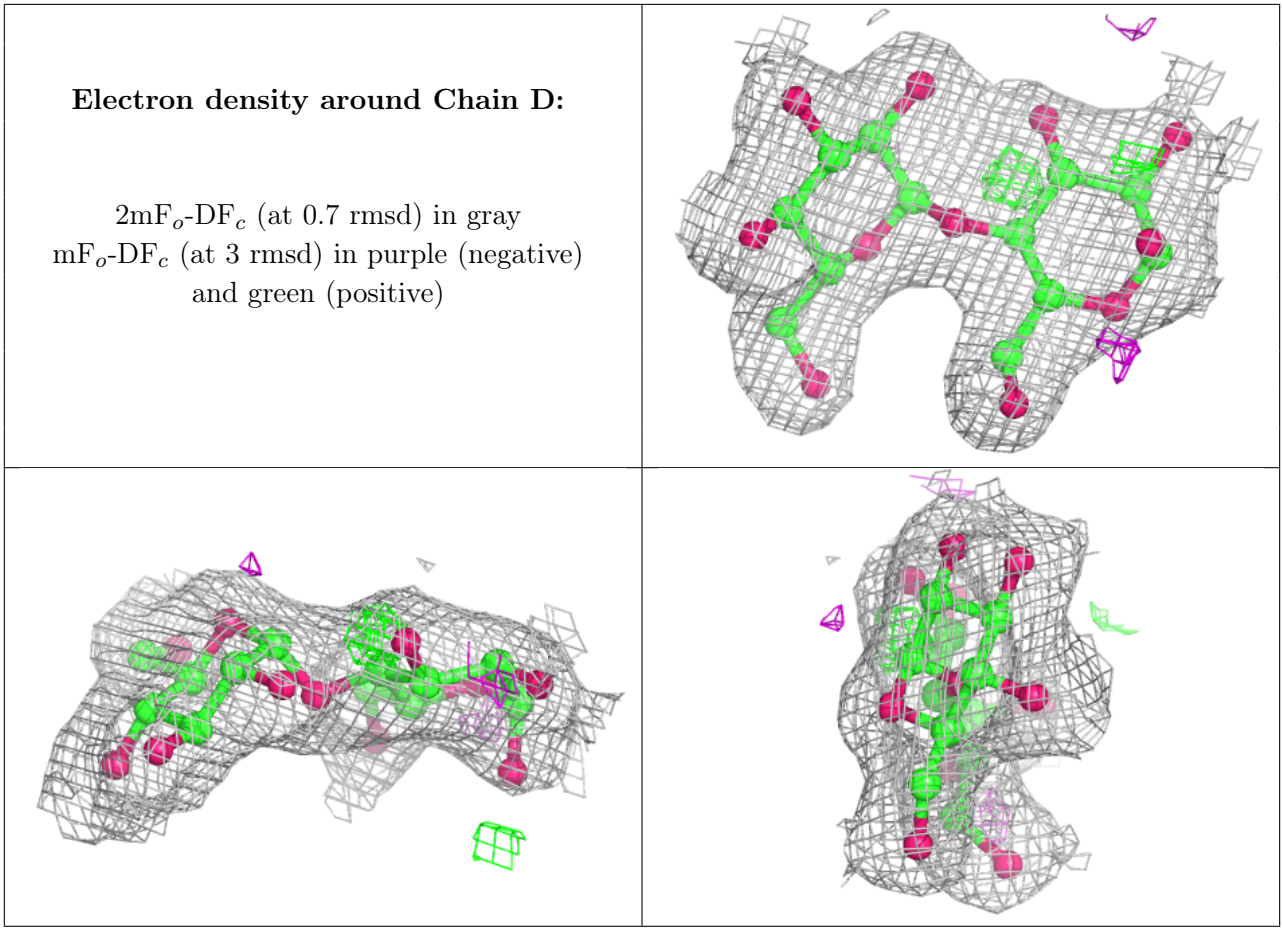
Electron density around Chain C:

$2mF_o$-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 D:

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



6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.