



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 10, 2020 – 04:25 AM BST

PDB ID : 4PQK
Title : C-Terminal domain of DNA binding protein
Authors : Schumacher, M.A.; Chinnam, N.; Tonthat, N.K.
Deposited on : 2014-03-03
Resolution : 3.40 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

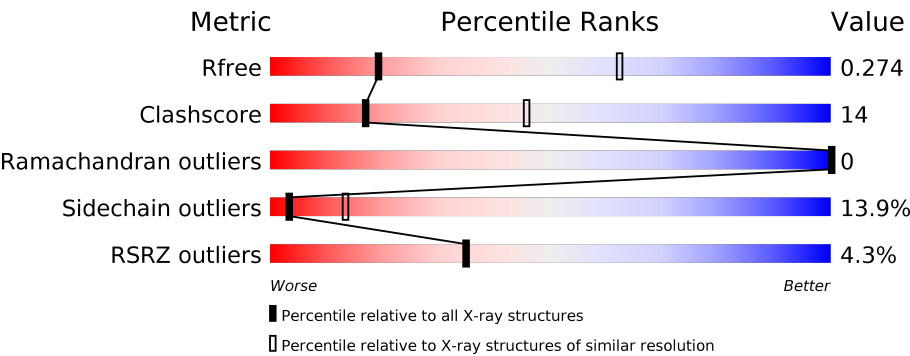
MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.13.1
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.13.1

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 3.40 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	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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 on 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	487	<div><div>3%</div><div>64%</div><div>30%</div><div>• •</div></div>
1	B	487	<div><div>4%</div><div>70%</div><div>26%</div><div>• •</div></div>
1	C	487	<div><div>5%</div><div>68%</div><div>25%</div><div>5%</div><div>•</div></div>
1	D	487	<div><div>5%</div><div>63%</div><div>28%</div><div>6%</div><div>•</div></div>
2	E	3	<div><div>33%</div><div>67%</div></div>
2	F	3	<div><div>100%</div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	3	 <div>67%33%</div>
2	H	3	 <div>100%</div>

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
2	GLC	E	3	-	-	X	-

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 14876 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 ABC transporter periplasmic protein, Truncated replication protein RepA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	478	Total	C	N	O	S	0	0	0
			3700	2371	603	713	13			
1	B	479	Total	C	N	O	S	0	0	0
			3722	2392	615	703	12			
1	C	476	Total	C	N	O	S	0	0	0
			3648	2346	601	690	11			
1	D	472	Total	C	N	O	S	0	0	0
			3670	2351	601	707	11			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP U6NJU2
B	1	MET	-	expression tag	UNP U6NJU2
C	1	MET	-	expression tag	UNP U6NJU2
D	1	MET	-	expression tag	UNP U6NJU2

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

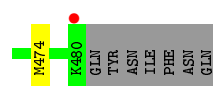


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	E	3	Total	C	O	0	0	0
			34	18	16			
2	F	3	Total	C	O	0	0	0
			34	18	16			
2	G	3	Total	C	O	0	0	0
			34	18	16			

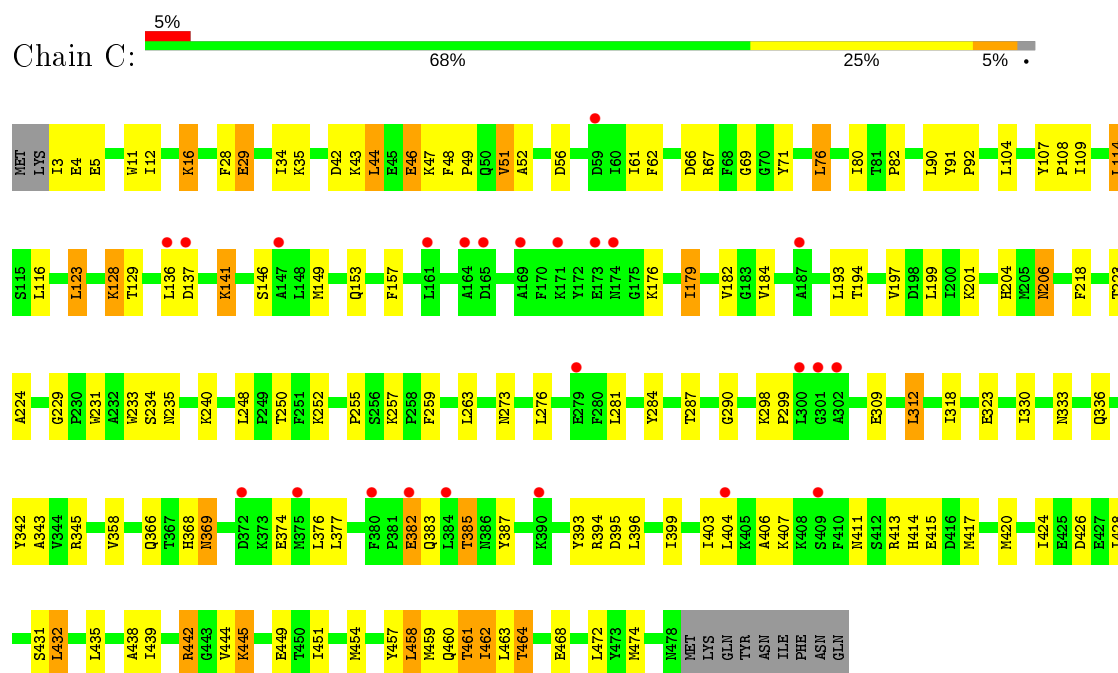
Continued on next page...

Continued from previous page...

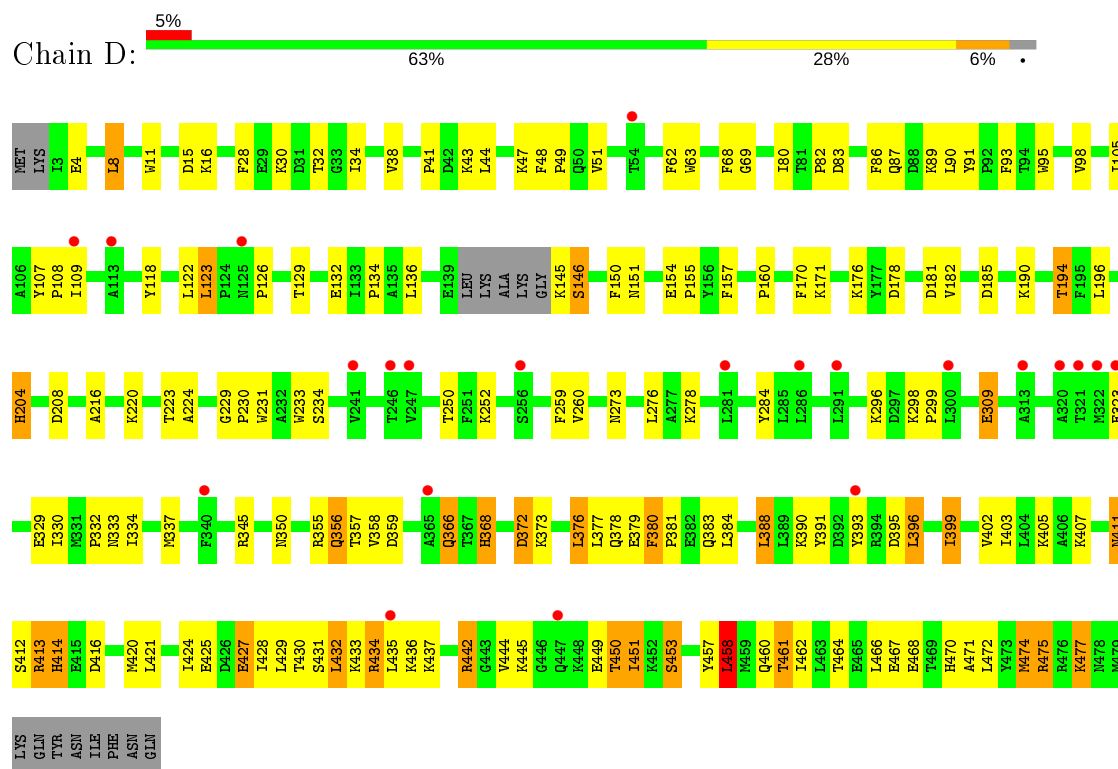
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	H	3	Total	C	O	0	0	0
			34	18	16			



- Molecule 1: Maltose ABC transporter periplasmic protein, Truncated replication protein RepA



- Molecule 1: Maltose ABC transporter periplasmic protein, Truncated replication protein RepA



- Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranos
- e

Chain E:  33% 67%

GLC1
GLC2
GLC3

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

Chain F:  100%

GLC1
GLC2
GLC3

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

Chain G:  67% 33%

GLC1
GLC2
GLC3

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

Chain H:  100%

GLC1
GLC2
GLC3

4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	64.22Å 93.67Å 100.81Å 107.67° 108.30° 84.93°	Depositor
Resolution (Å)	46.62 – 3.40 46.62 – 3.40	Depositor EDS
% Data completeness (in resolution range)	(Not available) (46.62-3.40) 67.2 (46.62-3.40)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.43 (at 3.40Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.237 , 0.275 0.241 , 0.274	Depositor DCC
R_{free} test set	1963 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å ²)	73.9	Xtriage
Anisotropy	0.378	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 25.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.85	EDS
Total number of atoms	14876	wwPDB-VP
Average B, all atoms (Å ²)	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.50% 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 [i](#)

5.1 Standard geometry [i](#)

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	0.45	0/3784	0.57	0/5142
1	B	0.44	0/3805	0.54	0/5157
1	C	0.43	0/3730	0.63	1/5063 (0.0%)
1	D	0.45	0/3751	0.65	2/5090 (0.0%)
All	All	0.44	0/15070	0.60	3/20452 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	8	LEU	CA-CB-CG	5.36	127.62	115.30
1	D	458	LEU	CA-CB-CG	5.14	127.12	115.30
1	C	463	LEU	CA-CB-CG	5.02	126.84	115.30

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	3700	0	3562	116	0
1	B	3722	0	3663	87	0
1	C	3648	0	3540	85	0
1	D	3670	0	3560	108	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	34	0	29	7	0
2	F	34	0	30	7	0
2	G	34	0	29	5	0
2	H	34	0	30	6	0
All	All	14876	0	14443	411	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 14.

The worst 5 of 411 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:G:3:GLC:O5	2:G:3:GLC:C5	1.63	1.46
2:E:3:GLC:O5	2:E:3:GLC:C5	1.64	1.45
2:H:3:GLC:O5	2:H:3:GLC:C5	1.64	1.43
2:F:3:GLC:C5	2:F:3:GLC:O5	1.64	1.41
2:F:3:GLC:C1	2:F:3:GLC:O5	1.72	1.38

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	476/487 (98%)	463 (97%)	13 (3%)	0	100	100
1	B	477/487 (98%)	472 (99%)	5 (1%)	0	100	100
1	C	474/487 (97%)	469 (99%)	5 (1%)	0	100	100
1	D	468/487 (96%)	464 (99%)	4 (1%)	0	100	100
All	All	1895/1948 (97%)	1868 (99%)	27 (1%)	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	375/407 (92%)	326 (87%)	49 (13%)	4	15
1	B	380/407 (93%)	337 (89%)	43 (11%)	6	21
1	C	363/407 (89%)	308 (85%)	55 (15%)	3	11
1	D	375/407 (92%)	315 (84%)	60 (16%)	2	10
All	All	1493/1628 (92%)	1286 (86%)	207 (14%)	3	13

5 of 207 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	4	GLU
1	C	201	LYS
1	D	434	ARG
1	C	42	ASP
1	C	128	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	324	ASN
1	B	336	GLN
1	B	378	GLN
1	C	369	ASN
1	D	470	HIS

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 ⓘ

12 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	E	1	2	12,12,12	1.27	1 (8%)	17,17,17	1.61	4 (23%)
2	GLC	E	2	2	11,11,12	2.03	4 (36%)	15,15,17	1.72	5 (33%)
2	GLC	E	3	2	11,11,12	7.45	7 (63%)	15,15,17	4.87	8 (53%)
2	GLC	F	1	2	12,12,12	1.22	0	17,17,17	1.31	3 (17%)
2	GLC	F	2	2	11,11,12	1.70	3 (27%)	15,15,17	1.23	1 (6%)
2	GLC	F	3	2	11,11,12	7.22	8 (72%)	15,15,17	3.45	5 (33%)
2	GLC	G	1	2	12,12,12	1.21	2 (16%)	17,17,17	1.80	5 (29%)
2	GLC	G	2	2	11,11,12	1.72	3 (27%)	15,15,17	1.71	4 (26%)
2	GLC	G	3	2	11,11,12	7.07	7 (63%)	15,15,17	3.77	4 (26%)
2	GLC	H	1	2	12,12,12	1.25	1 (8%)	17,17,17	1.41	2 (11%)
2	GLC	H	2	2	11,11,12	1.77	3 (27%)	15,15,17	1.12	2 (13%)
2	GLC	H	3	2	11,11,12	7.14	8 (72%)	15,15,17	2.89	5 (33%)

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	E	1	2	-	1/2/22/22	0/1/1/1
2	GLC	E	2	2	-	0/2/19/22	0/1/1/1
2	GLC	E	3	2	-	2/2/19/22	0/1/1/1
2	GLC	F	1	2	-	2/2/22/22	0/1/1/1
2	GLC	F	2	2	-	0/2/19/22	0/1/1/1
2	GLC	F	3	2	-	2/2/19/22	0/1/1/1
2	GLC	G	1	2	-	0/2/22/22	0/1/1/1
2	GLC	G	2	2	-	1/2/19/22	0/1/1/1
2	GLC	G	3	2	-	2/2/19/22	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GLC	H	1	2	-	2/2/22/22	0/1/1/1
2	GLC	H	2	2	-	0/2/19/22	0/1/1/1
2	GLC	H	3	2	-	2/2/19/22	0/1/1/1

The worst 5 of 47 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	3	GLC	O5-C1	18.58	1.73	1.43
2	F	3	GLC	O5-C1	17.70	1.72	1.43
2	G	3	GLC	O5-C1	17.43	1.71	1.43
2	H	3	GLC	O5-C1	17.42	1.71	1.43
2	H	3	GLC	O5-C5	10.51	1.64	1.43

The worst 5 of 48 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	3	GLC	C1-O5-C5	-11.46	96.66	112.19
2	E	3	GLC	C2-C3-C4	10.20	128.55	110.89
2	G	3	GLC	C1-O5-C5	-9.80	98.92	112.19
2	F	3	GLC	C1-O5-C5	-9.59	99.20	112.19
2	G	3	GLC	C2-C3-C4	8.63	125.83	110.89

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	F	3	GLC	O5-C5-C6-O6
2	G	3	GLC	O5-C5-C6-O6
2	H	3	GLC	O5-C5-C6-O6
2	E	3	GLC	O5-C5-C6-O6
2	F	3	GLC	C4-C5-C6-O6

There are no ring outliers.

9 monomers are involved in 25 short contacts:

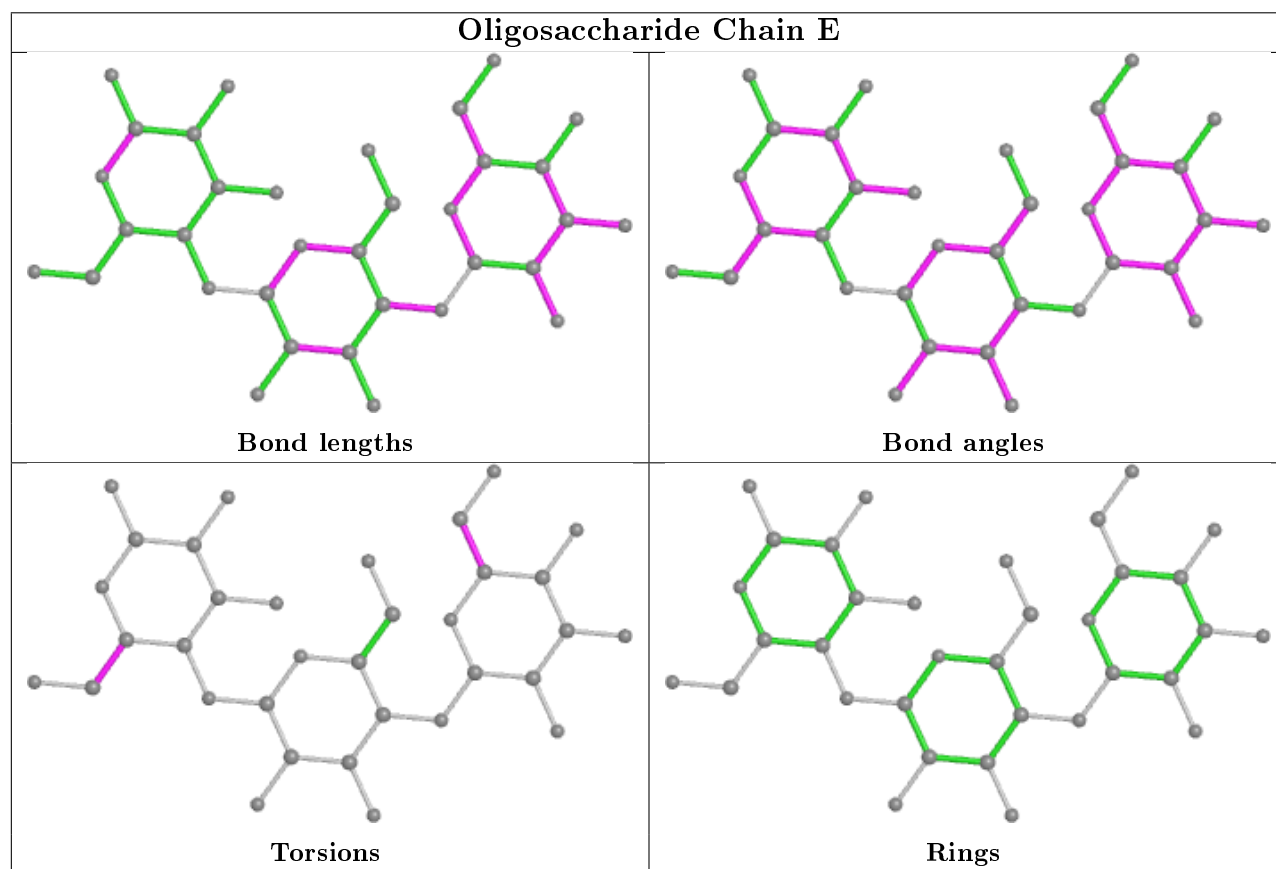
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	H	1	GLC	1	0
2	F	3	GLC	5	0
2	H	3	GLC	4	0
2	E	3	GLC	6	0

Continued on next page...

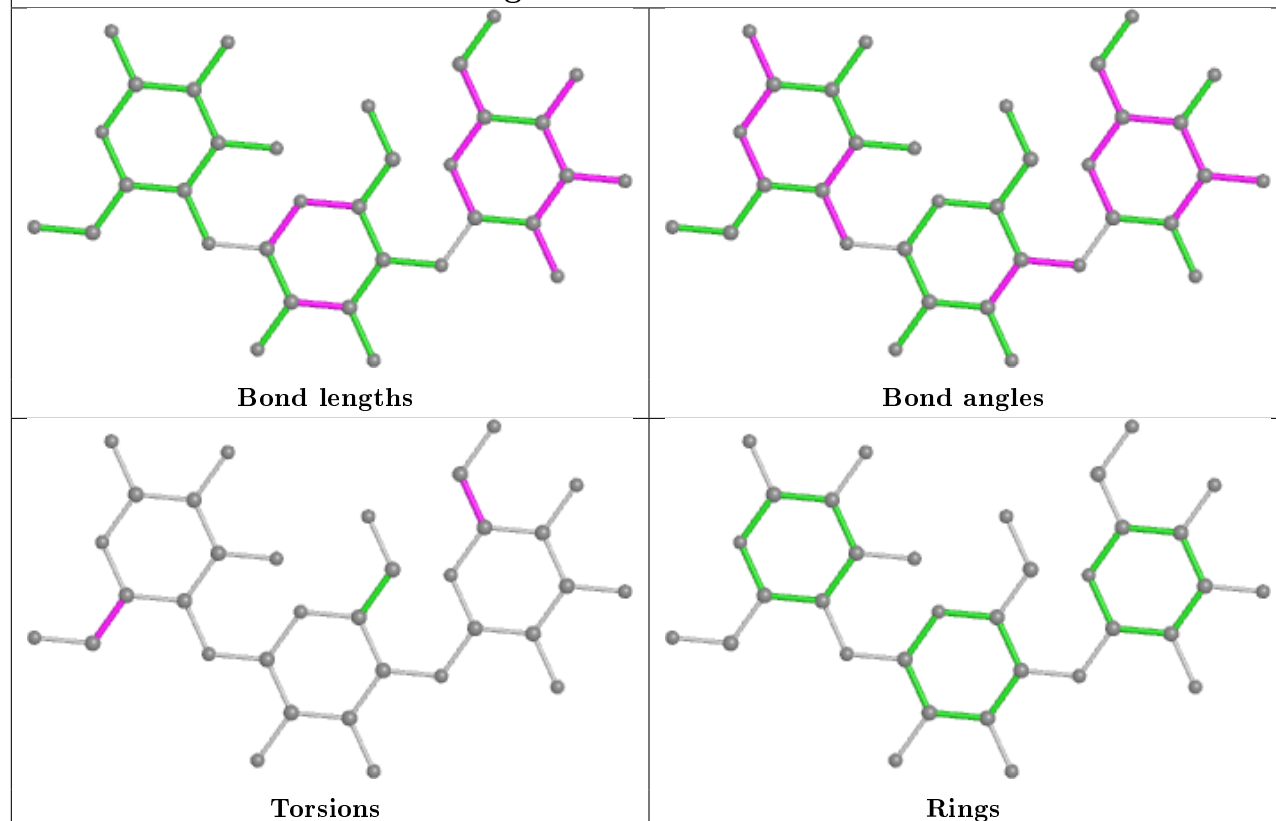
Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	2	GLC	2	0
2	F	2	GLC	1	0
2	F	1	GLC	1	0
2	G	3	GLC	5	0
2	H	2	GLC	1	0

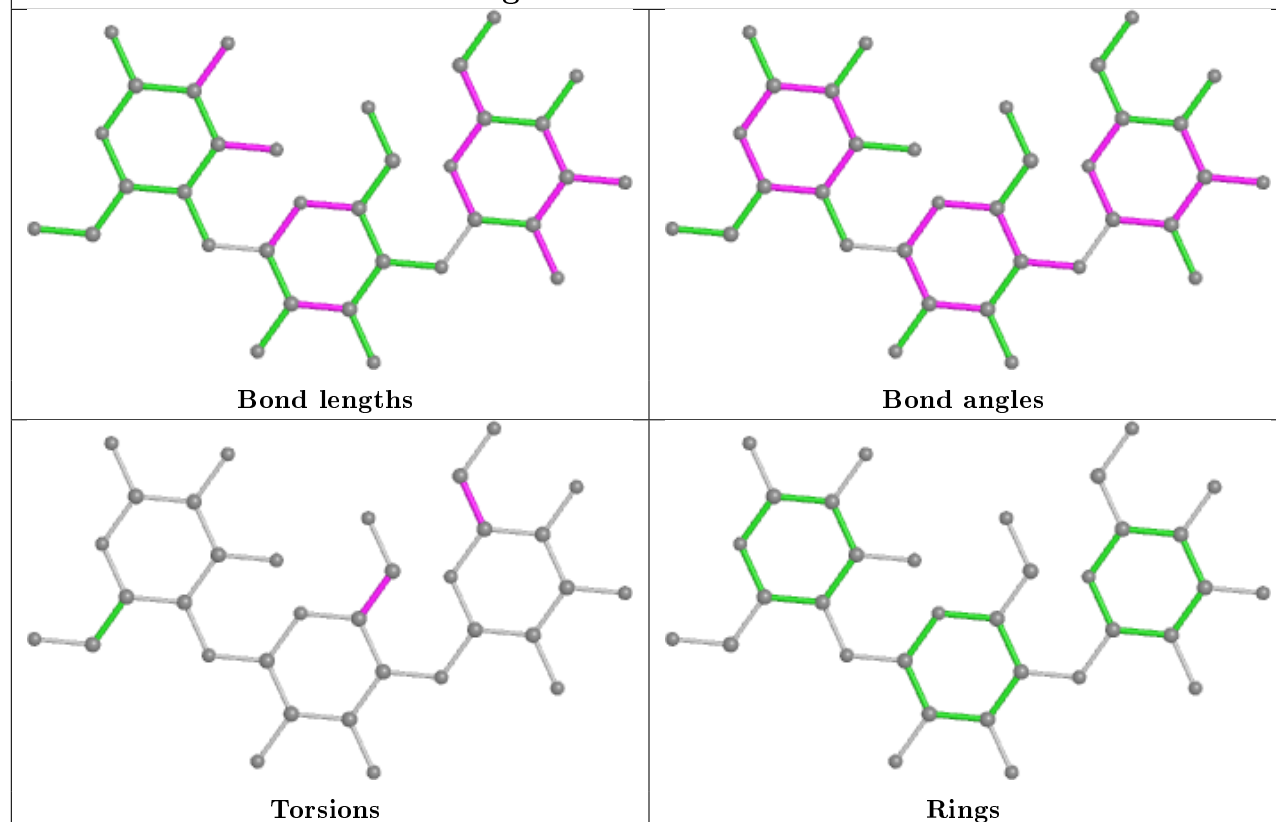
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

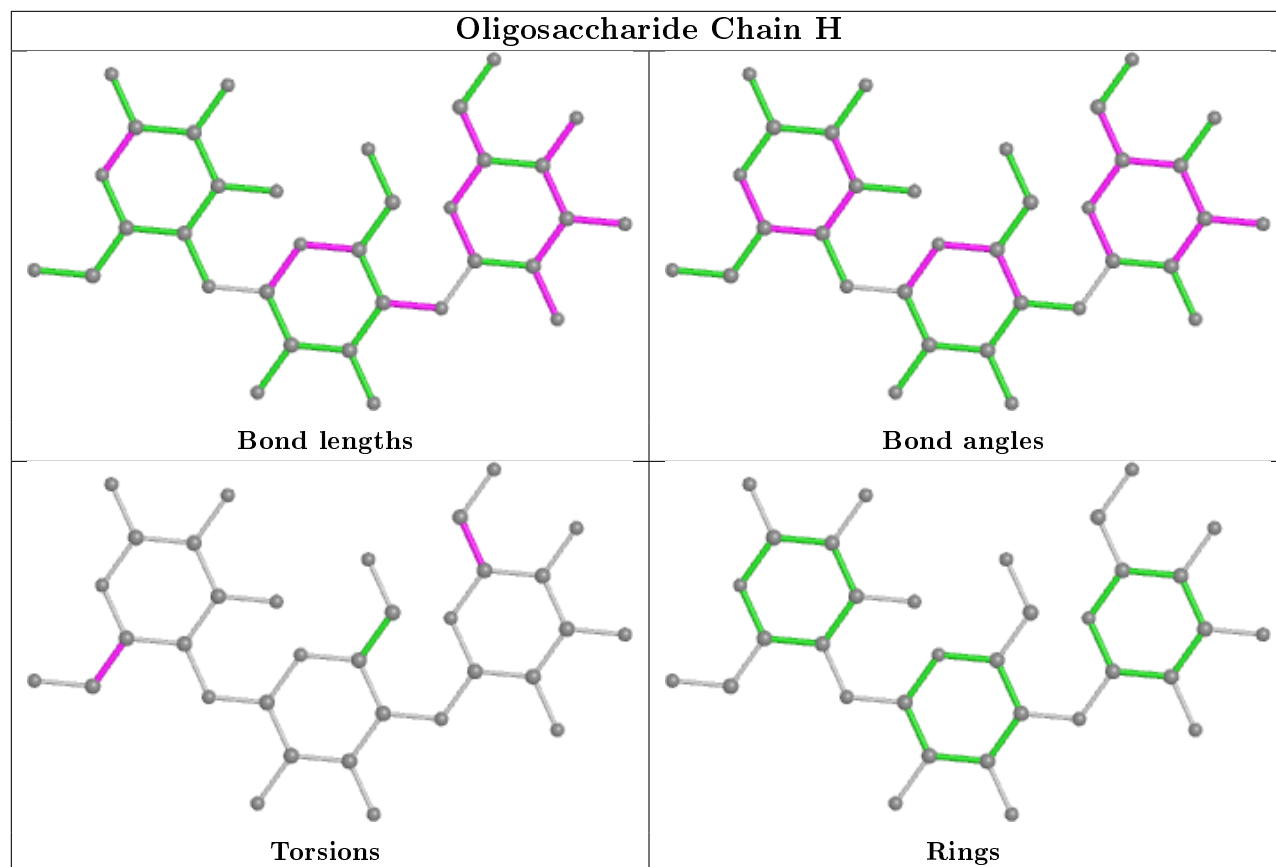


Oligosaccharide Chain F



Oligosaccharide Chain G





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](#)

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	478/487 (98%)	0.08	17 (3%) 42 42	31, 61, 95, 121	0
1	B	479/487 (98%)	0.10	19 (3%) 38 37	39, 66, 99, 135	0
1	C	476/487 (97%)	0.06	24 (5%) 28 29	41, 68, 105, 140	0
1	D	472/487 (96%)	0.15	22 (4%) 31 31	45, 73, 100, 138	0
All	All	1905/1948 (97%)	0.10	82 (4%) 35 35	31, 67, 101, 140	0

The worst 5 of 82 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	125	ASN	4.9
1	C	165	ASP	4.6
1	C	375	MET	4.5
1	C	372	ASP	4.4
1	C	173	GLU	4.3

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.

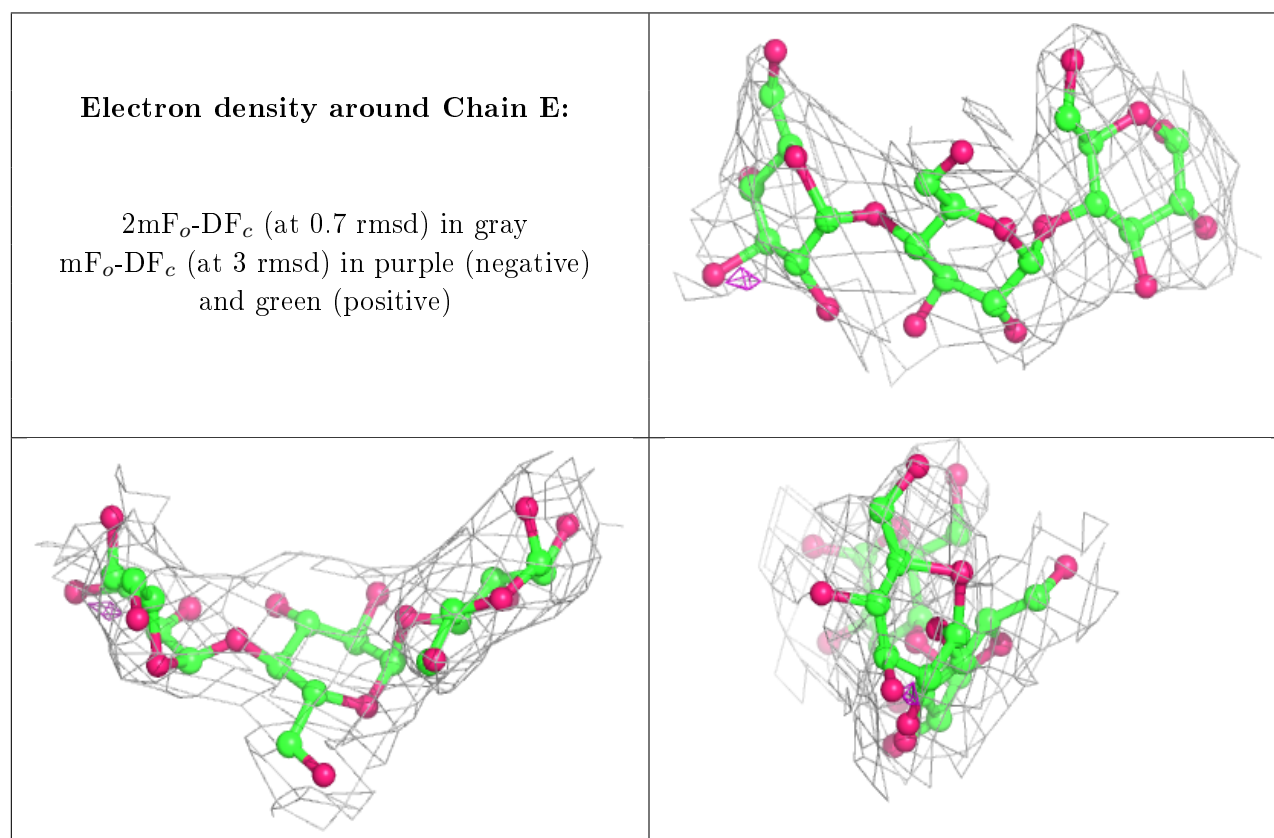
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	GLC	H	3	11/12	0.79	0.35	45,65,74,74	0
2	GLC	H	1	12/12	0.85	0.27	49,60,68,72	0

Continued on next page...

Continued from previous page...

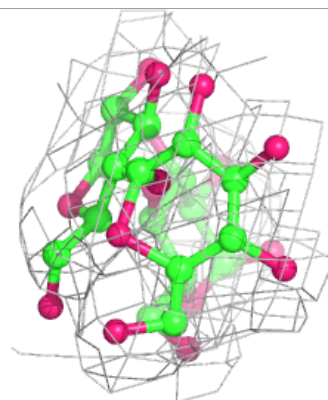
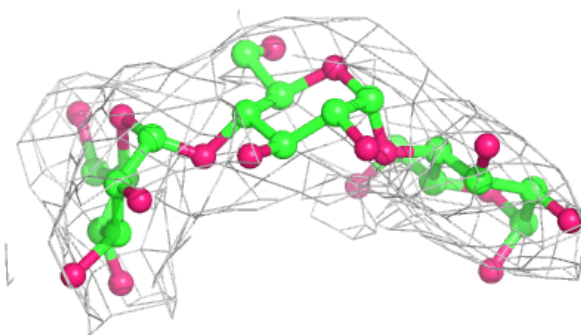
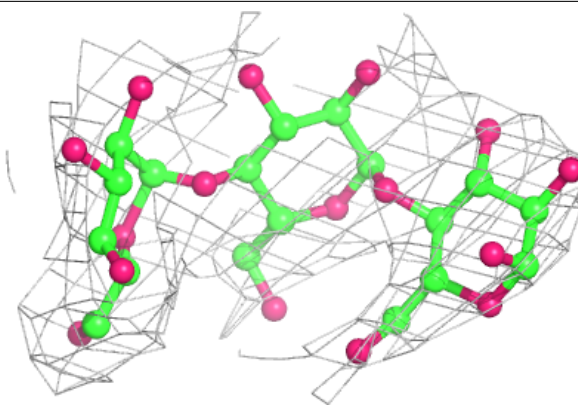
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	GLC	F	1	12/12	0.88	0.35	47,52,59,61	0
2	GLC	E	3	11/12	0.89	0.45	37,55,66,76	0
2	GLC	H	2	11/12	0.89	0.17	47,55,59,66	0
2	GLC	F	3	11/12	0.90	0.16	52,61,67,76	0
2	GLC	F	2	11/12	0.91	0.23	33,47,57,60	0
2	GLC	G	3	11/12	0.93	0.16	61,67,79,81	0
2	GLC	G	1	12/12	0.93	0.28	57,64,72,78	0
2	GLC	E	1	12/12	0.94	0.19	36,44,53,54	0
2	GLC	E	2	11/12	0.94	0.14	30,33,36,37	0
2	GLC	G	2	11/12	0.95	0.12	53,60,69,70	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.

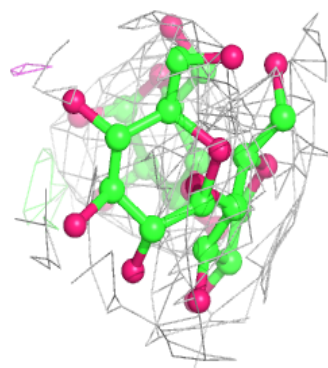
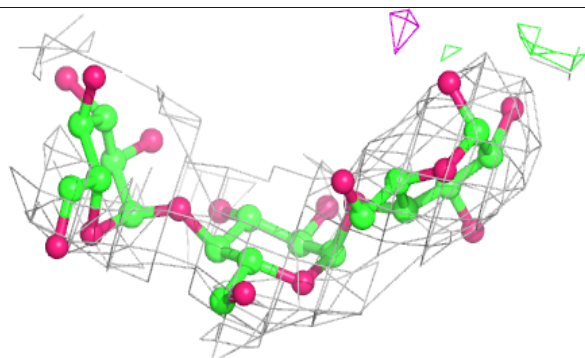
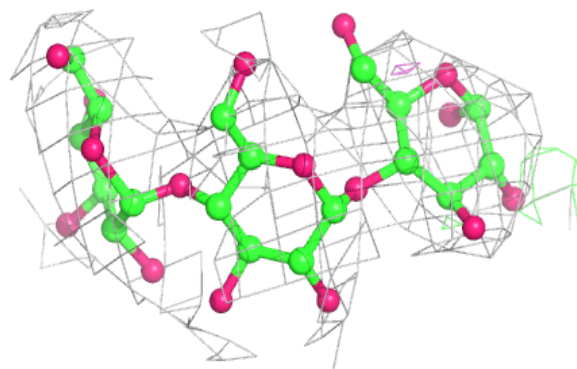


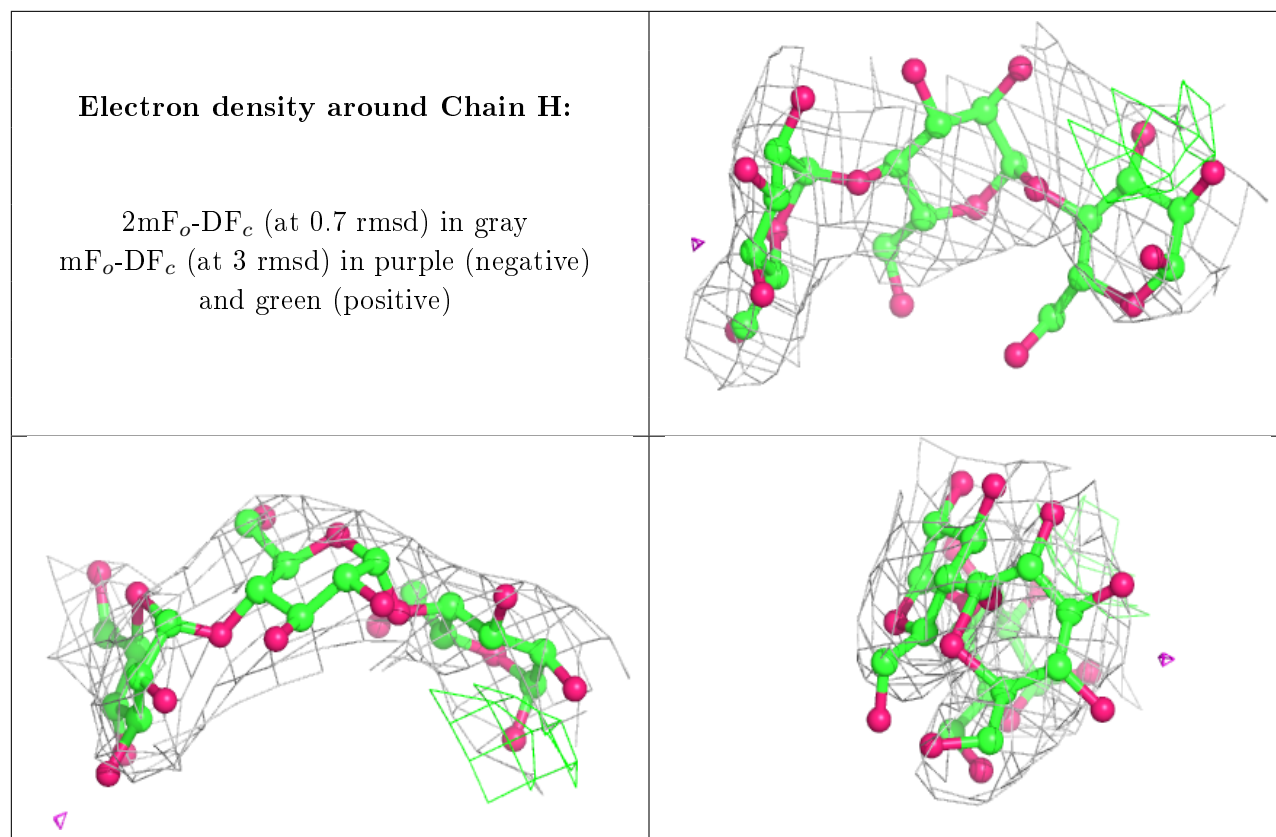
Electron density around Chain F:

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

**Electron density around Chain G:**

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





6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.