WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jan 7, 2021 – 06:07 PM GMT

PDB ID : 7AHX
Title : HIV-1 REVERSE TRANSCRIPTASE COMPLEX WITH DNA AND D-ASPARTATE TENOFOVIR WITH BOUND MANGANESE
Authors : Gu, W.; Martinez, S.E.; Nguyen, H.; Xu, H.; Herdewijn, P.; de Jonghe, S.; Das, K.
Deposited on : 2020-09-25
Resolution : 2.73 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 : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : FAILED
buster-report : 1.1.7 (2018)
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
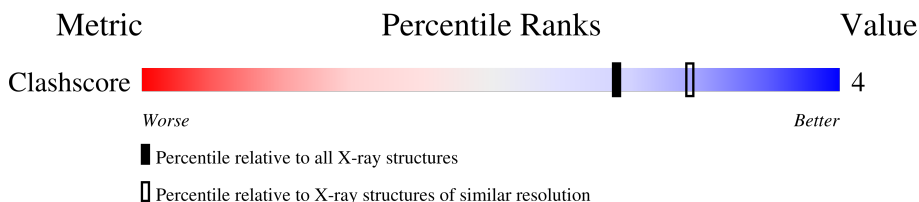Validation Pipeline (wwPDB-VP) : 2.16

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 2.73 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	Percentile Ranks	Value
Clashscore		4

Worse ... *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	1322 (2.76-2.72)

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%

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	556	90% / 10%
1	C	556	89% / 11%
2	B	428	89% / 7% •
2	D	428	87% / 9% •
3	E	27	44% / 44% / 11%
3	T	27	44% / 41% / 15%
4	F	21	62% / 33% / 5%
4	P	21	76% / 19% / 5%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 17778 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
1	A	555	Total 4513	C 2922	N 751	O 832	S 8	0	0	0
1	C	556	Total 4521	C 2927	N 752	O 833	S 9	0	0	0

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
2	B	411	Total 3401	C 2215	N 563	O 616	S 7	0	0	0
2	D	412	Total 3400	C 2212	N 563	O 619	S 6	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	engineered mutation	UNP P03366
D	280	SER	CYS	engineered mutation	UNP P03366

- Molecule 3 is a DNA chain called DNA (5'-D(*TP*GP*GP*TP*CP*GP*GP*CP*GP*CP* CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*CP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	23	Total	C	N	O	P	0	0	0
			477	223	95	136	23			
3	E	24	Total	C	N	O	P	0	0	0
			494	233	97	141	23			

- Molecule 4 is a DNA chain called DNA (5'-D(*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP* TP*CP*GP*GP*(MRG)*CP*GP*CP*CP*(DDG))-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	P	20	Total	C	N	O	P	0	0	0
			403	192	72	120	19			
4	F	20	Total	C	N	O	P	0	0	0
			403	192	72	120	19			

- Molecule 5 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Mn	0	0
			2	2		
5	C	2	Total	Mn	0	0
			2	2		

- Molecule 6 is D-Aspartate Tenofovir (three-letter code: REK) (formula: $C_{13}H_{19}N_6O_7P$) (labeled as "Ligand of Interest" by depositor).
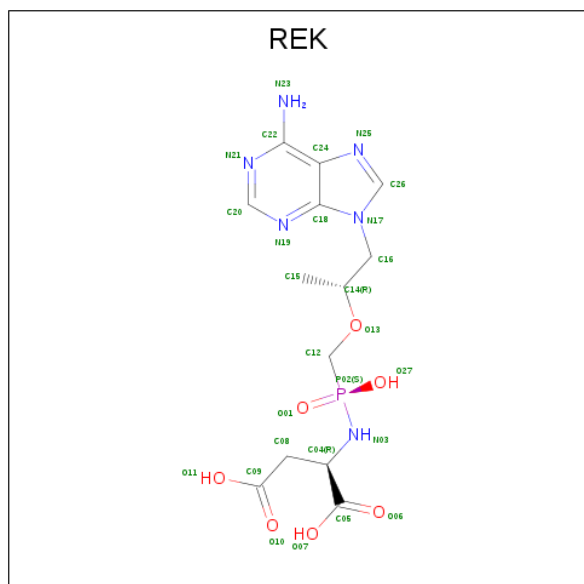
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	A	1	Total 27	C 13	N 6	O 7	P 1	0	0
6	C	1	Total 27	C 13	N 6	O 7	P 1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	36	Total 36	O 36	0	0
7	B	31	Total 31	O 31	0	0
7	C	16	Total 16	O 16	0	0
7	D	15	Total 15	O 15	0	0
7	T	3	Total 3	O 3	0	0
7	P	4	Total 4	O 4	0	0
7	E	1	Total 1	O 1	0	0
7	F	2	Total 2	O 2	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. 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.

Note EDS failed to run properly.
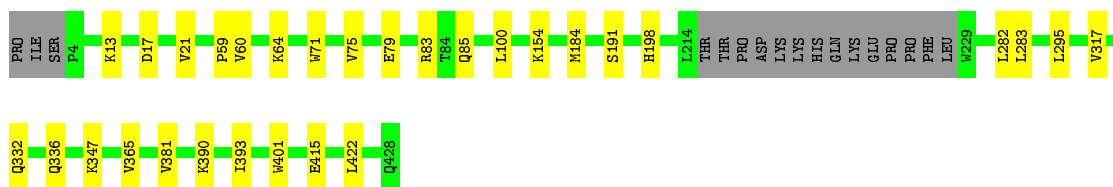
• Molecule 1: Gag-Pol polyprotein

Chain A:
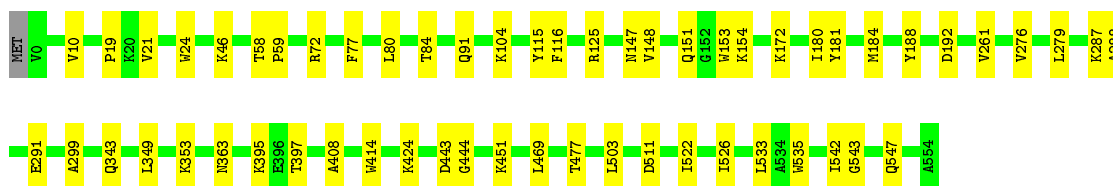
• Molecule 1: Gag-Pol polyprotein

Chain C:
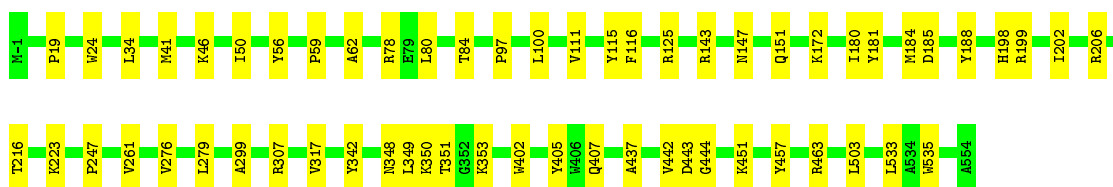
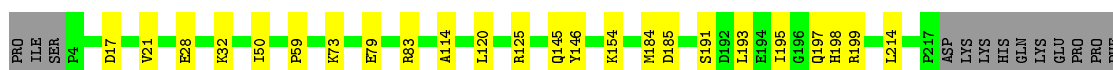• Molecule 2: Gag-Pol polyprotein
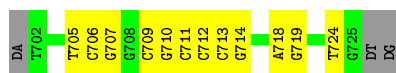
Chain B:

• Molecule 2: Gag-Pol polyprotein

Chain D:

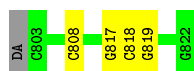LEU TRP MET G231 I244 V245 L246 V254 K263 L264 L282 L283 I293 K331 Q332 R356 K395 F416 P421 L422 V423 K424 L425 Q428

● Molecule 3: DNA (5'-D(*TP*GP*GP*TP*CP*GP*GP*CP*GP*CP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*CP*TP*G)-3')

Chain T:

44%   41%   15%

DA DT G703 C706 G707 G708 C709 G710 C711 C712 C713 G714 G720 G721 G725 DT DG

● Molecule 3: DNA (5'-D(*TP*GP*GP*TP*CP*GP*GP*CP*GP*CP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*CP*TP*G)-3')

Chain E:

44%   44%   11%

DA T702 T705 C706 G707 G708 C709 G710 C711 C712 C713 G714 A718 G719 T724 G725 DT DG

● Molecule 4: DNA (5'-D(*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)*CP*GP*CP*CP*(DDG))-3')

Chain P:

76%   19%   5%

DA C803 C808 G817 C818 G819 G822

● Molecule 4: DNA (5'-D(*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)*CP*GP*CP*CP*(DDG))-3')

Chain F:

62%   33%   5%

DA C803 A804 C808 C814 G815 G816 G817 C818 G819 G822

4　Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	89.42Å　　132.37Å　　138.74Å 90.00°　　98.48°　　90.00°	Depositor
Resolution (Å)	69.80　–　2.73	Depositor
% Data completeness (in resolution range)	99.6 (69.80-2.73)	Depositor
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.34 (at 2.73Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.214　,　0.254	Depositor
Wilson B-factor (Å^2)	58.6	Xtriage
Anisotropy	0.114	Xtriage
L-test for twinning[2]	$<|L|> = 0.47, <L^2> = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	17778	wwPDB-VP
Average B, all atoms (Å^2)	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.90% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: REK, MRG, DDG, MN

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	0.24	0/4631	0.41	0/6293
1	C	0.23	0/4639	0.41	0/6303
2	B	0.24	0/3499	0.41	0/4752
2	D	0.23	0/3497	0.40	0/4751
3	E	0.49	0/555	0.79	0/856
3	T	0.48	0/536	0.78	0/826
4	F	0.54	0/400	0.92	0/612
4	P	0.54	0/400	0.88	0/612
All	All	0.28	0/18157	0.48	0/25005

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	4513	0	4578	36	0
1	C	4521	0	4587	36	0
2	B	3401	0	3431	18	0
2	D	3400	0	3433	23	0
3	E	494	0	269	8	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	T	477	0	256	8	0
4	F	403	0	224	5	0
4	P	403	0	224	4	0
5	A	2	0	0	0	0
5	C	2	0	0	0	0
6	A	27	0	0	1	0
6	C	27	0	0	0	0
7	A	36	0	0	0	0
7	B	31	0	0	0	0
7	C	16	0	0	1	0
7	D	15	0	0	1	0
7	E	1	0	0	0	0
7	F	2	0	0	0	0
7	P	4	0	0	0	0
7	T	3	0	0	0	0
All	All	17778	0	17002	129	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 4.

The worst 5 of 129 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:C:181:TYR:HB2	1:C:188:TYR:HB3	1.74	0.70
1:A:125:ARG:HD3	1:A:147:ASN:HA	1.76	0.67
1:A:181:TYR:HB2	1:A:188:TYR:HB3	1.78	0.66
1:C:172:LYS:HE2	1:C:180:ILE:HB	1.80	0.63
2:D:244:ILE:HD13	2:D:425:LEU:HD11	1.80	0.63

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

There are no protein backbone outliers to report in this entry.

5.3.2 Protein sidechains (i)

There are no protein residues with a non-rotameric sidechain to report in this entry.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

4 non-standard protein/DNA/RNA residues 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.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 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.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier.

The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand REK C 603

Ligand REK A 603
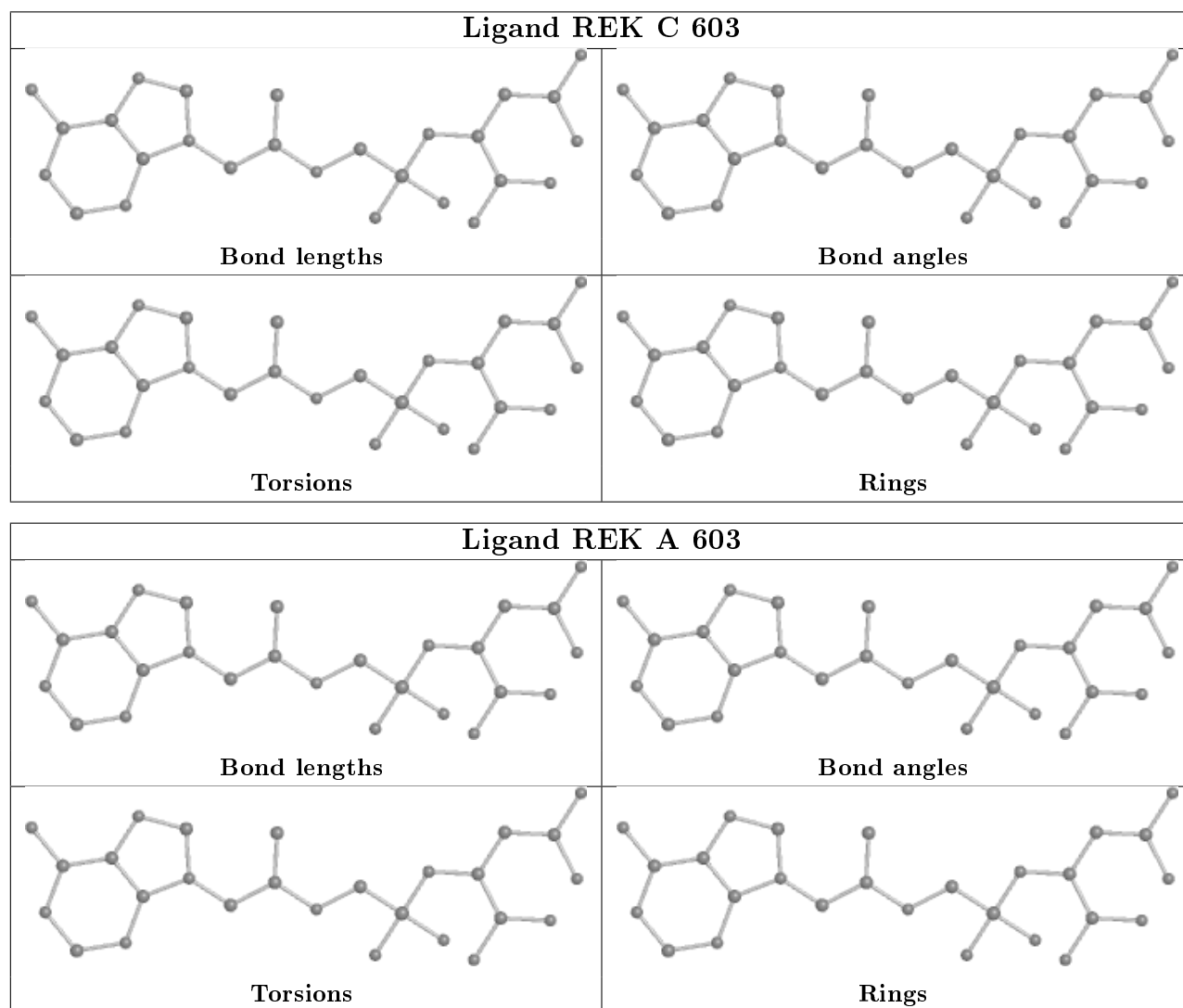
5.7 Other polymers (i)

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)

EDS failed to run properly - this section is therefore empty.

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

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates (i)

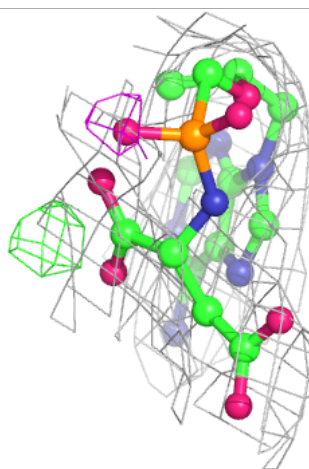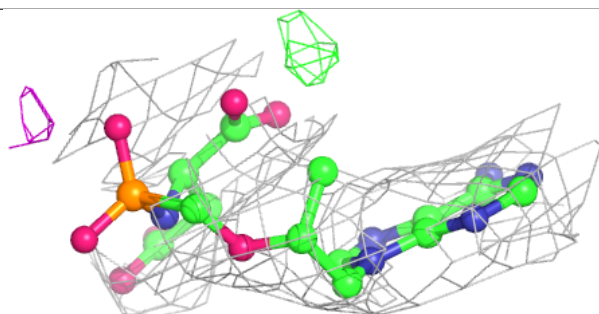EDS failed to run properly - this section is therefore empty.
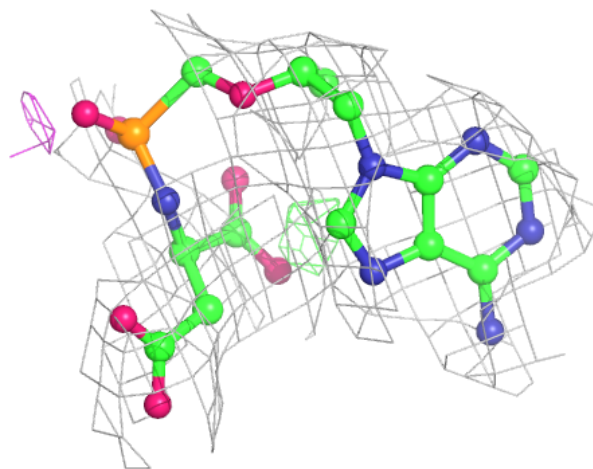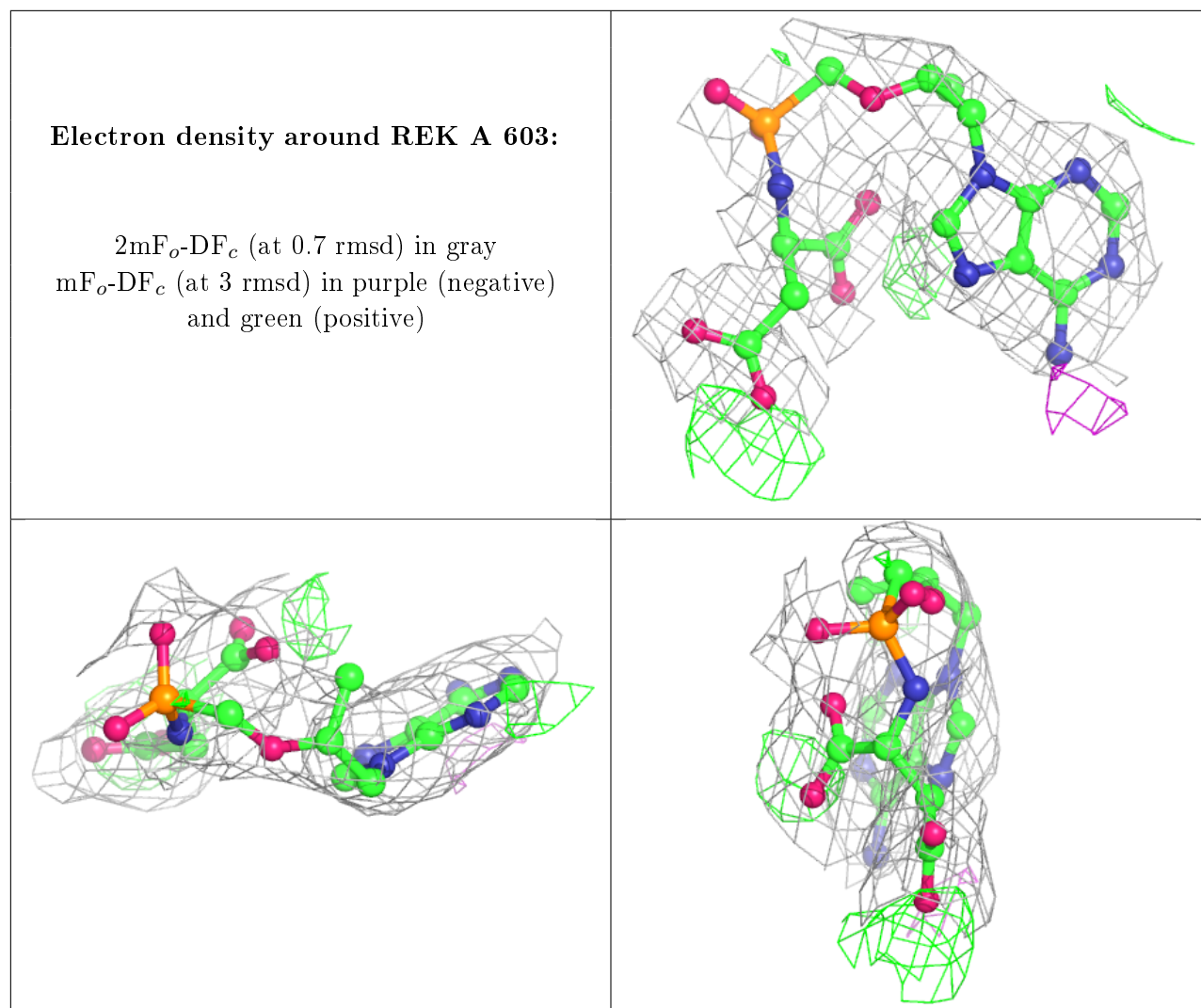
6.4 Ligands (i)

EDS failed to run properly - this section is therefore empty.

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 REK C 603:

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

Electron density around REK A 603:

$2mF_o$-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 ⓘ

EDS failed to run properly - this section is therefore empty.