



wwPDB X-ray Structure Validation Summary Report

Mar 24, 2022 – 12:28 pm GMT

PDB ID : 6HAK
Title : Crystal structure of HIV-1 reverse transcriptase (RT) in complex with a double stranded RNA represents the RT transcription initiation complex prior to nucleotide incorporation
Authors : Das, K.; Martinez, S.E.; Arnold, E.
Deposited on : 2018-08-07
Resolution : 3.95 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 : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.27
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) : 2.27

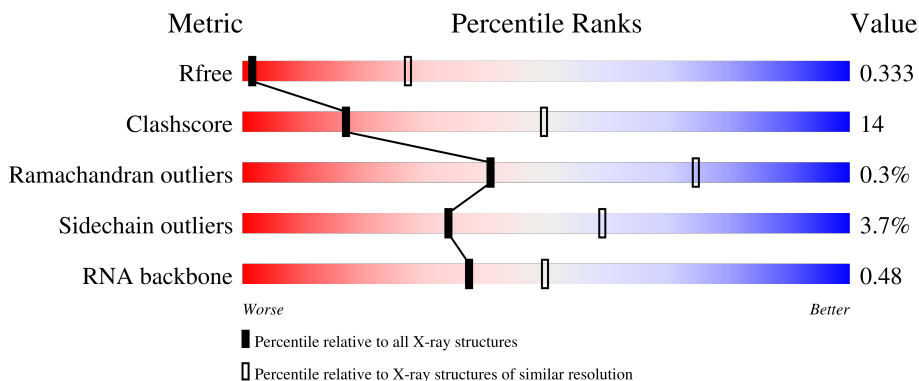
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.95 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	1025 (4.22-3.70)
Clashscore	141614	1085 (4.22-3.70)
Ramachandran outliers	138981	1047 (4.22-3.70)
Sidechain outliers	138945	1039 (4.22-3.70)
RNA backbone	3102	1041 (4.84-3.00)

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	556	65% (green), 33% (yellow), .. (orange/red)
1	C	556	65% (green), 32% (yellow), .. (orange/red)
2	B	444	60% (green), 32% (yellow), 7% (orange/red), . (grey)
2	D	444	61% (green), 30% (yellow), 7% (orange/red), . (grey)
3	E	23	39% (green), 43% (yellow), 13% (orange/red), . (grey)
3	T	23	35% (green), 39% (yellow), 13% (orange/red), 13% (grey)

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	F	17	 41% 41% 18%
4	P	17	 47% 47% 6%
5	G	2	 50% 50%

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 17251 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 Gag-Pol polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	553	Total	C	N	O	S	0	0	0
			4412	2847	738	819	8			
1	C	553	Total	C	N	O	S	0	0	0
			4456	2883	743	822	8			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	initiating methionine	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	258	CYS	GLN	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366
A	498	ASN	ASP	engineered mutation	UNP P03366
C	-1	MET	-	initiating methionine	UNP P03366
C	0	VAL	-	expression tag	UNP P03366
C	258	CYS	GLN	engineered mutation	UNP P03366
C	280	SER	CYS	engineered mutation	UNP P03366
C	498	ASN	ASP	engineered mutation	UNP P03366

- Molecule 2 is a protein called Gag-Pol polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	412	Total	C	N	O	S	0	0	0
			3380	2200	562	612	6			
2	D	415	Total	C	N	O	S	0	0	0
			3396	2207	565	617	7			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-15	MET	-	initiating methionine	UNP P03366
B	-14	ALA	-	expression tag	UNP P03366

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-13	HIS	-	expression tag	UNP P03366
B	-12	HIS	-	expression tag	UNP P03366
B	-11	HIS	-	expression tag	UNP P03366
B	-10	HIS	-	expression tag	UNP P03366
B	-9	HIS	-	expression tag	UNP P03366
B	-8	HIS	-	expression tag	UNP P03366
B	-7	ALA	-	expression tag	UNP P03366
B	-6	LEU	-	expression tag	UNP P03366
B	-5	GLU	-	expression tag	UNP P03366
B	-4	VAL	-	expression tag	UNP P03366
B	-3	LEU	-	expression tag	UNP P03366
B	-2	PHE	-	expression tag	UNP P03366
B	-1	GLN	-	expression tag	UNP P03366
B	0	GLY	-	expression tag	UNP P03366
B	280	SER	CYS	engineered mutation	UNP P03366
D	-15	MET	-	initiating methionine	UNP P03366
D	-14	ALA	-	expression tag	UNP P03366
D	-13	HIS	-	expression tag	UNP P03366
D	-12	HIS	-	expression tag	UNP P03366
D	-11	HIS	-	expression tag	UNP P03366
D	-10	HIS	-	expression tag	UNP P03366
D	-9	HIS	-	expression tag	UNP P03366
D	-8	HIS	-	expression tag	UNP P03366
D	-7	ALA	-	expression tag	UNP P03366
D	-6	LEU	-	expression tag	UNP P03366
D	-5	GLU	-	expression tag	UNP P03366
D	-4	VAL	-	expression tag	UNP P03366
D	-3	LEU	-	expression tag	UNP P03366
D	-2	PHE	-	expression tag	UNP P03366
D	-1	GLN	-	expression tag	UNP P03366
D	0	GLY	-	expression tag	UNP P03366
D	280	SER	CYS	engineered mutation	UNP P03366

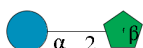
- Molecule 3 is a RNA chain called RNA (5'-R(P*AP*GP*UP*GP*GP*CP*GP*GP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	20	Total	C	N	O	P	0	0	0
			437	194	87	136	20			
3	E	20	Total	C	N	O	P	0	0	0
			437	194	87	136	20			

- Molecule 4 is a RNA chain called RNA (5'-R(P*UP*CP*CP*CP*UP*GP*UP*UP*CP*GP*GP*CP*CP*GP*CP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	P	17	Total	C	N	O	P	0	0	0
			354	158	57	122	17			
4	F	17	Total	C	N	O	P	0	0	0
			354	158	57	122	17			

- Molecule 5 is an oligosaccharide called beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose.



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

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	Mg	0	0
			1	1		

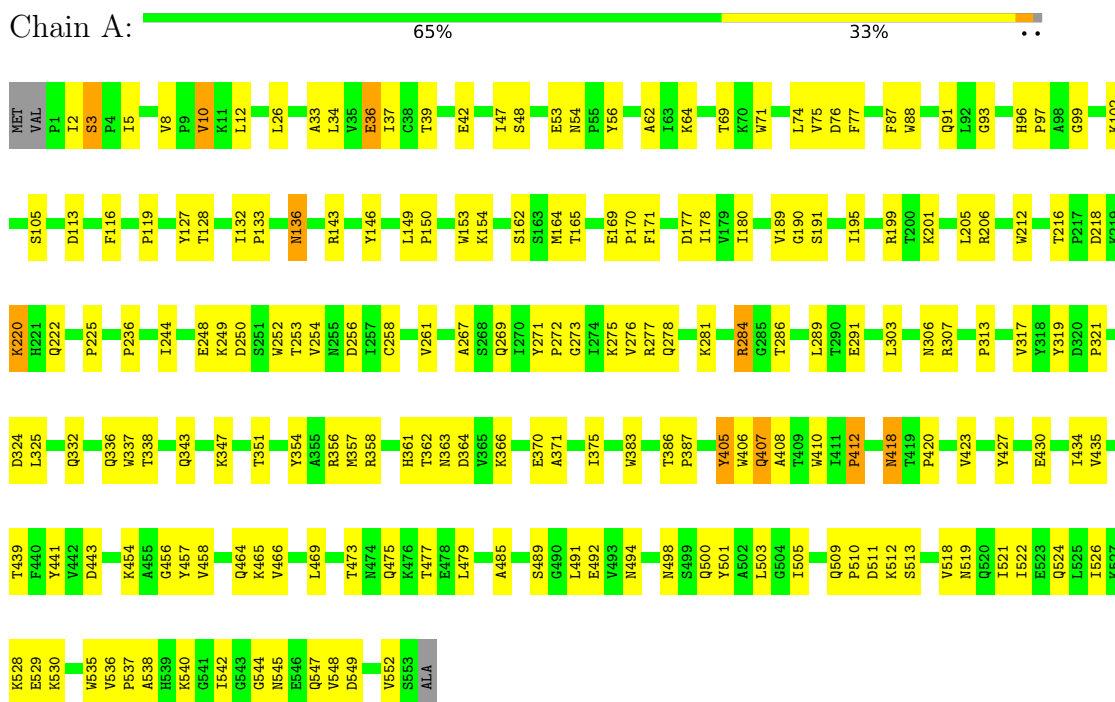
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	1	Total	O	0	0
			1	1		

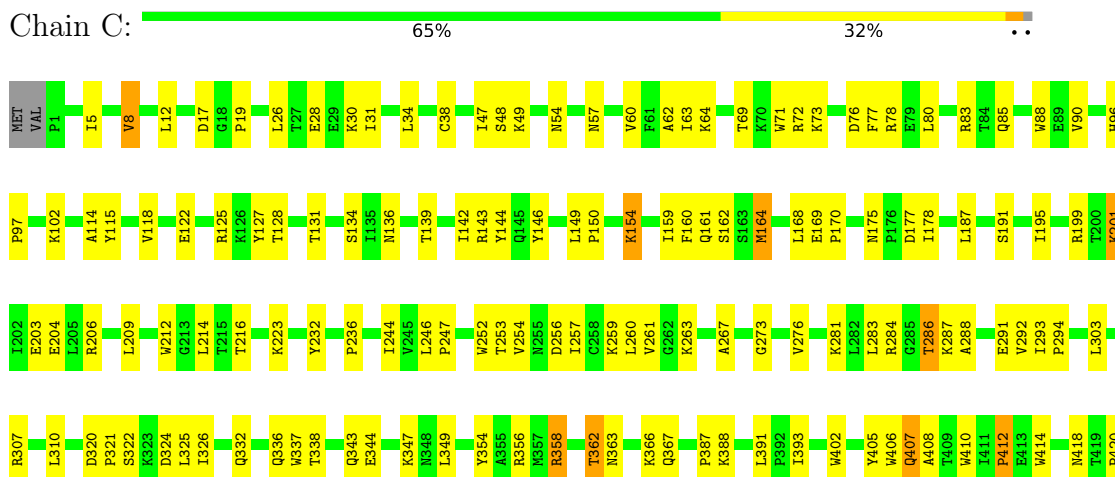
3 Residue-property plots [i](#)

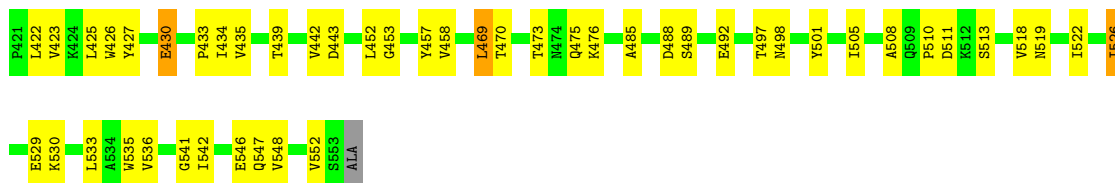
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded 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. 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: Gag-Pol polyprotein

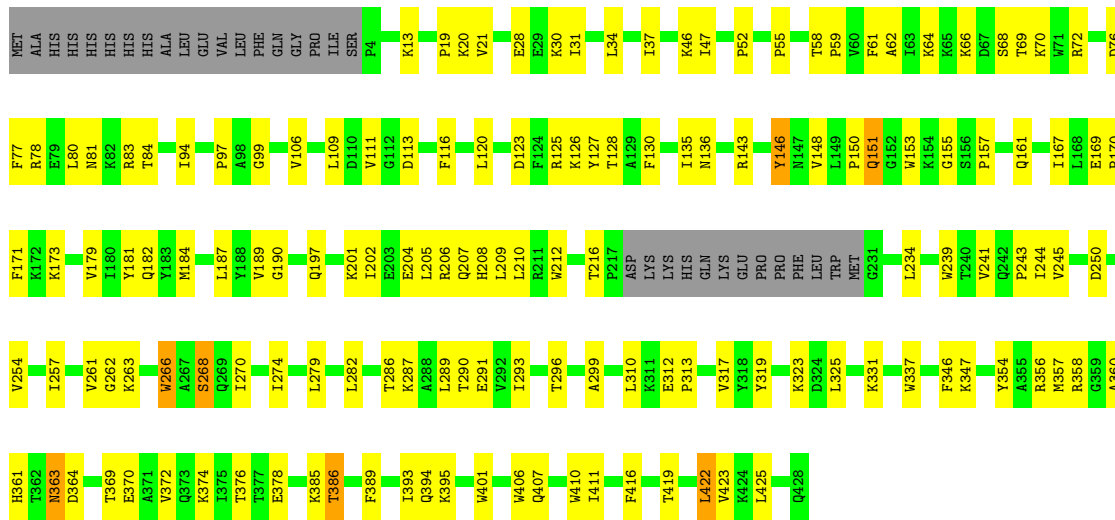


- Molecule 1: Gag-Pol polyprotein

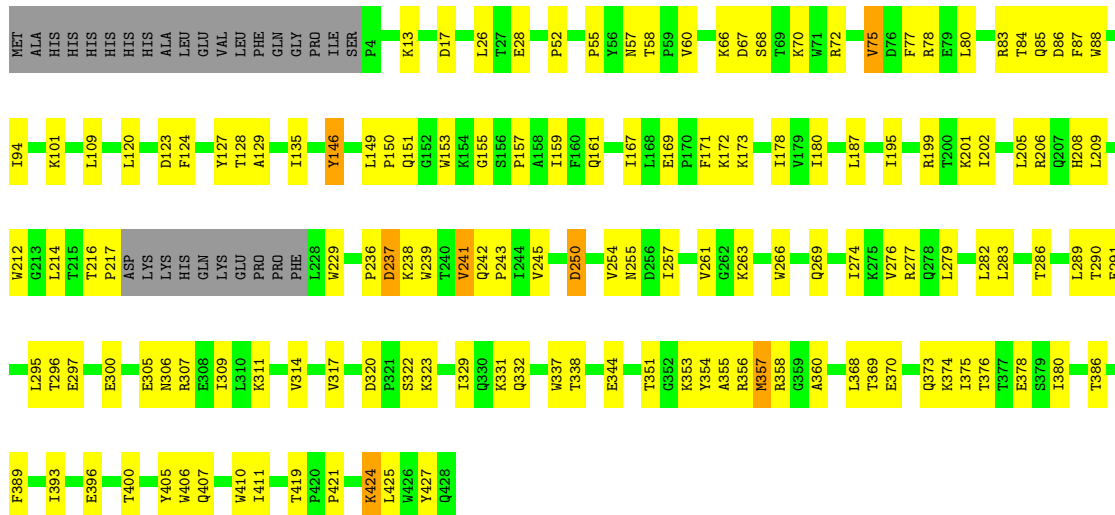




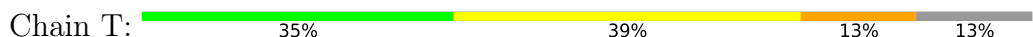
• Molecule 2: Gag-Pol polyprotein

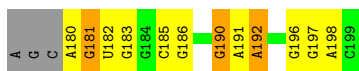


• Molecule 2: Gag-Pol polyprotein

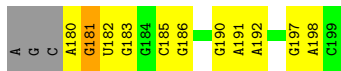


• Molecule 3: RNA (5'-R(P*AP*GP*UP*GP*GP*CP*GP*GP*CP*CP*GP*AP*AP*CP*AP*G P*GP*GP*AP*C)-3')





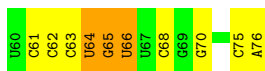
- Molecule 3: RNA (5'-R(P*AP*GP*UP*GP*GP*CP*GP*GP*CP*CP*GP*AP*AP*CP*AP*G P*GP*GP*AP*C)-3')



- Molecule 4: RNA (5'-R(P*UP*CP*CP*CP*UP*GP*UP*UP*CP*GP*GP*CP*CP*GP*CP*C P*A)-3')



- Molecule 4: RNA (5'-R(P*UP*CP*CP*CP*UP*GP*UP*UP*CP*GP*GP*CP*CP*GP*CP*C P*A)-3')



- Molecule 5: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	175.24Å 175.24Å 225.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	87.62 – 3.95 87.62 – 3.95	Depositor EDS
% Data completeness (in resolution range)	99.7 (87.62-3.95) 99.8 (87.62-3.95)	Depositor EDS
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.32 (at 4.01Å)	Xtriage
Refinement program	PHENIX (1.12_2829: ???)	Depositor
R, R_{free}	0.293 , 0.329 0.296 , 0.333	Depositor DCC
R_{free} test set	1505 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å ²)	152.1	Xtriage
Anisotropy	0.167	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.70	EDS
Total number of atoms	17251	wwPDB-VP
Average B, all atoms (Å ²)	191.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.53% 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: MG, GLC, FRU

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	0.35	0/4526	0.58	0/6156
1	C	0.35	0/4571	0.57	0/6214
2	B	0.35	0/3477	0.56	0/4726
2	D	0.36	0/3492	0.59	0/4746
3	E	0.33	0/490	0.85	0/764
3	T	0.47	0/490	1.03	0/764
4	F	0.41	0/392	1.03	0/607
4	P	0.51	0/392	1.19	0/607
All	All	0.36	0/17830	0.64	0/24584

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 [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	4412	0	4391	143	0
1	C	4456	0	4497	131	0
2	B	3380	0	3392	112	0
2	D	3396	0	3411	96	0
3	E	437	0	220	7	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	T	437	0	220	11	0
4	F	354	0	184	5	0
4	P	354	0	184	1	0
5	G	23	0	21	0	0
6	A	1	0	0	0	0
7	A	1	0	0	0	0
All	All	17251	0	16520	468	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 468 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:361:HIS:HD1	1:A:513:SER:HG	1.15	0.87
2:B:266:TRP:HE1	2:B:425:LEU:HD22	1.39	0.87
2:D:78:ARG:HD3	2:D:411:ILE:HG22	1.59	0.84
1:A:503:LEU:HD22	1:A:535:TRP:HB2	1.59	0.84
1:A:54:ASN:HB3	1:A:143:ARG:HH12	1.44	0.82

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	551/556 (99%)	515 (94%)	33 (6%)	3 (0%)	29	66
1	C	551/556 (99%)	519 (94%)	30 (5%)	2 (0%)	34	70
2	B	408/444 (92%)	387 (95%)	21 (5%)	0	100	100
2	D	411/444 (93%)	384 (93%)	26 (6%)	1 (0%)	47	79
All	All	1921/2000 (96%)	1805 (94%)	110 (6%)	6 (0%)	41	74

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	321	PRO
1	C	412	PRO
2	D	357	MET
1	A	284	ARG
1	A	412	PRO

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	473/497 (95%)	456 (96%)	17 (4%)	35	60
1	C	485/497 (98%)	470 (97%)	15 (3%)	40	63
2	B	367/403 (91%)	350 (95%)	17 (5%)	27	54
2	D	370/403 (92%)	357 (96%)	13 (4%)	36	61
All	All	1695/1800 (94%)	1633 (96%)	62 (4%)	34	59

5 of 62 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	386	THR
2	D	237	ASP
1	C	154	LYS
2	D	151	GLN
2	D	410	TRP

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

Mol	Chain	Res	Type
1	A	500	GLN
2	B	208	HIS
2	B	363	ASN
2	B	407	GLN
1	C	500	GLN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	E	19/23 (82%)	3 (15%)	0
3	T	19/23 (82%)	3 (15%)	0
4	F	16/17 (94%)	8 (50%)	0
4	P	16/17 (94%)	8 (50%)	0
All	All	70/80 (87%)	22 (31%)	0

5 of 22 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	T	181	G
3	T	190	G
3	T	192	A
4	P	61	C
4	P	62	C

There are no RNA pucker outliers to report.

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	$\# Z > 2$
5	GLC	G	1	5	11,11,12	0.60	0	15,15,17	1.14	1 (6%)
5	FRU	G	2	5	11,12,12	0.57	0	10,18,18	0.73	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
5	GLC	G	1	5	-	0/2/19/22	0/1/1/1
5	FRU	G	2	5	-	5/5/24/24	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
5	G	1	GLC	C3-C4-C5	-2.04	106.59	110.24

There are no chirality outliers.

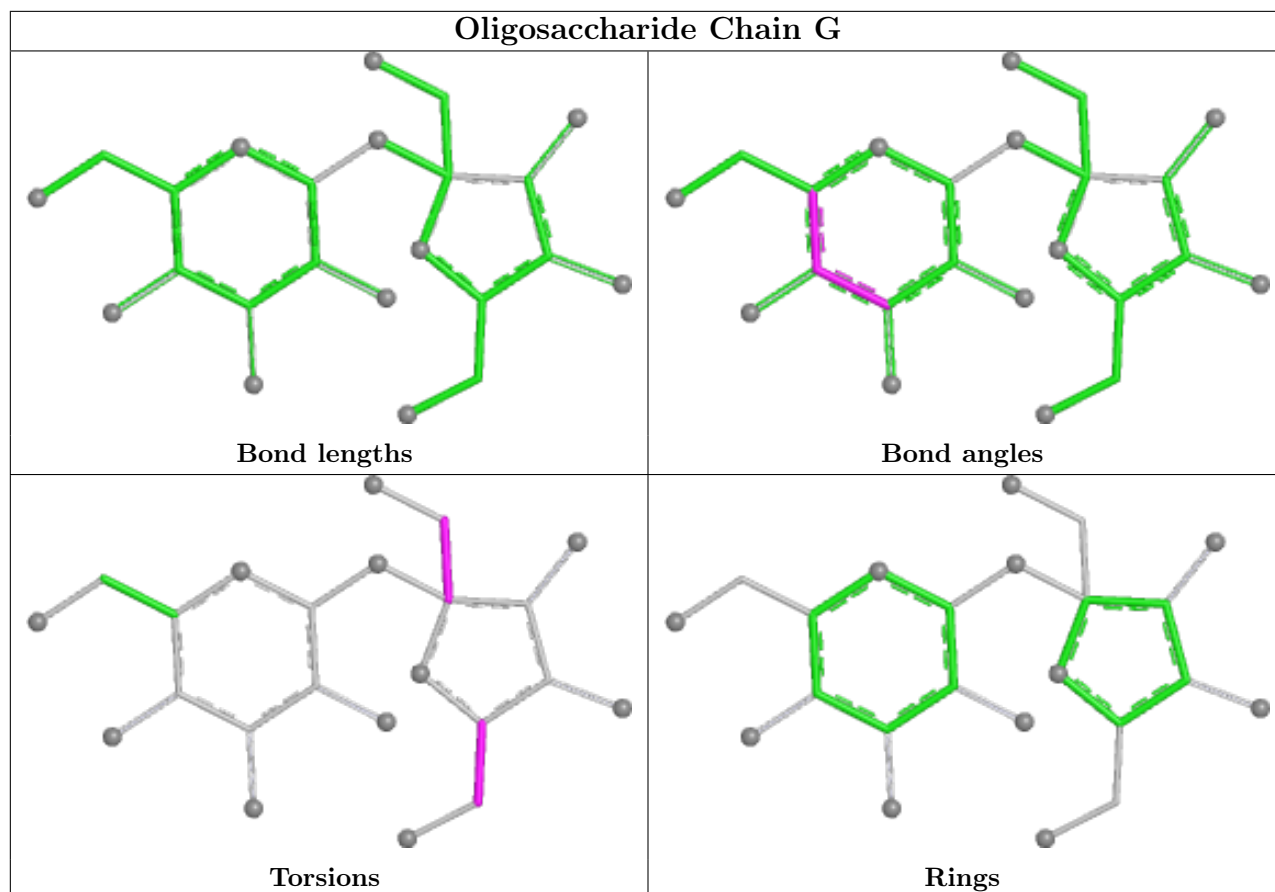
All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	G	2	FRU	O1-C1-C2-C3
5	G	2	FRU	O1-C1-C2-O2
5	G	2	FRU	O1-C1-C2-O5
5	G	2	FRU	O5-C5-C6-O6
5	G	2	FRU	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 1 ligands modelled in this entry, 1 is 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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.

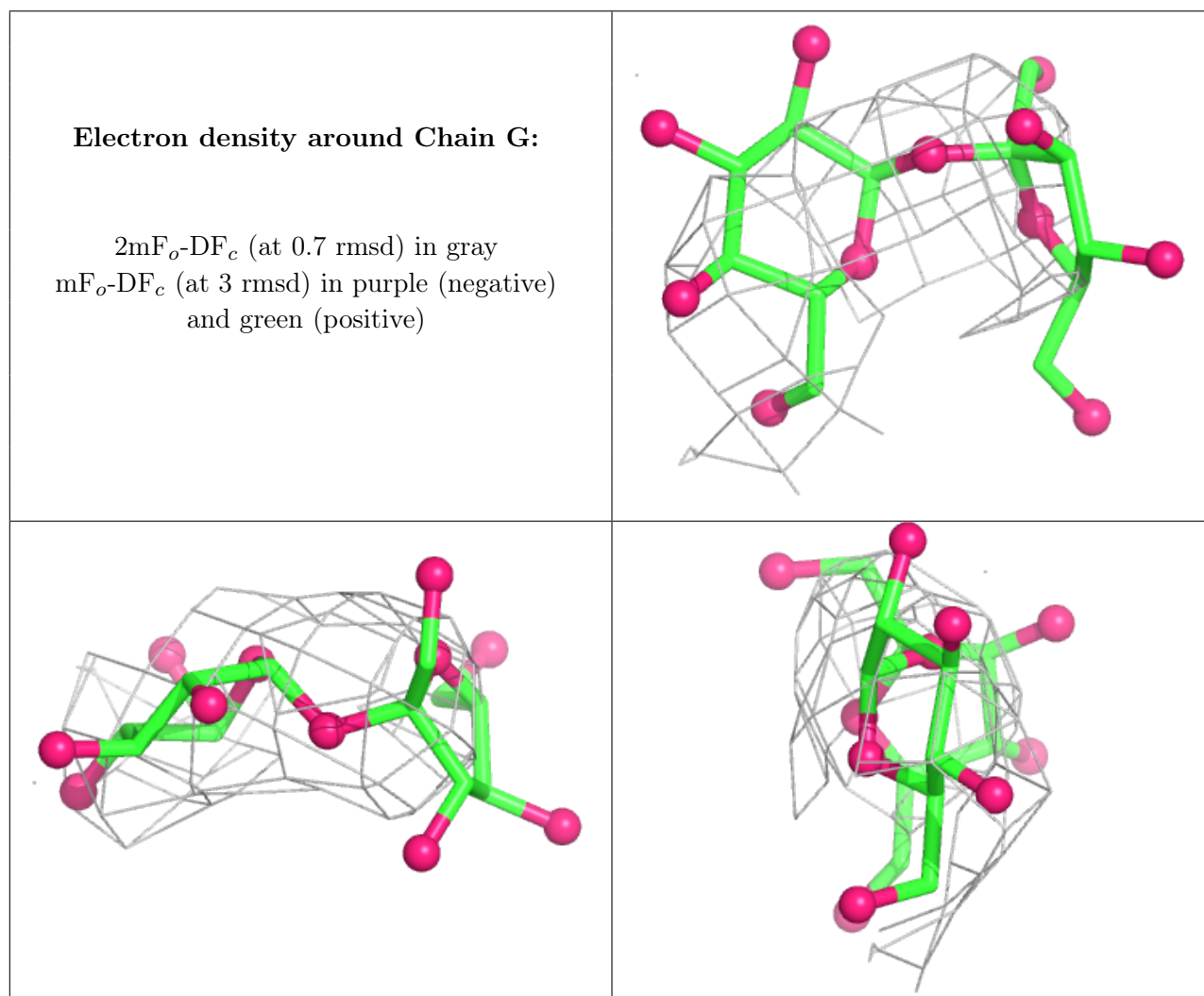
6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.