



Full wwPDB X-ray Structure Validation Report ⓘ

Jul 12, 2022 – 04:17 pm BST

PDB ID : 6SJA
Title : Structure of HPV16 E6 oncoprotein in complex with IRF3 LxxLL motif
Authors : Suarez, I.P.; Cousido-Siah, A.; Bonhoure, A.; Mitschler, A.; Podjarny, A.; Trave, G.
Deposited on : 2019-08-13
Resolution : 1.50 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

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 : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.29
buster-report : 1.1.7 (2018)
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) : 2.29

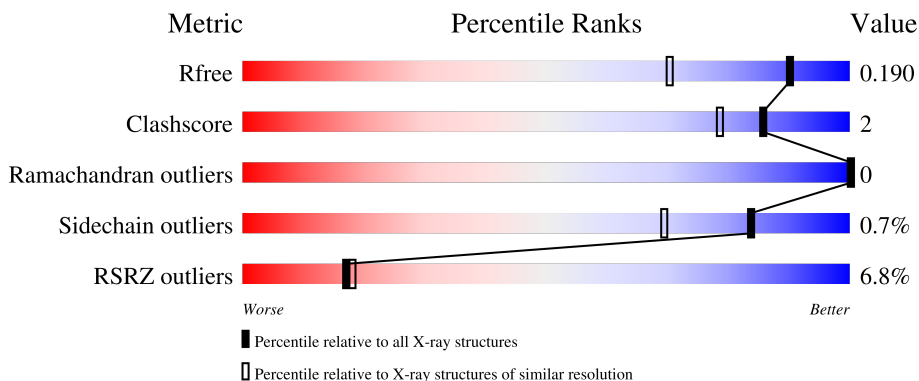
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.50 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	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)
RSRZ outliers	127900	2884 (1.50-1.50)

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 $\leq 5\%$. 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	383	
2	B	153	
3	C	5	

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 4668 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, Interferon regulatory factor 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	383	2977	1914	483	572	8	0	5	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P0AEX9
A	83	ALA	ASP	conflict	UNP P0AEX9
A	84	ALA	LYS	engineered mutation	UNP P0AEX9
A	240	ALA	LYS	engineered mutation	UNP P0AEX9
A	360	ALA	GLU	engineered mutation	UNP P0AEX9
A	363	ALA	-	linker	UNP P0AEX9
A	364	ALA	-	linker	UNP P0AEX9
A	365	ALA	-	linker	UNP P0AEX9
A	366	GLN	-	linker	UNP P0AEX9
A	367	THR	-	linker	UNP P0AEX9
A	368	ASN	-	linker	UNP P0AEX9
A	369	ALA	-	linker	UNP P0AEX9
A	370	ALA	-	linker	UNP P0AEX9
A	371	ALA	-	linker	UNP P0AEX9

- Molecule 2 is a protein called Protein E6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	134	1147	719	213	204	11	0	2	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	999	GLY	-	expression tag	UNP P03126

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1047	ARG	PHE	engineered mutation	UNP P03126
B	1080	SER	CYS	engineered mutation	UNP P03126
B	1097	SER	CYS	engineered mutation	UNP P03126
B	1111	SER	CYS	engineered mutation	UNP P03126
B	1140	SER	CYS	engineered mutation	UNP P03126

- Molecule 3 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	Trace	
3	C	5	Total	C	O	0	0	0
			56	30	26			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total	Zn	0	0
			2	2		

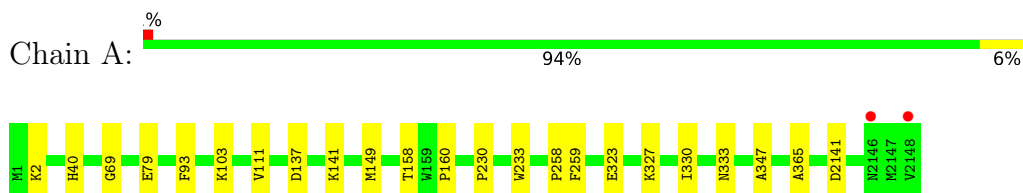
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	405	Total	O	0	4
			409	409		
5	B	76	Total	O	0	1
			77	77		

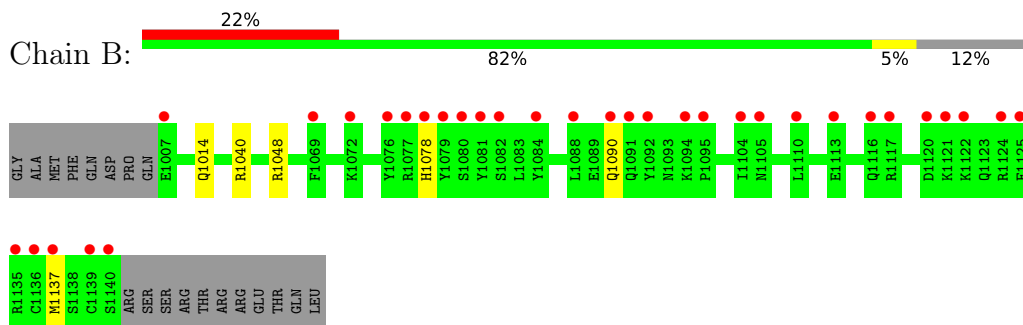
3 Residue-property plots [i](#)

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, Interferon regulatory factor 3



- Molecule 2: Protein E6



- Molecule 3: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	97.66Å 132.87Å 43.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.83 – 1.50 45.83 – 1.50	Depositor EDS
% Data completeness (in resolution range)	98.1 (45.83-1.50) 97.9 (45.83-1.50)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.01 (at 1.50Å)	Xtrriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.159 , 0.191 0.159 , 0.190	Depositor DCC
R_{free} test set	4446 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	19.0	Xtrriage
Anisotropy	0.314	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	4668	wwPDB-VP
Average B, all atoms (Å ²)	29.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.95% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, 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	0.41	0/3055	0.57	0/4148
2	B	0.31	0/1171	0.50	0/1574
All	All	0.39	0/4226	0.55	0/5722

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	2977	0	2953	14	0
2	B	1147	0	1140	5	0
3	C	56	0	48	0	0
4	B	2	0	0	0	0
5	A	409	0	0	5	0
5	B	77	0	0	3	0
All	All	4668	0	4141	18	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.

All (18) 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)
2:B:1040[A]:ARG:NH1	5:B:1305:HOH:O	2.36	0.59
1:A:158:THR:HG23	5:A:2205:HOH:O	2.05	0.56
1:A:2141:ASP:OD1	2:B:1131:ARG:NH1	2.38	0.56
2:B:1048:ARG:NH1	5:B:1307:HOH:O	2.39	0.56
2:B:1078:HIS:HB3	2:B:1129:ARG:HG3	1.87	0.56
1:A:2:LYS:NZ	5:A:2203:HOH:O	2.29	0.54
1:A:40:HIS:HD2	5:A:2508:HOH:O	1.93	0.51
1:A:137:ASP:OD2	1:A:141:LYS:HE3	2.10	0.51
1:A:323:GLU:HG2	1:A:327[B]:LYS:HE2	1.93	0.50
1:A:93:PHE:CZ	1:A:111:VAL:HG21	2.52	0.45
1:A:79:GLU:CD	1:A:103:LYS:HE3	2.38	0.44
1:A:230:PRO:HA	1:A:233:TRP:CE2	2.54	0.43
1:A:149:MET:HG2	5:A:2470[B]:HOH:O	2.19	0.42
1:A:160:PRO:HG3	1:A:258:PRO:HA	2.00	0.42
2:B:1014:GLN:NE2	5:B:1301:HOH:O	2.34	0.41
1:A:69:GLY:HA3	1:A:333:ASN:O	2.21	0.41
1:A:330:ILE:HG22	5:A:2402:HOH:O	2.20	0.40
1:A:347:ALA:HB2	1:A:365:ALA:HB2	2.04	0.40

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	386/383 (101%)	381 (99%)	5 (1%)	0	100	100
2	B	134/153 (88%)	131 (98%)	3 (2%)	0	100	100
All	All	520/536 (97%)	512 (98%)	8 (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	A	305/300 (102%)	304 (100%)	1 (0%)	92	85
2	B	130/145 (90%)	128 (98%)	2 (2%)	65	39
All	All	435/445 (98%)	432 (99%)	3 (1%)	84	69

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	259	PHE
2	B	1090	GLN
2	B	1137	MET

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	19	ASN
1	A	40	HIS
2	B	1014	GLN
2	B	1090	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](#)

5 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
3	GLC	C	1	3	12,12,12	1.05	1 (8%)	17,17,17	0.64	0
3	GLC	C	2	3	11,11,12	2.06	3 (27%)	15,15,17	1.43	2 (13%)
3	GLC	C	3	3	11,11,12	2.20	3 (27%)	15,15,17	1.40	2 (13%)
3	GLC	C	4	3	11,11,12	2.50	7 (63%)	15,15,17	1.54	2 (13%)
3	GLC	C	5	3	11,11,12	2.65	6 (54%)	15,15,17	1.10	1 (6%)

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
3	GLC	C	1	3	-	0/2/22/22	0/1/1/1
3	GLC	C	2	3	-	0/2/19/22	0/1/1/1
3	GLC	C	3	3	-	0/2/19/22	0/1/1/1
3	GLC	C	4	3	-	0/2/19/22	0/1/1/1
3	GLC	C	5	3	-	2/2/19/22	0/1/1/1

All (20) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	5	GLC	O5-C5	5.54	1.54	1.43
3	C	4	GLC	O5-C5	5.50	1.54	1.43
3	C	3	GLC	O5-C5	5.15	1.53	1.43
3	C	2	GLC	O5-C5	4.93	1.53	1.43
3	C	5	GLC	C2-C3	-3.58	1.47	1.52
3	C	4	GLC	O5-C1	2.91	1.48	1.43
3	C	5	GLC	C6-C5	-2.86	1.42	1.51
3	C	5	GLC	O5-C1	2.68	1.48	1.43
3	C	4	GLC	O2-C2	2.57	1.48	1.43
3	C	3	GLC	C6-C5	-2.54	1.43	1.51
3	C	2	GLC	O3-C3	2.39	1.48	1.43
3	C	4	GLC	C6-C5	-2.37	1.43	1.51

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	4	GLC	C1-C2	-2.34	1.46	1.52
3	C	5	GLC	O3-C3	2.33	1.48	1.43
3	C	4	GLC	C2-C3	-2.28	1.49	1.52
3	C	4	GLC	O4-C4	2.28	1.48	1.43
3	C	2	GLC	C2-C3	-2.26	1.49	1.52
3	C	5	GLC	O2-C2	2.23	1.48	1.43
3	C	3	GLC	O2-C2	2.21	1.48	1.43
3	C	1	GLC	O4-C4	2.10	1.47	1.43

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	4	GLC	C1-C2-C3	4.83	115.60	109.67
3	C	2	GLC	C1-O5-C5	3.73	117.24	112.19
3	C	3	GLC	O4-C4-C5	-2.80	102.35	109.30
3	C	5	GLC	C1-C2-C3	2.50	112.75	109.67
3	C	3	GLC	C1-O5-C5	2.33	115.34	112.19
3	C	4	GLC	O5-C5-C4	-2.17	105.55	110.83
3	C	2	GLC	O5-C5-C4	-2.03	105.88	110.83

There are no chirality outliers.

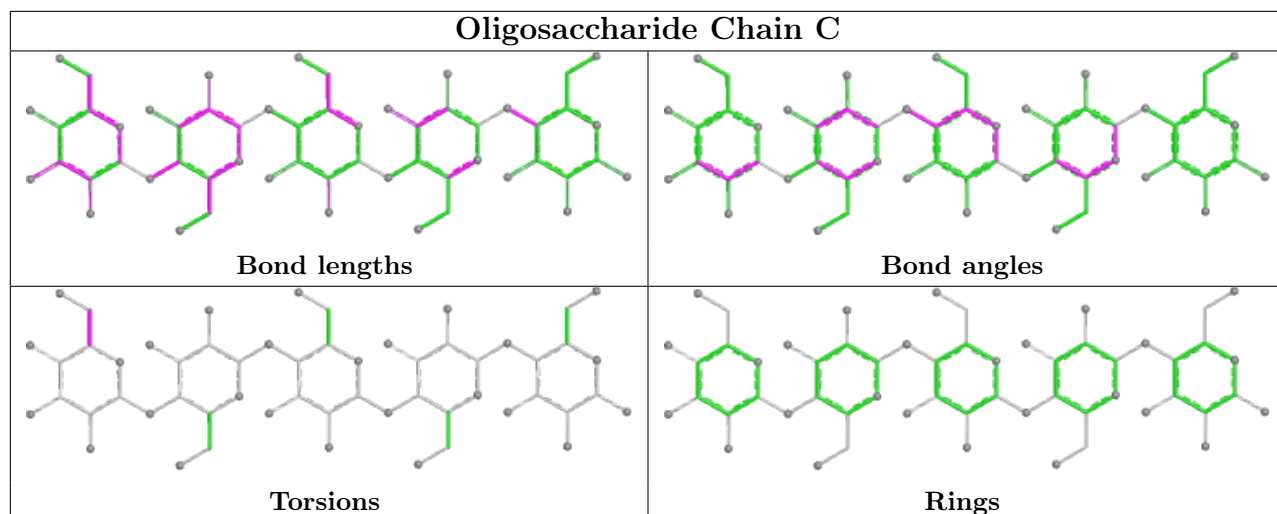
All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	5	GLC	O5-C5-C6-O6
3	C	5	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.



5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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, 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	A	383/383 (100%)	-0.11	2 (0%) 91 93	13, 21, 38, 71	1 (0%)
2	B	134/153 (87%)	1.04	33 (24%) 0 0	17, 39, 72, 79	0
All	All	517/536 (96%)	0.19	35 (6%) 17 18	13, 23, 55, 79	1 (0%)

All (35) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	2148	VAL	10.9
2	B	1076	TYR	9.5
2	B	1079	TYR	6.2
2	B	1139	CYS	5.2
2	B	1092	TYR	4.6
2	B	1136	CYS	4.4
2	B	1077	ARG	4.0
2	B	1124	ARG	3.8
2	B	1117	ARG	3.8
2	B	1084	TYR	3.7
2	B	1121	LYS	3.6
1	A	2146	ASN	3.3
2	B	1120	ASP	3.1
2	B	1137	MET	3.0
2	B	1072	LYS	3.0
2	B	1122	LYS	3.0
2	B	1078	HIS	2.8
2	B	1113	GLU	2.7
2	B	1104	ILE	2.7
2	B	1080	SER	2.7
2	B	1140	SER	2.6
2	B	1069	PHE	2.6
2	B	1088	LEU	2.5
2	B	1125	PHE	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	1091	GLN	2.5
2	B	1105	ASN	2.4
2	B	1094	LYS	2.4
2	B	1082	SER	2.4
2	B	1095	PRO	2.3
2	B	1090	GLN	2.3
2	B	1081	TYR	2.2
2	B	1116	GLN	2.2
2	B	1007	GLU	2.1
2	B	1110	LEU	2.1
2	B	1135	ARG	2.0

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

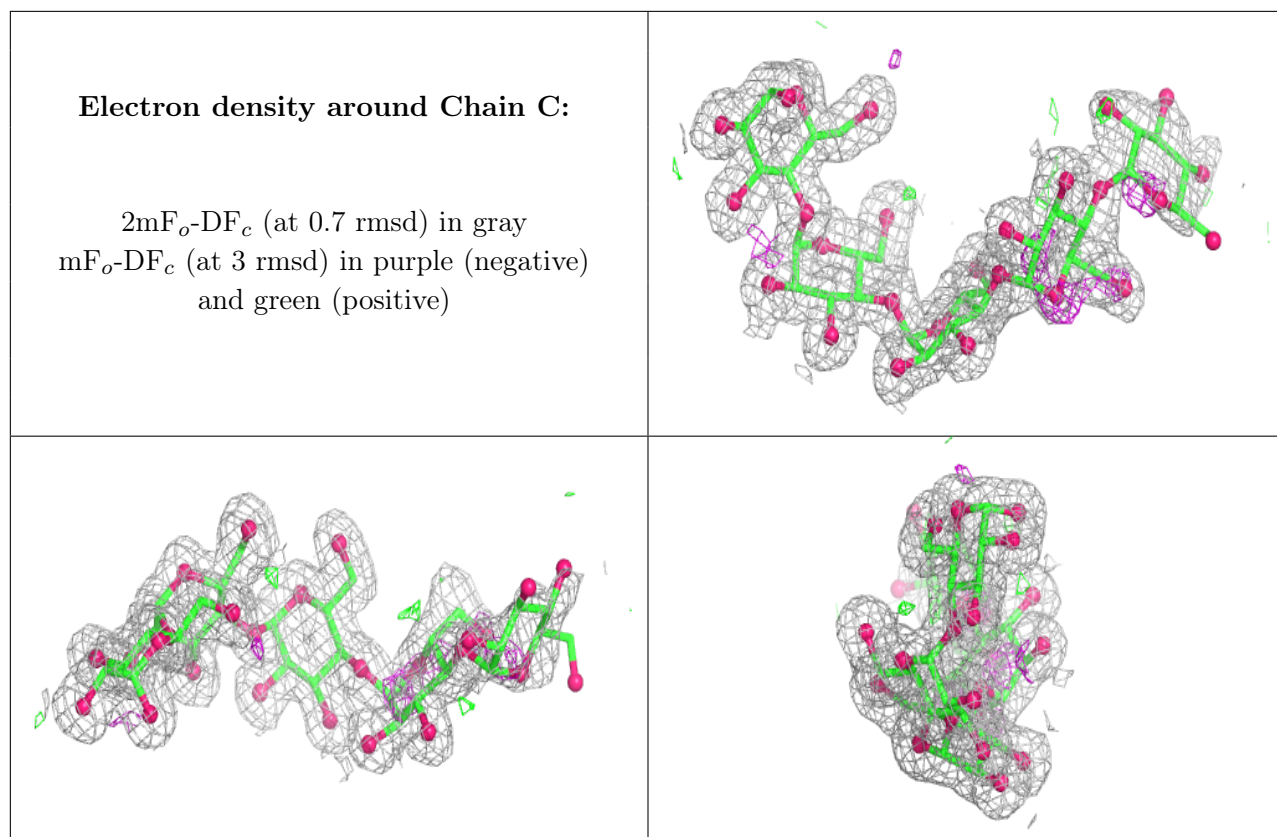
There are no non-standard protein/DNA/RNA residues in this entry.

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
3	GLC	C	5	11/12	0.75	0.36	52,66,79,79	0
3	GLC	C	4	11/12	0.80	0.17	26,30,35,40	0
3	GLC	C	2	11/12	0.97	0.10	11,13,13,13	0
3	GLC	C	3	11/12	0.97	0.07	14,16,19,20	0
3	GLC	C	1	12/12	0.98	0.09	12,14,15,16	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](#)

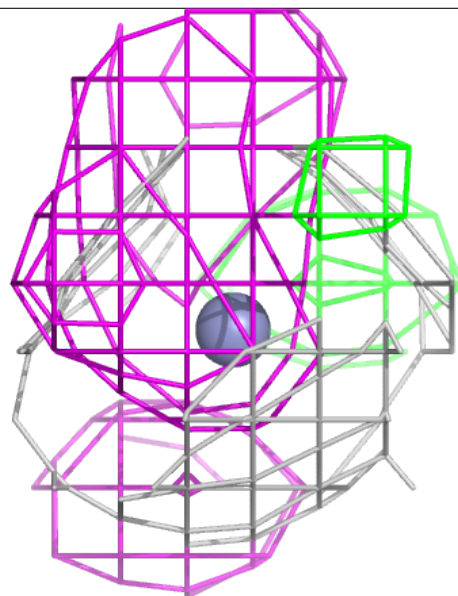
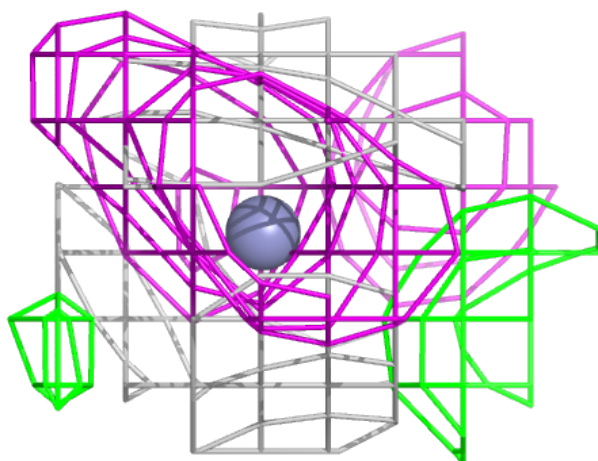
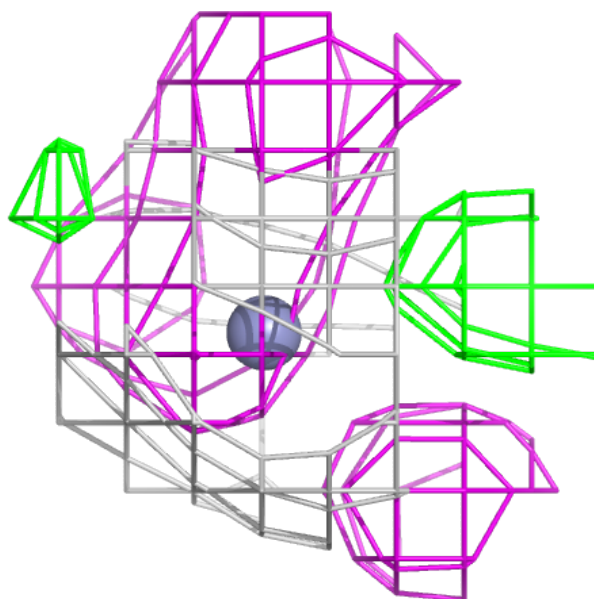
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

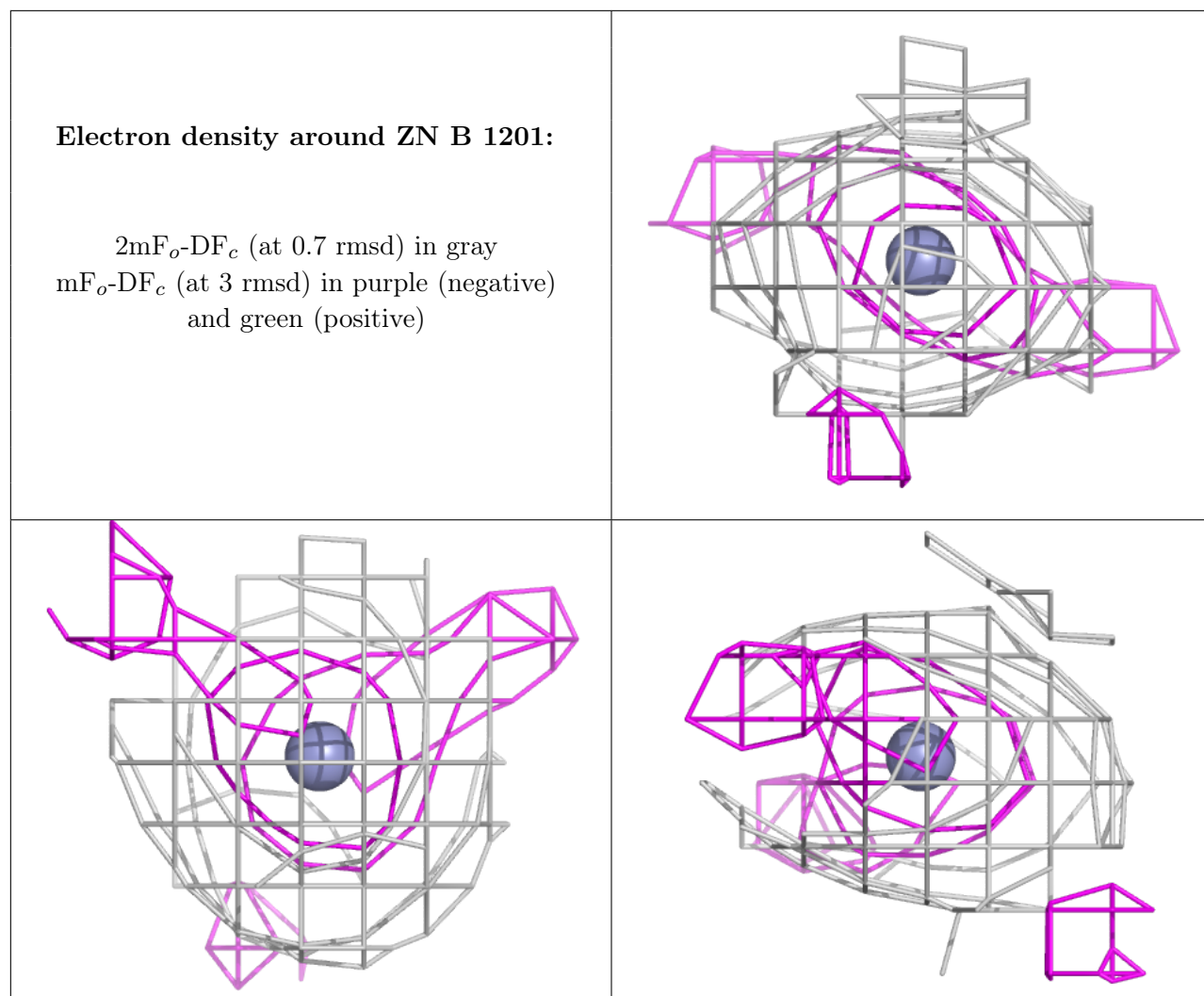
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
4	ZN	B	1200	1/1	0.98	0.05	45,45,45,45	0
4	ZN	B	1201	1/1	1.00	0.06	23,23,23,23	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around ZN B 1200:

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





6.5 Other polymers [i](#)

There are no such residues in this entry.