wwPDB X-ray Structure Validation Summary Report (i)

Dec 2, 2021 – 11:10 am GMT

PDB ID	:	7A3R
Title	:	Crystal structure of dengue 1 virus envelope glycoprotein
Authors	:	Sharma, A.; Vaney, M.C.; Guardado-Calvo, P.; Duquerroy, S.; Rouvinski, A.; Rey, F.A.
Deposited on	:	2020-08-18
Resolution	:	3.60 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 following versions of software and data (see references (i)) were used in the production of this report:

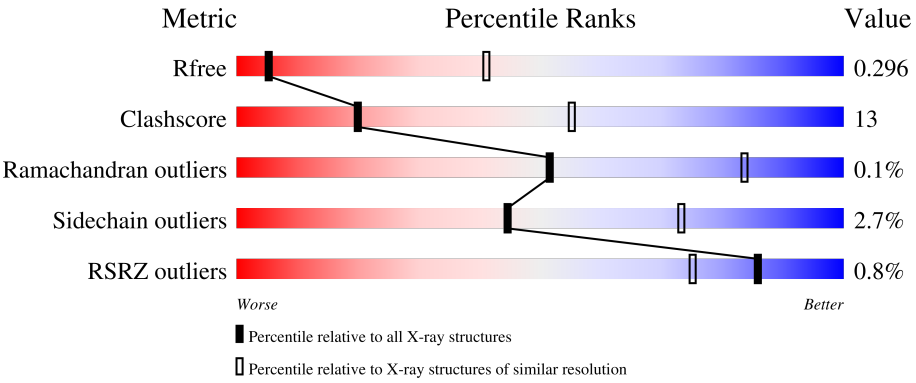
MolProbity	:	4.02b-467
Mogul	:	1.8.4 (270009), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 3.60 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	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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	433	66% 21% · 12%
1	B	433	% 59% 28% · 12%
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
3	NAG	B	501	-	-	-	X

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 5936 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 Core protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	379	Total	C	N	O	S	0	0	0
			2920	1842	495	563	20			
1	B	383	Total	C	N	O	S	0	0	0
			2964	1870	502	572	20			

There are 76 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	396	GLY	-	expression tag	UNP B1N6C5
A	397	ALA	-	expression tag	UNP B1N6C5
A	398	ALA	-	expression tag	UNP B1N6C5
A	399	ASP	-	expression tag	UNP B1N6C5
A	400	ASP	-	expression tag	UNP B1N6C5
A	401	ASP	-	expression tag	UNP B1N6C5
A	402	ASP	-	expression tag	UNP B1N6C5
A	403	LYS	-	expression tag	UNP B1N6C5
A	404	ALA	-	expression tag	UNP B1N6C5
A	405	GLY	-	expression tag	UNP B1N6C5
A	406	TRP	-	expression tag	UNP B1N6C5
A	407	SER	-	expression tag	UNP B1N6C5
A	408	HIS	-	expression tag	UNP B1N6C5
A	409	PRO	-	expression tag	UNP B1N6C5
A	410	GLN	-	expression tag	UNP B1N6C5
A	411	PHE	-	expression tag	UNP B1N6C5
A	412	GLU	-	expression tag	UNP B1N6C5
A	413	LYS	-	expression tag	UNP B1N6C5
A	414	GLY	-	expression tag	UNP B1N6C5
A	415	GLY	-	expression tag	UNP B1N6C5
A	416	GLY	-	expression tag	UNP B1N6C5
A	417	SER	-	expression tag	UNP B1N6C5
A	418	GLY	-	expression tag	UNP B1N6C5
A	419	GLY	-	expression tag	UNP B1N6C5
A	420	GLY	-	expression tag	UNP B1N6C5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	421	SER	-	expression tag	UNP B1N6C5
A	422	GLY	-	expression tag	UNP B1N6C5
A	423	GLY	-	expression tag	UNP B1N6C5
A	424	GLY	-	expression tag	UNP B1N6C5
A	425	SER	-	expression tag	UNP B1N6C5
A	426	TRP	-	expression tag	UNP B1N6C5
A	427	SER	-	expression tag	UNP B1N6C5
A	428	HIS	-	expression tag	UNP B1N6C5
A	429	PRO	-	expression tag	UNP B1N6C5
A	430	GLN	-	expression tag	UNP B1N6C5
A	431	PHE	-	expression tag	UNP B1N6C5
A	432	GLU	-	expression tag	UNP B1N6C5
A	433	LYS	-	expression tag	UNP B1N6C5
B	396	GLY	-	expression tag	UNP B1N6C5
B	397	ALA	-	expression tag	UNP B1N6C5
B	398	ALA	-	expression tag	UNP B1N6C5
B	399	ASP	-	expression tag	UNP B1N6C5
B	400	ASP	-	expression tag	UNP B1N6C5
B	401	ASP	-	expression tag	UNP B1N6C5
B	402	ASP	-	expression tag	UNP B1N6C5
B	403	LYS	-	expression tag	UNP B1N6C5
B	404	ALA	-	expression tag	UNP B1N6C5
B	405	GLY	-	expression tag	UNP B1N6C5
B	406	TRP	-	expression tag	UNP B1N6C5
B	407	SER	-	expression tag	UNP B1N6C5
B	408	HIS	-	expression tag	UNP B1N6C5
B	409	PRO	-	expression tag	UNP B1N6C5
B	410	GLN	-	expression tag	UNP B1N6C5
B	411	PHE	-	expression tag	UNP B1N6C5
B	412	GLU	-	expression tag	UNP B1N6C5
B	413	LYS	-	expression tag	UNP B1N6C5
B	414	GLY	-	expression tag	UNP B1N6C5
B	415	GLY	-	expression tag	UNP B1N6C5
B	416	GLY	-	expression tag	UNP B1N6C5
B	417	SER	-	expression tag	UNP B1N6C5
B	418	GLY	-	expression tag	UNP B1N6C5
B	419	GLY	-	expression tag	UNP B1N6C5
B	420	GLY	-	expression tag	UNP B1N6C5
B	421	SER	-	expression tag	UNP B1N6C5
B	422	GLY	-	expression tag	UNP B1N6C5
B	423	GLY	-	expression tag	UNP B1N6C5
B	424	GLY	-	expression tag	UNP B1N6C5

Continued on next page...

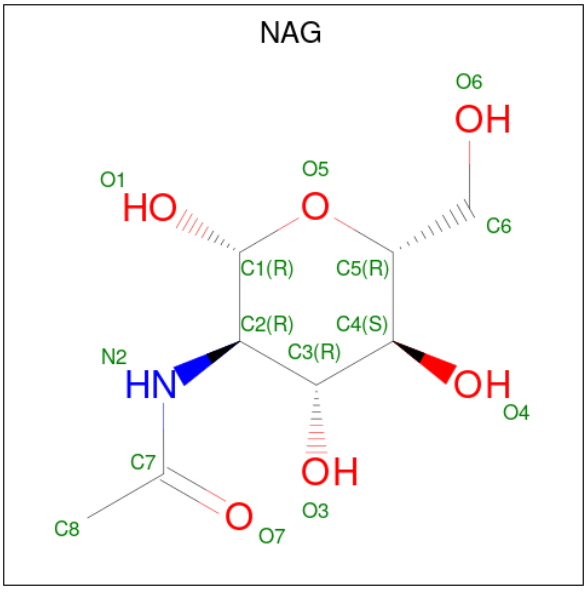
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	425	SER	-	expression tag	UNP B1N6C5
B	426	TRP	-	expression tag	UNP B1N6C5
B	427	SER	-	expression tag	UNP B1N6C5
B	428	HIS	-	expression tag	UNP B1N6C5
B	429	PRO	-	expression tag	UNP B1N6C5
B	430	GLN	-	expression tag	UNP B1N6C5
B	431	PHE	-	expression tag	UNP B1N6C5
B	432	GLU	-	expression tag	UNP B1N6C5
B	433	LYS	-	expression tag	UNP B1N6C5

- Molecule 2 is an oligosaccharide called alpha-L-fucopyranose-(1-6)-2-acetamido-2-deoxy-beta-D-glucopyranose.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	C	2	24	14	1	9	0	0	0

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



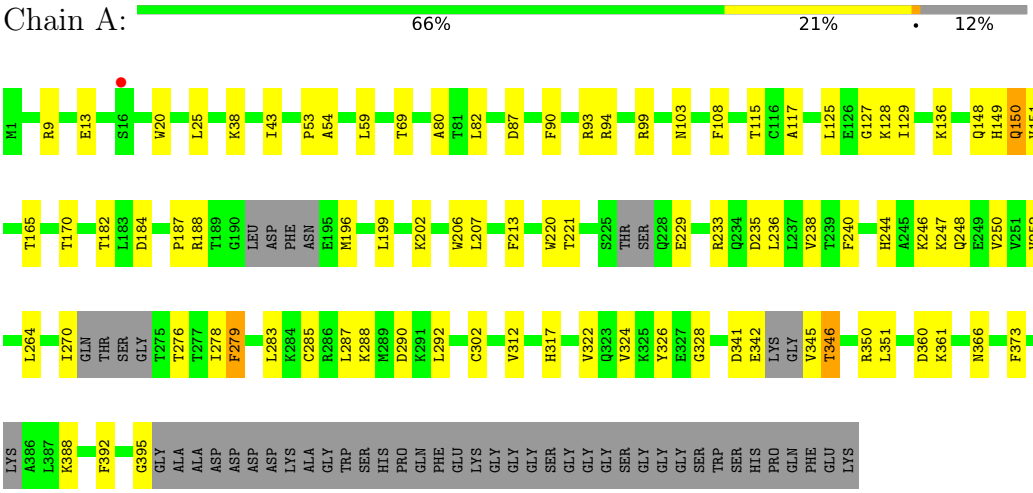
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		

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.

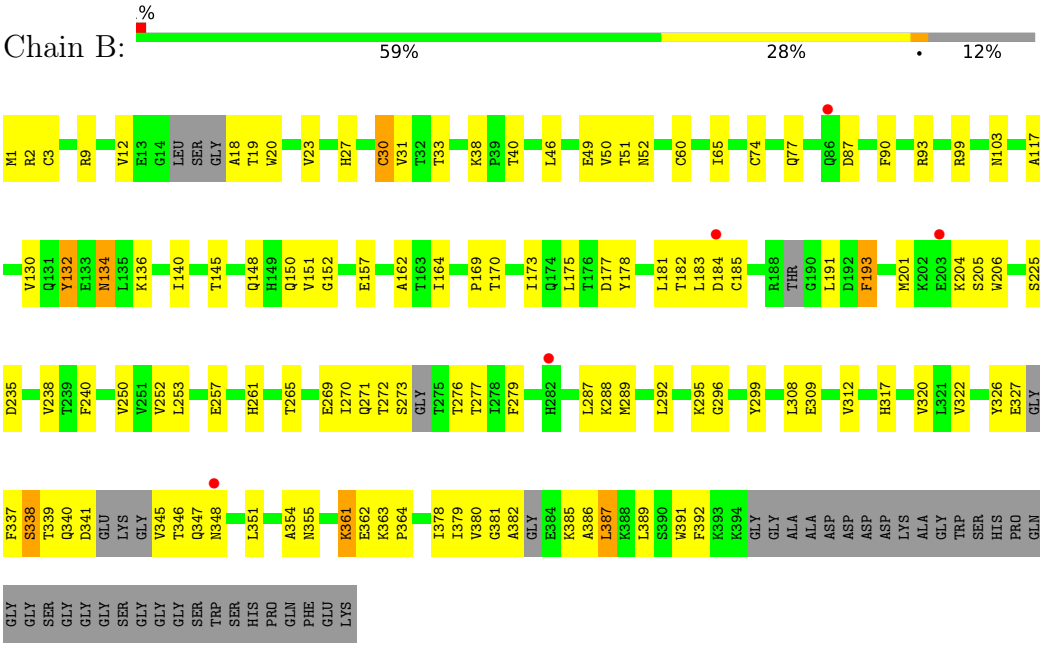
• Molecule 1: Core protein

Chain A:



• Molecule 1: Core protein

Chain B:



• Molecule 2: alpha-L-fucopyranose-(1-6)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain C:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	105.19Å 105.19Å 180.04Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.85 – 3.60 19.85 – 3.50	Depositor EDS
% Data completeness (in resolution range)	99.9 (19.85-3.60) 99.9 (19.85-3.50)	Depositor EDS
R_{merge}	0.22	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.49 (at 3.52Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.250 , 0.295 0.250 , 0.296	Depositor DCC
R_{free} test set	960 reflections (7.25%)	wwPDB-VP
Wilson B-factor (Å^2)	69.6	Xtriage
Anisotropy	0.649	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.38$, $< L^2 > = 0.20$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	5936	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.90% 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: FUC, NAG

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	0.32	0/2968	0.55	1/4018 (0.0%)
1	B	0.31	0/3012	0.55	0/4076
All	All	0.31	0/5980	0.55	1/8094 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	199	LEU	CA-CB-CG	5.56	128.09	115.30

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	2920	0	2936	64	0
1	B	2964	0	2979	86	0
2	C	24	0	22	1	0
3	A	14	0	13	0	0
3	B	14	0	13	1	0
All	All	5936	0	5963	150	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 150 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:382:ALA:H	1:B:386:ALA:HB2	1.45	0.82
1:A:43:ILE:HD11	1:A:287:LEU:HD21	1.62	0.81
1:A:278:ILE:HG22	1:A:279:PHE:H	1.50	0.76
1:A:270:ILE:HG23	1:A:276:THR:O	1.86	0.75
1:B:337:PHE:HB3	1:B:351:LEU:HD11	1.67	0.75

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	367/433 (85%)	354 (96%)	13 (4%)	0	100 100
1	B	369/433 (85%)	346 (94%)	22 (6%)	1 (0%)	41 75
All	All	736/866 (85%)	700 (95%)	35 (5%)	1 (0%)	51 83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	193	PHE

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	328/364 (90%)	321 (98%)	7 (2%)	53 78
1	B	334/364 (92%)	323 (97%)	11 (3%)	38 69
All	All	662/728 (91%)	644 (97%)	18 (3%)	44 73

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

Mol	Chain	Res	Type
1	B	348	ASN
1	B	392	PHE
1	B	387	LEU
1	B	132	TYR
1	B	338	SER

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)

2 monosaccharides are modelled in this entry.

There are no bond length outliers.

There are no bond angle outliers.

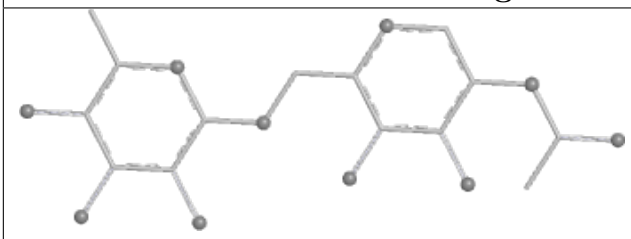
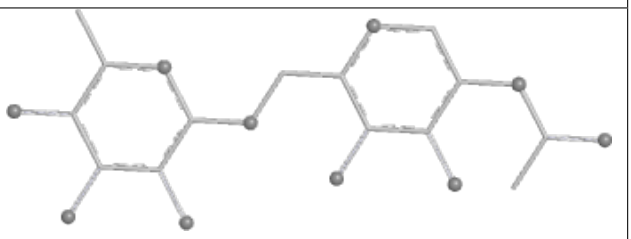
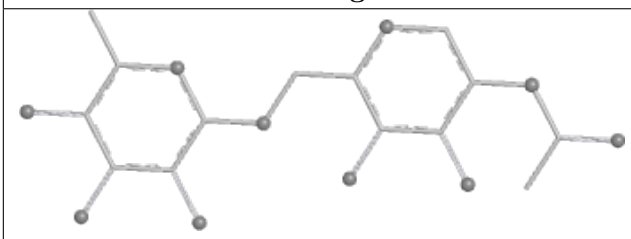
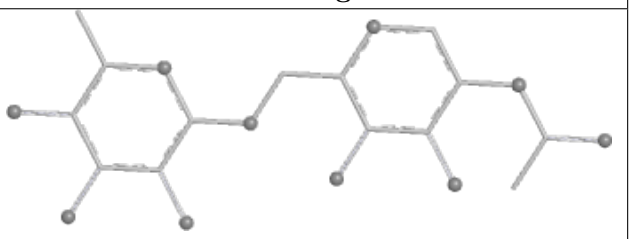
There are no chirality outliers.

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.

Oligosaccharide Chain C	
	
Bond lengths	Bond angles
	
Torsions	Rings

5.6 Ligand geometry (i)

2 ligands are modelled in this entry.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	379/433 (87%)	-0.22	1 (0%)	94	88	37, 64, 90, 106	0
1	B	383/433 (88%)	-0.07	5 (1%)	77	63	44, 78, 110, 125	0
All	All	762/866 (87%)	-0.15	6 (0%)	86	75	37, 70, 104, 125	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	348	ASN	2.9
1	B	282	HIS	2.8
1	B	86	GLN	2.6
1	B	203	GLU	2.5
1	B	184	ASP	2.3

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

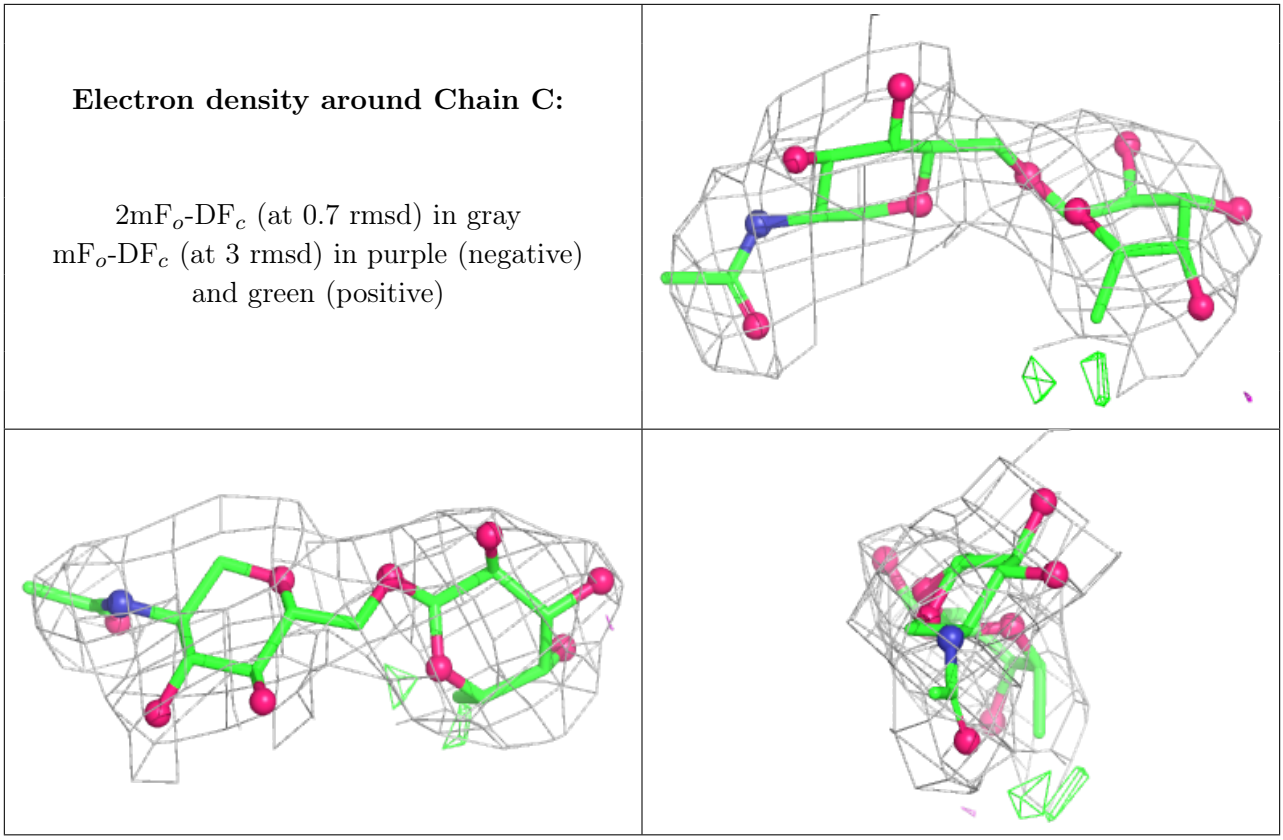
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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	NAG	C	1	14/15	0.88	0.28	61,65,70,74	0
2	FUC	C	2	10/11	0.88	0.32	71,74,76,76	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)



6.4 Ligands

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
3	NAG	B	501	14/15	0.74	0.41	117,122,124,124	0
3	NAG	A	501	14/15	0.83	0.47	100,107,110,111	0

6.5 Other polymers

There are no such residues in this entry.