



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 13, 2024 – 06:34 AM EDT

PDB ID : 4HPH  
Title : The crystal structure of isomaltulose synthase mutant E295Q from *Erwinia rhapontici* NX5 in complex with its natural substrate sucrose  
Authors : Xu, Z.; Li, S.; Xu, H.; Zhou, J.  
Deposited on : 2012-10-23  
Resolution : 1.70 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 1.20.1  
EDS : 2.36.2  
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) : 2.36.2

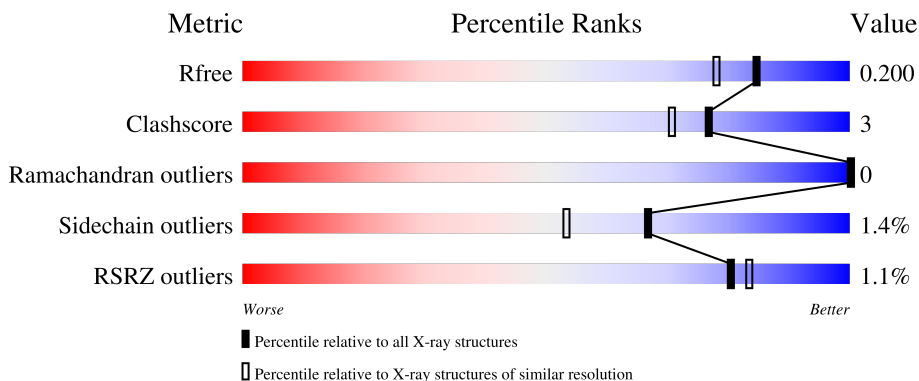
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*


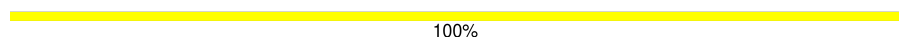
The reported resolution of this entry is 1.70 Å.

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	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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  $\geq 3$ , 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	A	559	 88% 10%
2	B	2	 100%

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 5287 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.

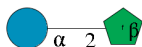
- Molecule 1 is a protein called Sucrose isomerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	559	4701	2994	817	881	9	0	7	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	295	GLN	GLU	engineered mutation	UNP D9MPF2

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



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

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Ca	0	0
			1	1		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



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

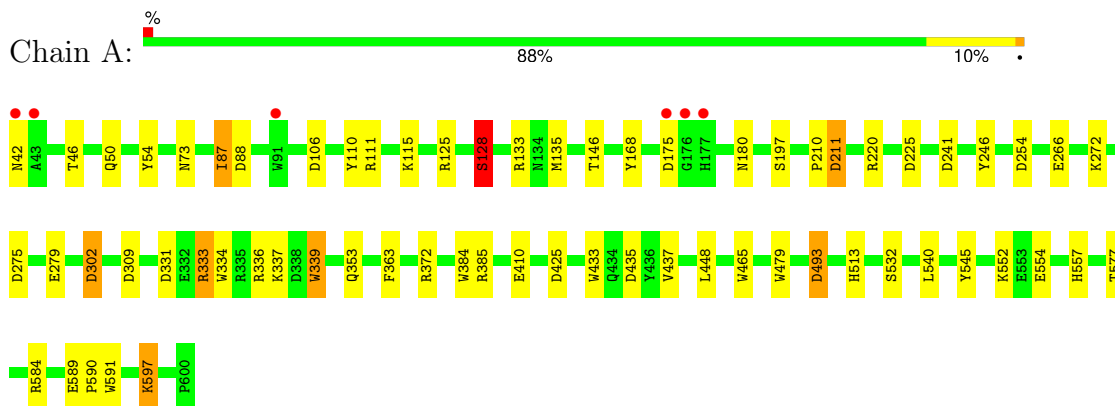
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	556	Total O 556 556	0	0

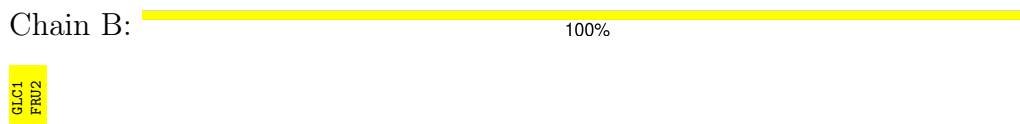
### 3 Residue-property plots [i](#)

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: Sucrose isomerase



- Molecule 2: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	58.55Å 81.09Å 138.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.98 – 1.70 34.98 – 1.70	Depositor EDS
% Data completeness (in resolution range)	99.2 (34.98-1.70) 99.2 (34.98-1.70)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.93 (at 1.70Å)	Xtrriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.161 , 0.200 0.162 , 0.200	Depositor DCC
$R_{free}$ test set	3660 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	19.7	Xtrriage
Anisotropy	0.085	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 41.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	5287	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.41% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: GOL, FRU, CA, 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	1.21	13/4852 (0.3%)	1.19	29/6578 (0.4%)

All (13) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	479	TRP	CD2-CE2	7.28	1.50	1.41
1	A	384	TRP	CD2-CE2	7.07	1.49	1.41
1	A	168	TYR	CE1-CZ	6.19	1.46	1.38
1	A	410	GLU	CD-OE2	5.91	1.32	1.25
1	A	339	TRP	CG-CD1	5.90	1.45	1.36
1	A	433	TRP	CD2-CE2	5.74	1.48	1.41
1	A	110	TYR	CG-CD2	5.67	1.46	1.39
1	A	334	TRP	CD2-CE2	5.48	1.48	1.41
1	A	339	TRP	CD2-CE2	5.38	1.47	1.41
1	A	465	TRP	CD2-CE2	5.31	1.47	1.41
1	A	128	SER	CB-OG	-5.25	1.35	1.42
1	A	591	TRP	CD2-CE2	5.16	1.47	1.41
1	A	279	GLU	CD-OE1	5.10	1.31	1.25

All (29) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	425[A]	ASP	CB-CG-OD2	8.37	125.83	118.30
1	A	425[B]	ASP	CB-CG-OD2	8.37	125.83	118.30
1	A	333	ARG	NE-CZ-NH2	7.52	124.06	120.30
1	A	241	ASP	CB-CG-OD1	-7.40	111.64	118.30
1	A	372	ARG	NE-CZ-NH2	-7.10	116.75	120.30
1	A	493	ASP	CB-CG-OD2	7.05	124.65	118.30
1	A	133	ARG	NE-CZ-NH2	-7.02	116.79	120.30
1	A	133	ARG	NE-CZ-NH1	6.99	123.80	120.30

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	333	ARG	NE-CZ-NH1	-6.80	116.90	120.30
1	A	597	LYS	CD-CE-NZ	-6.70	96.29	111.70
1	A	106	ASP	CB-CG-OD1	6.68	124.31	118.30
1	A	309	ASP	CB-CG-OD1	6.60	124.24	118.30
1	A	246	TYR	CB-CG-CD2	-6.56	117.07	121.00
1	A	309	ASP	CB-CG-OD2	-6.50	112.45	118.30
1	A	385	ARG	NE-CZ-NH2	-5.94	117.33	120.30
1	A	211	ASP	CB-CG-OD2	5.89	123.60	118.30
1	A	135	MET	CG-SD-CE	-5.89	90.77	100.20
1	A	331	ASP	CB-CG-OD1	5.89	123.60	118.30
1	A	220	ARG	NE-CZ-NH1	5.73	123.17	120.30
1	A	363	PHE	CB-CG-CD2	-5.69	116.82	120.80
1	A	448	LEU	CB-CG-CD1	-5.51	101.63	111.00
1	A	435	ASP	CB-CG-OD1	5.37	123.13	118.30
1	A	584	ARG	NE-CZ-NH1	5.31	122.95	120.30
1	A	302	ASP	CB-CG-OD2	-5.29	113.54	118.30
1	A	302	ASP	CB-CG-OD1	5.25	123.03	118.30
1	A	111	ARG	NE-CZ-NH1	-5.09	117.76	120.30
1	A	540	LEU	CB-CG-CD1	-5.05	102.42	111.00
1	A	246	TYR	CB-CG-CD1	5.05	124.03	121.00
1	A	437	VAL	CG1-CB-CG2	-5.01	102.88	110.90

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	4701	0	4488	28	0
2	B	23	0	21	0	0
3	A	1	0	0	0	0
4	A	6	0	8	1	0
5	A	556	0	0	16	0
All	All	5287	0	4517	28	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 3.

All (28) 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:266:GLU:HG2	5:A:1354:HOH:O	1.72	0.89
1:A:225:ASP:HB3	5:A:1116:HOH:O	1.71	0.88
1:A:554:GLU:HG2	5:A:1355:HOH:O	1.76	0.85
1:A:577:THR:HG23	5:A:969:HOH:O	1.86	0.75
1:A:513:HIS:HE1	5:A:1160:HOH:O	1.71	0.71
1:A:353:GLN:HG3	5:A:1111:HOH:O	1.92	0.69
1:A:557:HIS:HD2	5:A:1141:HOH:O	1.76	0.67
1:A:493:ASP:HB2	5:A:1273:HOH:O	1.95	0.67
1:A:339:TRP:O	4:A:702:GOL:H2	1.99	0.62
1:A:353:GLN:CG	5:A:1111:HOH:O	2.47	0.62
1:A:197:SER:HB3	5:A:1102:HOH:O	2.00	0.61
1:A:54:TYR:HB2	1:A:87:ILE:HG13	1.82	0.59
1:A:545:TYR:OH	1:A:597:LYS:HE2	2.03	0.58
1:A:197:SER:CB	5:A:1102:HOH:O	2.51	0.58
1:A:73:ASN:OD1	1:A:125[B]:ARG:NH2	2.38	0.57
1:A:88:ASP:OD2	5:A:1160:HOH:O	2.18	0.56
1:A:42:ASN:HB2	5:A:1112:HOH:O	2.08	0.52
1:A:333:ARG:HG3	1:A:333:ARG:HH11	1.79	0.48
1:A:557:HIS:CD2	5:A:1141:HOH:O	2.59	0.47
1:A:128:SER:OG	5:A:1337:HOH:O	2.20	0.47
1:A:545:TYR:CZ	1:A:597:LYS:HE2	2.51	0.45
1:A:532:SER:HA	1:A:552:LYS:HE3	1.99	0.44
1:A:146:THR:O	1:A:210:PRO:HD2	2.19	0.43
1:A:46:THR:O	1:A:50:GLN:HG2	2.18	0.43
1:A:589:GLU:HB3	1:A:590:PRO:HD2	2.02	0.42
1:A:180:ASN:HB2	1:A:254:ASP:OD1	2.20	0.41
1:A:337:LYS:HE3	5:A:1176:HOH:O	2.20	0.41
1:A:272:LYS:CE	1:A:275:ASP:OD2	2.69	0.41

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	564/559 (101%)	547 (97%)	17 (3%)	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	A	507/500 (101%)	500 (99%)	7 (1%)	67	53

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	87	ILE
1	A	115	LYS
1	A	128	SER
1	A	175	ASP
1	A	211	ASP
1	A	302	ASP
1	A	336	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	499	ASN
1	A	513	HIS
1	A	557	HIS
1	A	585	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](#)

2 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	B	1	2	11,11,12	1.30	2 (18%)	15,15,17	1.95	3 (20%)
2	FRU	B	2	2	11,12,12	1.08	0	10,18,18	2.63	5 (50%)

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	B	1	2	-	0/2/19/22	0/1/1/1
2	FRU	B	2	2	-	0/5/24/24	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1	GLC	O2-C2	2.42	1.48	1.43
2	B	1	GLC	O4-C4	2.01	1.47	1.43

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1	GLC	C1-C2-C3	-5.34	101.86	109.64
2	B	2	FRU	O1-C1-C2	-4.74	101.19	111.67
2	B	2	FRU	C6-C5-C4	-3.85	106.01	115.10
2	B	1	GLC	C1-O5-C5	3.61	117.02	112.19
2	B	2	FRU	O4-C4-C5	-2.95	102.61	111.08

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1	GLC	C2-C3-C4	-2.90	105.77	110.86
2	B	2	FRU	O3-C3-C4	-2.68	103.78	113.25
2	B	2	FRU	O2-C2-O5	-2.26	104.98	109.33

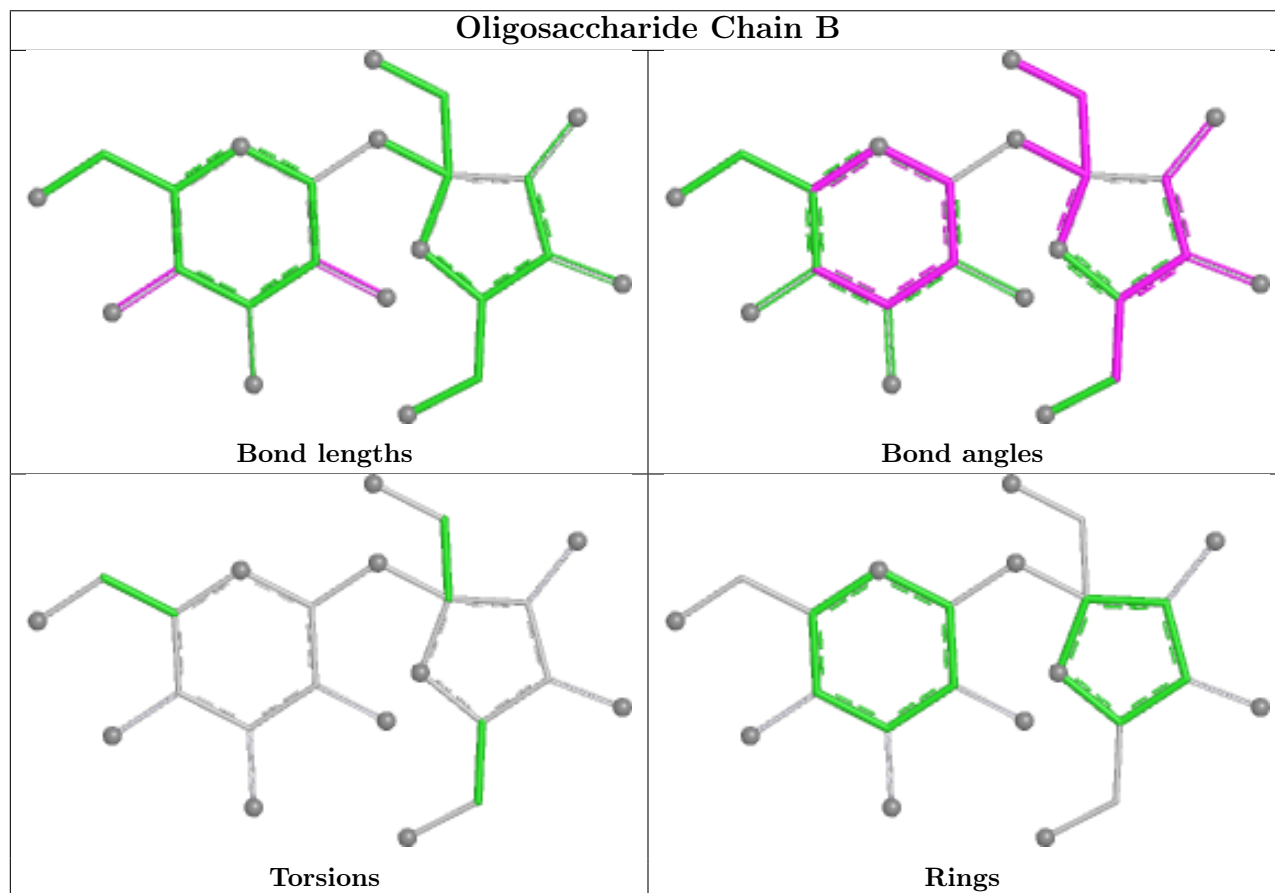
There are no chirality outliers.

There are no torsion outliers.

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.



## 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 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
4	GOL	A	702	-	5,5,5	0.55	0	5,5,5	0.97	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
4	GOL	A	702	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	702	GOL	O1-C1-C2-C3
4	A	702	GOL	O1-C1-C2-O2
4	A	702	GOL	O2-C2-C3-O3

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	702	GOL	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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	559/559 (100%)	-0.26	6 (1%) 80   83	13, 20, 36, 64	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	42	ASN	4.7
1	A	177	HIS	3.8
1	A	175	ASP	3.1
1	A	91	TRP	2.7
1	A	43	ALA	2.2
1	A	176	GLY	2.2

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

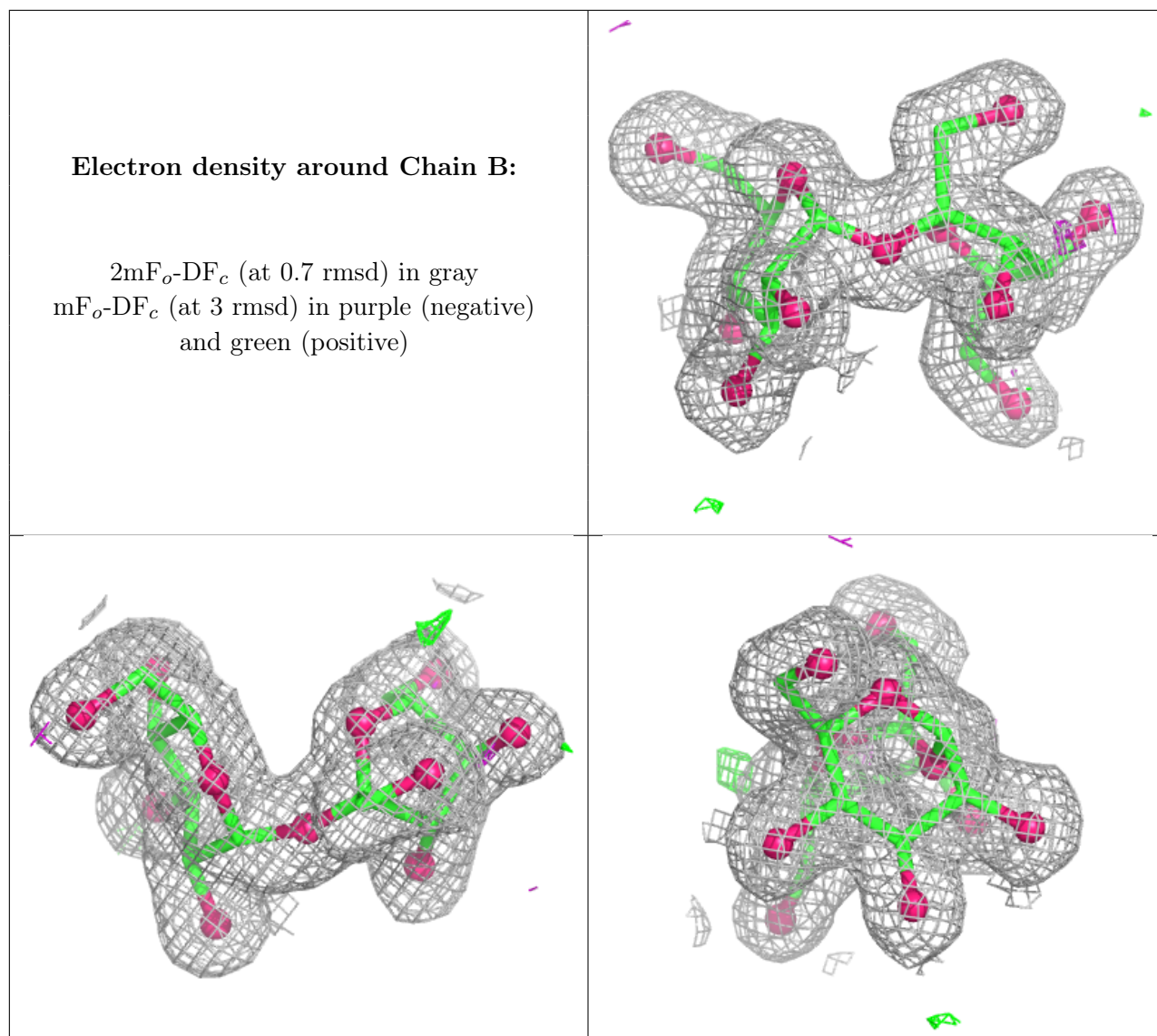
There are no non-standard protein/DNA/RNA residues in this entry.

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
2	FRU	B	2	12/12	0.95	0.12	17,20,24,25	0
2	GLC	B	1	11/12	0.99	0.17	15,16,18,19	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.



## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q < 0.9
4	GOL	A	702	6/6	0.82	0.12	37,41,49,51	0
3	CA	A	701	1/1	0.99	0.05	24,24,24,24	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.