



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 20, 2020 – 01:55 PM BST

PDB ID : 4K4Y  
Title : Cocksackievirus B3 polymerase elongation complex (r2+1\_form)  
Authors : Gong, P.; Peersen, O.B.  
Deposited on : 2013-04-12  
Resolution : 2.72 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](mailto: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  
buster-report : 1.1.7 (2018)  
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

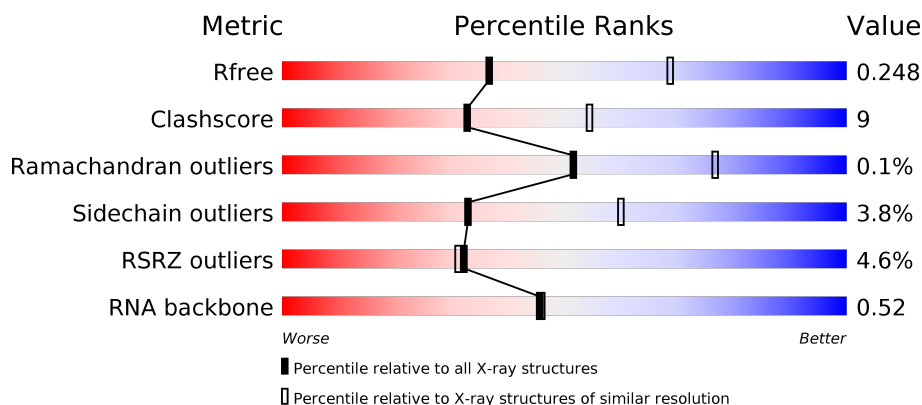
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.72 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	3359 (2.74-2.70)
Clashscore	141614	3686 (2.74-2.70)
Ramachandran outliers	138981	3622 (2.74-2.70)
Sidechain outliers	138945	3623 (2.74-2.70)
RSRZ outliers	127900	3276 (2.74-2.70)
RNA backbone	3102	1067 (3.00-2.44)

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  $\geq 3$ , 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	472	<div> <div>4%</div> <div>75%</div> <div>21%</div> <div>..</div> </div>
1	E	472	<div> <div>4%</div> <div>75%</div> <div>22%</div> <div>..</div> </div>
1	I	472	<div> <div>5%</div> <div>74%</div> <div>23%</div> <div>..</div> </div>
1	M	472	<div> <div>4%</div> <div>78%</div> <div>19%</div> <div>..</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	B	24	
2	F	24	
2	J	24	
2	N	24	
3	C	15	
3	G	15	
3	K	15	
3	O	15	
4	D	9	
4	H	9	
4	L	9	
4	P	9	

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
6	ACT	A	502	-	-	X	-
6	ACT	E	503	-	-	X	-
6	ACT	I	503	-	-	X	-
7	GOL	I	504	-	-	X	-
7	GOL	M	503	-	-	X	-

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 18284 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 RNA-dependent RNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	463	Total	C	N	O	S	0	0	0
			3693	2372	614	685	22			
1	E	463	Total	C	N	O	S	0	0	0
			3693	2372	614	685	22			
1	I	463	Total	C	N	O	S	0	0	0
			3693	2372	614	685	22			
1	M	463	Total	C	N	O	S	0	0	0
			3693	2372	614	685	22			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	252	ILE	LEU	VARIANT	UNP Q66338
A	463	GLY	-	EXPRESSION TAG	UNP Q66338
A	464	SER	-	EXPRESSION TAG	UNP Q66338
A	465	SER	-	EXPRESSION TAG	UNP Q66338
A	466	SER	-	EXPRESSION TAG	UNP Q66338
A	467	HIS	-	EXPRESSION TAG	UNP Q66338
A	468	HIS	-	EXPRESSION TAG	UNP Q66338
A	469	HIS	-	EXPRESSION TAG	UNP Q66338
A	470	HIS	-	EXPRESSION TAG	UNP Q66338
A	471	HIS	-	EXPRESSION TAG	UNP Q66338
A	472	HIS	-	EXPRESSION TAG	UNP Q66338
E	252	ILE	LEU	VARIANT	UNP Q66338
E	463	GLY	-	EXPRESSION TAG	UNP Q66338
E	464	SER	-	EXPRESSION TAG	UNP Q66338
E	465	SER	-	EXPRESSION TAG	UNP Q66338
E	466	SER	-	EXPRESSION TAG	UNP Q66338
E	467	HIS	-	EXPRESSION TAG	UNP Q66338
E	468	HIS	-	EXPRESSION TAG	UNP Q66338
E	469	HIS	-	EXPRESSION TAG	UNP Q66338
E	470	HIS	-	EXPRESSION TAG	UNP Q66338
E	471	HIS	-	EXPRESSION TAG	UNP Q66338

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	472	HIS	-	EXPRESSION TAG	UNP Q66338
I	252	ILE	LEU	VARIANT	UNP Q66338
I	463	GLY	-	EXPRESSION TAG	UNP Q66338
I	464	SER	-	EXPRESSION TAG	UNP Q66338
I	465	SER	-	EXPRESSION TAG	UNP Q66338
I	466	SER	-	EXPRESSION TAG	UNP Q66338
I	467	HIS	-	EXPRESSION TAG	UNP Q66338
I	468	HIS	-	EXPRESSION TAG	UNP Q66338
I	469	HIS	-	EXPRESSION TAG	UNP Q66338
I	470	HIS	-	EXPRESSION TAG	UNP Q66338
I	471	HIS	-	EXPRESSION TAG	UNP Q66338
I	472	HIS	-	EXPRESSION TAG	UNP Q66338
M	252	ILE	LEU	VARIANT	UNP Q66338
M	463	GLY	-	EXPRESSION TAG	UNP Q66338
M	464	SER	-	EXPRESSION TAG	UNP Q66338
M	465	SER	-	EXPRESSION TAG	UNP Q66338
M	466	SER	-	EXPRESSION TAG	UNP Q66338
M	467	HIS	-	EXPRESSION TAG	UNP Q66338
M	468	HIS	-	EXPRESSION TAG	UNP Q66338
M	469	HIS	-	EXPRESSION TAG	UNP Q66338
M	470	HIS	-	EXPRESSION TAG	UNP Q66338
M	471	HIS	-	EXPRESSION TAG	UNP Q66338
M	472	HIS	-	EXPRESSION TAG	UNP Q66338

- Molecule 2 is a RNA chain called RNA (5'-R(\*AP\*AP\*GP\*UP\*CP\*UP\*CP\*CP\*AP\*GP\*GP\*UP\*CP\*UP\*CP\*UP\*CP\*GP\*UP\*CP\*GP\*AP\*AP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	18	Total	C	N	O	P	0	0	0
			376	168	61	129	18			
2	F	17	Total	C	N	O	P	0	0	0
			356	159	58	122	17			
2	J	18	Total	C	N	O	P	0	0	0
			376	168	61	129	18			
2	N	17	Total	C	N	O	P	0	0	0
			356	159	58	122	17			

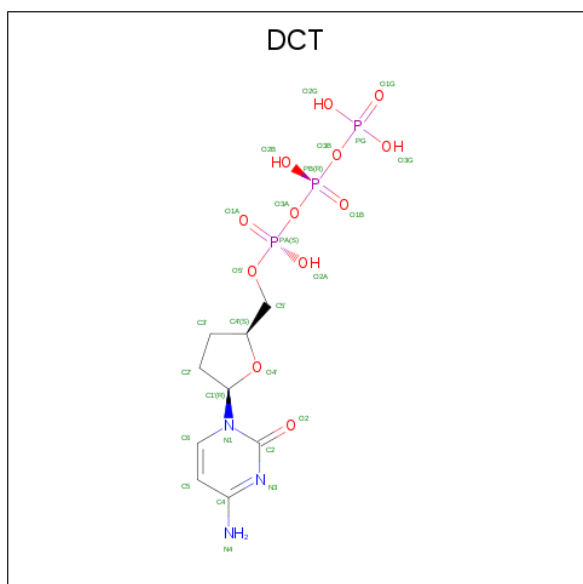
- Molecule 3 is DNA/RNA hybrid called DNA/RNA (5'-R(\*UP\*GP\*UP\*UP\*CP\*GP\*AP\*C P\*GP\*AP\*GP\*AP\*GP\*AP\*(DOC))-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	12	Total	C	N	O	P	0	0	0
			258	116	51	79	12			
3	G	12	Total	C	N	O	P	0	0	0
			258	116	51	79	12			
3	K	12	Total	C	N	O	P	0	0	0
			258	116	51	79	12			
3	O	12	Total	C	N	O	P	0	0	0
			258	116	51	79	12			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	5	Total	C	N	O	P	0	0	0
			113	50	25	33	5			
4	H	5	Total	C	N	O	P	0	0	0
			113	50	25	33	5			
4	L	5	Total	C	N	O	P	0	0	0
			113	50	25	33	5			
4	P	5	Total	C	N	O	P	0	0	0
			113	50	25	33	5			

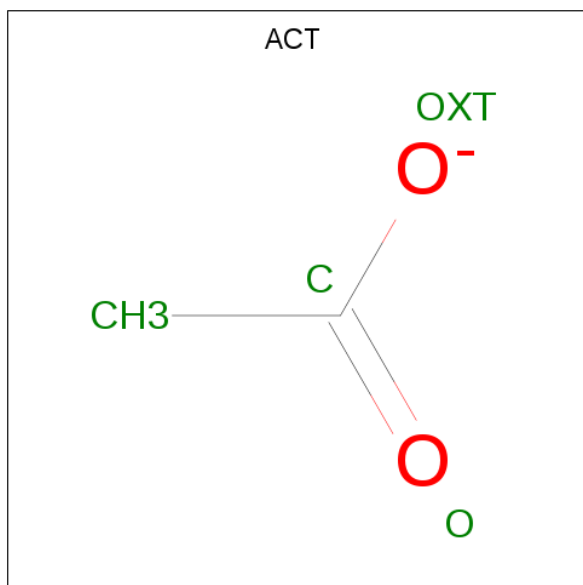
- Molecule 5 is 2',3'-DIDEOXYCYTIDINE 5'-TRIPHOSPHATE (three-letter code: DCT) (formula: C<sub>9</sub>H<sub>16</sub>N<sub>3</sub>O<sub>12</sub>P<sub>3</sub>).



Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	E	1	Total	C	N	O	P	0	0
			27	9	3	12	3		
5	I	1	Total	C	N	O	P	0	0
			27	9	3	12	3		
5	M	1	Total	C	N	O	P	0	0
			27	9	3	12	3		

- Molecule 6 is ACETATE ION (three-letter code: ACT) (formula:  $C_2H_3O_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			4	2	2		
6	E	1	Total	C	O	0	0
			4	2	2		
6	I	1	Total	C	O	0	0
			4	2	2		

- Molecule 7 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	C	O	0	0
			6	3	3		
7	A	1	Total	C	O	0	0
			6	3	3		
7	A	1	Total	C	O	0	0
			6	3	3		
7	E	1	Total	C	O	0	0
			6	3	3		
7	E	1	Total	C	O	0	0
			6	3	3		
7	E	1	Total	C	O	0	0
			6	3	3		
7	I	1	Total	C	O	0	0
			6	3	3		
7	I	1	Total	C	O	0	0
			6	3	3		
7	I	1	Total	C	O	0	0
			6	3	3		
7	M	1	Total	C	O	0	0
			6	3	3		
7	M	1	Total	C	O	0	0
			6	3	3		
7	M	1	Total	C	O	0	0
			6	3	3		

- Molecule 8 is water.

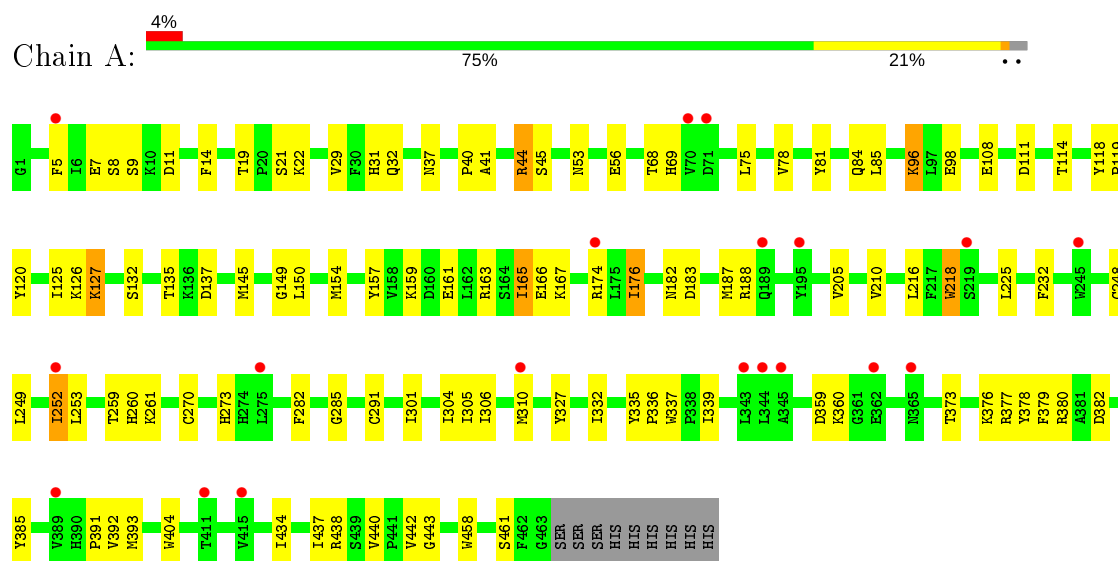


Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	80	Total 80	O 80	0	0
8	B	9	Total 9	O 9	0	0
8	C	6	Total 6	O 6	0	0
8	E	71	Total 71	O 71	0	0
8	F	13	Total 13	O 13	0	0
8	G	7	Total 7	O 7	0	0
8	I	73	Total 73	O 73	0	0
8	J	16	Total 16	O 16	0	0
8	K	5	Total 5	O 5	0	0
8	L	2	Total 2	O 2	0	0
8	M	69	Total 69	O 69	0	0
8	N	12	Total 12	O 12	0	0
8	O	9	Total 9	O 9	0	0

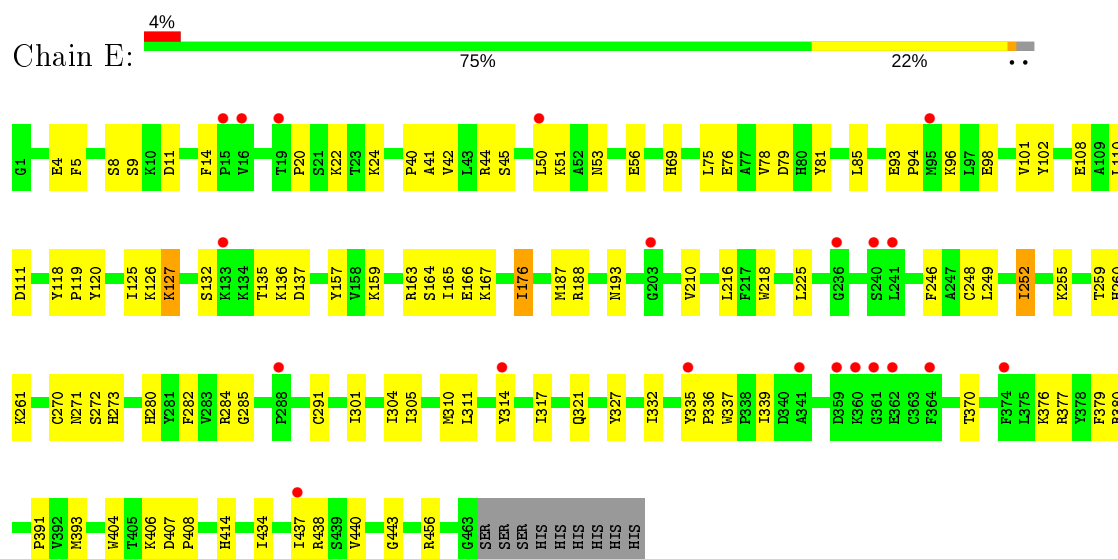
### 3 Residue-property plots [i](#)

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 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.

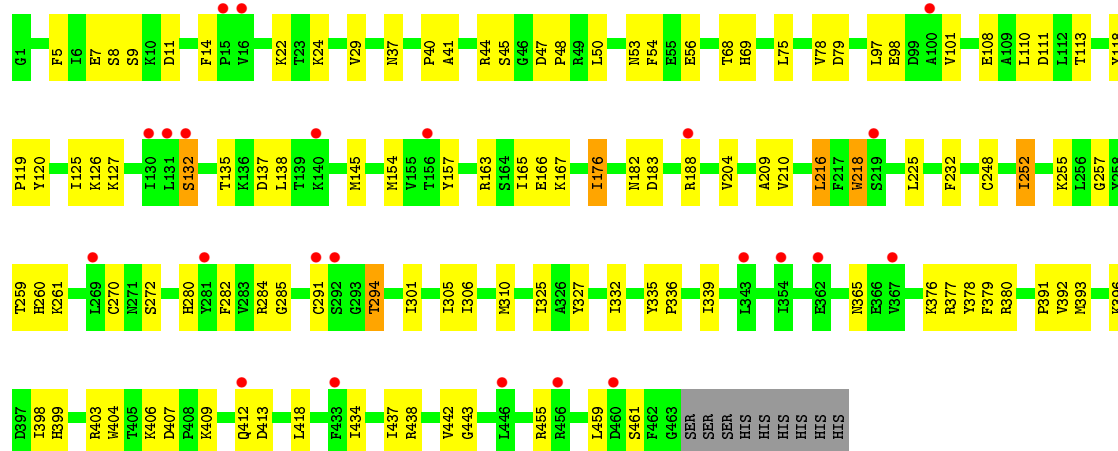
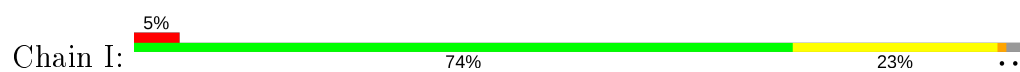
#### • Molecule 1: RNA-dependent RNA polymerase



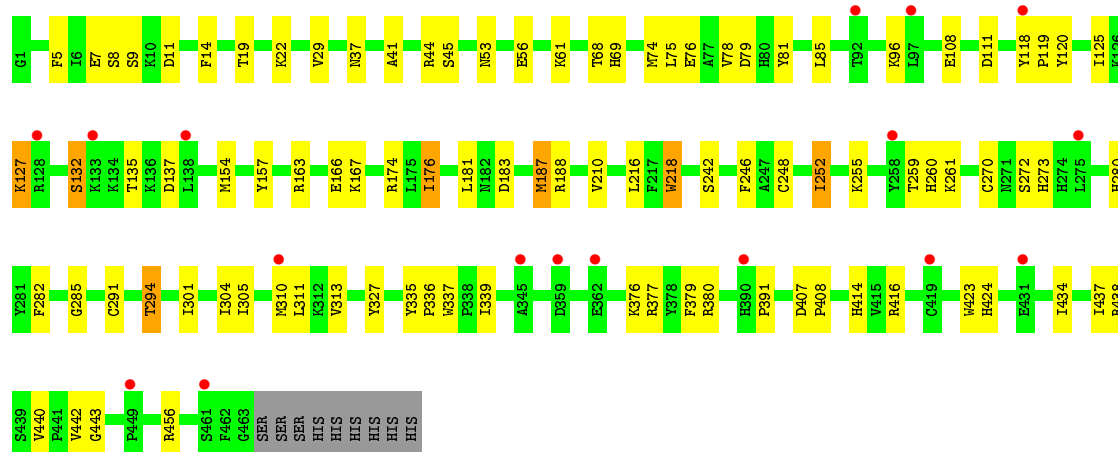
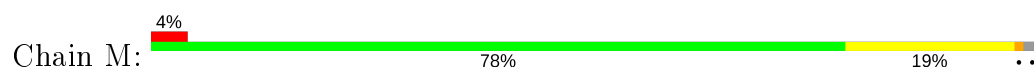
#### • Molecule 1: RNA-dependent RNA polymerase



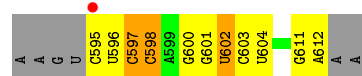
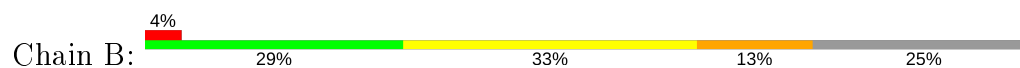
#### • Molecule 1: RNA-dependent RNA polymerase



- Molecule 1: RNA-dependent RNA polymerase



- Molecule 2: RNA (5'-R(\*AP\*AP\*GP\*UP\*CP\*UP\*CP\*CP\*AP\*GP\*GP\*UP\*CP\*UP\*CP\*UP\*CP\*GP\*UP\*CP\*GP\*AP\*AP\*A)-3')



- Molecule 2: RNA (5'-R(\*AP\*AP\*GP\*UP\*CP\*UP\*CP\*CP\*AP\*GP\*GP\*UP\*CP\*UP\*CP\*UP\*CP\*GP\*UP\*CP\*GP\*AP\*AP\*A)-3')




- Molecule 2: RNA (5'-R(\*AP\*AP\*GP\*UP\*CP\*UP\*CP\*CP\*AP\*GP\*GP\*UP\*CP\*UP\*CP\*UP\*CP\*GP\*UP\*CP\*GP\*AP\*AP\*A)-3')

Chain J: 



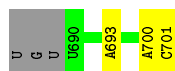
- Molecule 2: RNA (5'-R(\*AP\*AP\*GP\*UP\*CP\*UP\*CP\*CP\*AP\*GP\*GP\*UP\*CP\*UP\*CP\*UP\*CP\*GP\*UP\*CP\*GP\*AP\*AP\*A)-3')

Chain N: 



- Molecule 3: DNA/RNA (5'-R(\*UP\*GP\*UP\*UP\*CP\*GP\*AP\*CP\*GP\*AP\*GP\*AP\*GP\*AP\*(DOC))-3')

Chain C: 



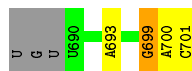
- Molecule 3: DNA/RNA (5'-R(\*UP\*GP\*UP\*UP\*CP\*GP\*AP\*CP\*GP\*AP\*GP\*AP\*GP\*AP\*(DOC))-3')

Chain G: 



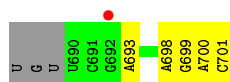
- Molecule 3: DNA/RNA (5'-R(\*UP\*GP\*UP\*UP\*CP\*GP\*AP\*CP\*GP\*AP\*GP\*AP\*GP\*AP\*(DOC))-3')

Chain K: 

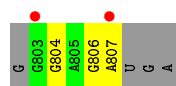
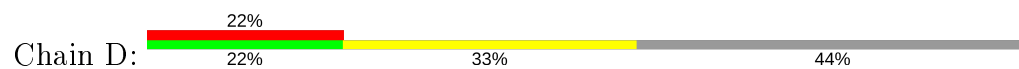


- Molecule 3: DNA/RNA (5'-R(\*UP\*GP\*UP\*UP\*CP\*GP\*AP\*CP\*GP\*AP\*GP\*AP\*GP\*AP\*(DOC))-3')

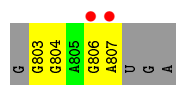
Chain O: 



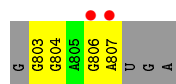
- Molecule 4: RNA (5'-R(\*GP\*GP\*GP\*AP\*GP\*AP\*UP\*GP\*A)-3')



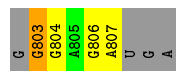
- Molecule 4: RNA (5'-R(\*GP\*GP\*GP\*AP\*GP\*AP\*UP\*GP\*A)-3')



- Molecule 4: RNA (5'-R(\*GP\*GP\*GP\*AP\*GP\*AP\*UP\*GP\*A)-3')



- Molecule 4: RNA (5'-R(\*GP\*GP\*GP\*AP\*GP\*AP\*UP\*GP\*A)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	60.37Å 60.40Å 194.03Å 89.92° 89.96° 79.75°	Depositor
Resolution (Å)	43.76 – 2.72 43.76 – 2.72	Depositor EDS
% Data completeness (in resolution range)	97.6 (43.76-2.72) 97.3 (43.76-2.72)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.91 (at 2.73Å)	Xtriage
Refinement program	PHENIX 1.8.2 _1309	Depositor
R, $R_{free}$	0.199 , 0.247 0.198 , 0.248	Depositor DCC
$R_{free}$ test set	3563 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	43.2	Xtriage
Anisotropy	0.809	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 26.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.53$ , $\langle L^2 \rangle = 0.37$	Xtriage
Estimated twinning fraction	0.387 for -h,-k,l 0.387 for k,h,-l 0.397 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	18284	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.54% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: GOL, DCT, DOC, ACT

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	0/3786	0.59	0/5125
1	E	0.44	0/3786	0.58	0/5125
1	I	0.45	0/3786	0.59	0/5125
1	M	0.46	0/3786	0.59	0/5125
2	B	0.52	0/417	1.05	0/646
2	F	0.56	0/395	1.13	1/612 (0.2%)
2	J	0.56	0/417	1.02	0/646
2	N	0.55	0/395	1.05	1/612 (0.2%)
3	C	0.54	0/269	1.10	0/418
3	G	0.59	0/269	1.16	0/418
3	K	0.58	0/269	1.23	1/418 (0.2%)
3	O	0.58	0/269	1.15	0/418
4	D	0.54	0/127	1.05	0/197
4	H	0.50	0/127	1.09	0/197
4	L	0.40	0/127	1.07	0/197
4	P	0.48	0/127	0.95	0/197
All	All	0.47	0/18352	0.71	3/25476 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	E	0	1
1	I	0	1
1	M	0	1
All	All	0	3

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	598	C	N1-C2-O2	6.22	122.63	118.90
2	N	598	C	N1-C2-O2	5.41	122.15	118.90
3	K	699	G	N7-C8-N9	-5.24	110.48	113.10

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	272	SER	Peptide
1	I	272	SER	Peptide
1	M	272	SER	Peptide

## 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	3693	0	3661	68	0
1	E	3693	0	3661	72	0
1	I	3693	0	3661	78	0
1	M	3693	0	3661	58	1
2	B	376	0	194	7	0
2	F	356	0	183	6	0
2	J	376	0	194	5	0
2	N	356	0	183	6	0
3	C	258	0	132	1	0
3	G	258	0	132	3	0
3	K	258	0	132	2	0
3	O	258	0	132	3	0
4	D	113	0	56	2	0
4	H	113	0	56	1	0
4	L	113	0	56	2	0
4	P	113	0	56	2	0
5	A	27	0	12	5	0
5	E	27	0	12	0	0
5	I	27	0	12	0	0
5	M	27	0	12	3	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	4	0	3	3	0
6	E	4	0	3	2	0
6	I	4	0	3	2	0
7	A	18	0	24	3	0
7	E	18	0	24	4	0
7	I	18	0	24	6	0
7	M	18	0	24	5	0
8	A	80	0	0	1	0
8	B	9	0	0	0	0
8	C	6	0	0	0	0
8	E	71	0	0	3	0
8	F	13	0	0	1	0
8	G	7	0	0	1	0
8	I	73	0	0	6	0
8	J	16	0	0	0	0
8	K	5	0	0	0	0
8	L	2	0	0	0	0
8	M	69	0	0	4	0
8	N	12	0	0	0	0
8	O	9	0	0	0	0
All	All	18284	0	16303	296	1

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 9.

The worst 5 of 296 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:E:259:THR:HG22	1:E:261:LYS:H	1.33	0.94
1:A:259:THR:HG22	1:A:261:LYS:H	1.35	0.92
1:I:259:THR:HG22	1:I:261:LYS:H	1.36	0.90
1:M:259:THR:HG22	1:M:261:LYS:H	1.38	0.88
1:A:161:GLU:OE2	1:A:174:ARG:NH2	2.08	0.87

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:M:135:THR:OG1	1:M:313:VAL:O[1_545]	2.19	0.01

## 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	461/472 (98%)	439 (95%)	21 (5%)	1 (0%)	47	72
1	E	461/472 (98%)	445 (96%)	16 (4%)	0	100	100
1	I	461/472 (98%)	443 (96%)	17 (4%)	1 (0%)	47	72
1	M	461/472 (98%)	443 (96%)	18 (4%)	0	100	100
All	All	1844/1888 (98%)	1770 (96%)	72 (4%)	2 (0%)	51	77

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	360	LYS
1	I	365	ASN

### 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	404/413 (98%)	386 (96%)	18 (4%)	27	53
1	E	404/413 (98%)	390 (96%)	14 (4%)	36	63
1	I	404/413 (98%)	391 (97%)	13 (3%)	39	67
1	M	404/413 (98%)	387 (96%)	17 (4%)	30	56
All	All	1616/1652 (98%)	1554 (96%)	62 (4%)	33	60

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

Mol	Chain	Res	Type
1	E	252	ILE
1	I	44	ARG
1	M	218	TRP
1	E	440	VAL
1	I	75	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	32	GLN
1	E	32	GLN
1	I	32	GLN
1	M	32	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	17/24 (70%)	5 (29%)	1 (5%)
2	F	17/24 (70%)	5 (29%)	1 (5%)
2	J	17/24 (70%)	5 (29%)	1 (5%)
2	N	17/24 (70%)	5 (29%)	1 (5%)
3	C	10/15 (66%)	1 (10%)	0
3	G	10/15 (66%)	1 (10%)	0
3	K	10/15 (66%)	1 (10%)	0
3	O	10/15 (66%)	1 (10%)	0
4	D	4/9 (44%)	1 (25%)	0
4	H	5/9 (55%)	1 (20%)	1 (20%)
4	L	4/9 (44%)	1 (25%)	0
4	P	5/9 (55%)	1 (20%)	1 (20%)
All	All	126/192 (65%)	28 (22%)	6 (4%)

5 of 28 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	597	C
2	B	598	C
2	B	602	U
2	B	611	G
2	B	612	A

5 of 6 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
4	H	803	G
4	P	803	G
2	J	596	U
2	F	596	U
2	N	596	U

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

4 non-standard protein/DNA/RNA residues 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	DOC	O	701	3,2	14,19,20	0.74	0	13,26,29	1.55	4 (30%)
3	DOC	G	701	3,2	14,19,20	0.69	0	13,26,29	1.49	2 (15%)
3	DOC	K	701	3,2	14,19,20	0.65	0	13,26,29	1.26	2 (15%)
3	DOC	C	701	3,2	14,19,20	0.62	0	13,26,29	1.39	2 (15%)

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	DOC	O	701	3,2	-	0/4/18/19	0/2/2/2
3	DOC	G	701	3,2	-	0/4/18/19	0/2/2/2
3	DOC	K	701	3,2	-	0/4/18/19	0/2/2/2
3	DOC	C	701	3,2	-	0/4/18/19	0/2/2/2

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	K	701	DOC	C2-N3-C4	3.28	119.67	116.34
3	O	701	DOC	C2-N3-C4	3.13	119.52	116.34

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	G	701	DOC	C2-N3-C4	3.09	119.48	116.34
3	C	701	DOC	C2-N3-C4	2.72	119.10	116.34
3	C	701	DOC	N4-C4-N3	2.61	120.61	116.49

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](#)

19 ligands 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	$\# Z  > 2$
7	GOL	A	505	-	5,5,5	0.37	0	5,5,5	0.34	0
7	GOL	I	502	-	5,5,5	0.39	0	5,5,5	0.30	0
7	GOL	M	502	-	5,5,5	0.33	0	5,5,5	0.34	0
5	DCT	I	501	-	22,28,28	0.90	0	26,43,43	1.54	4 (15%)
7	GOL	I	505	-	5,5,5	0.24	0	5,5,5	0.44	0
5	DCT	M	501	-	22,28,28	0.90	0	26,43,43	1.79	7 (26%)
6	ACT	E	503	-	1,3,3	1.36	0	0,3,3	0.00	-
7	GOL	A	504	-	5,5,5	0.30	0	5,5,5	0.37	0
7	GOL	A	503	-	5,5,5	0.42	0	5,5,5	0.46	0
7	GOL	M	504	-	5,5,5	0.33	0	5,5,5	0.61	0
5	DCT	E	501	-	22,28,28	0.94	0	26,43,43	1.90	8 (30%)
7	GOL	E	504	-	5,5,5	0.38	0	5,5,5	0.40	0
7	GOL	I	504	-	5,5,5	0.35	0	5,5,5	0.82	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
6	ACT	A	502	-	1,3,3	1.62	0	0,3,3	0.00	-
6	ACT	I	503	-	1,3,3	1.19	0	0,3,3	0.00	-
5	DCT	A	501	-	22,28,28	0.92	0	26,43,43	1.51	5 (19%)
7	GOL	M	503	-	5,5,5	0.37	0	5,5,5	1.28	0
7	GOL	E	505	-	5,5,5	0.35	0	5,5,5	0.52	0
7	GOL	E	502	-	5,5,5	0.32	0	5,5,5	0.24	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
7	GOL	A	505	-	-	2/4/4/4	-
7	GOL	I	502	-	-	2/4/4/4	-
7	GOL	M	502	-	-	4/4/4/4	-
5	DCT	I	501	-	-	3/19/31/31	0/2/2/2
7	GOL	I	505	-	-	0/4/4/4	-
5	DCT	A	501	-	-	7/19/31/31	0/2/2/2
5	DCT	M	501	-	-	3/19/31/31	0/2/2/2
7	GOL	A	504	-	-	4/4/4/4	-
7	GOL	A	503	-	-	2/4/4/4	-
7	GOL	M	504	-	-	2/4/4/4	-
5	DCT	E	501	-	-	3/19/31/31	0/2/2/2
7	GOL	E	504	-	-	4/4/4/4	-
7	GOL	I	504	-	-	2/4/4/4	-
7	GOL	M	503	-	-	3/4/4/4	-
7	GOL	E	505	-	-	2/4/4/4	-
7	GOL	E	502	-	-	4/4/4/4	-

There are no bond length outliers.

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	E	501	DCT	PB-O3B-PG	-4.13	118.64	132.83
5	E	501	DCT	C2-N3-C4	4.05	120.45	116.34
5	I	501	DCT	C2-N3-C4	3.92	120.31	116.34
5	I	501	DCT	PB-O3B-PG	-3.85	119.62	132.83

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	M	501	DCT	PB-O3B-PG	-3.66	120.26	132.83

There are no chirality outliers.

5 of 47 torsion outliers are listed below:

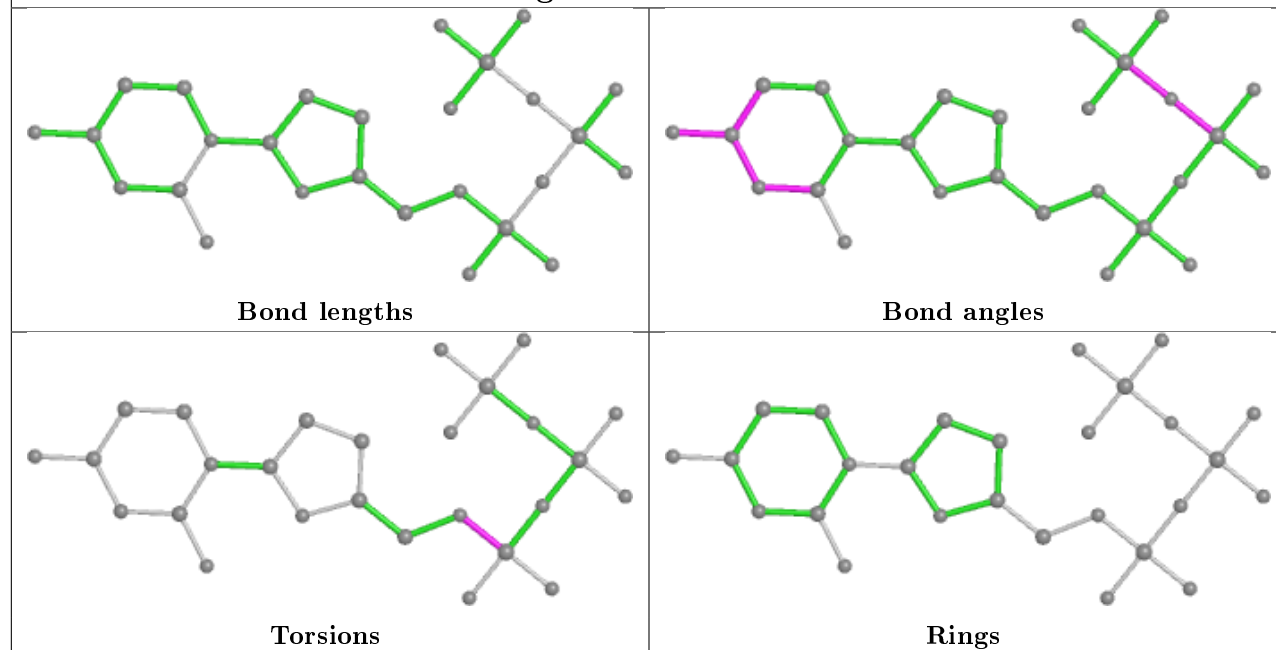
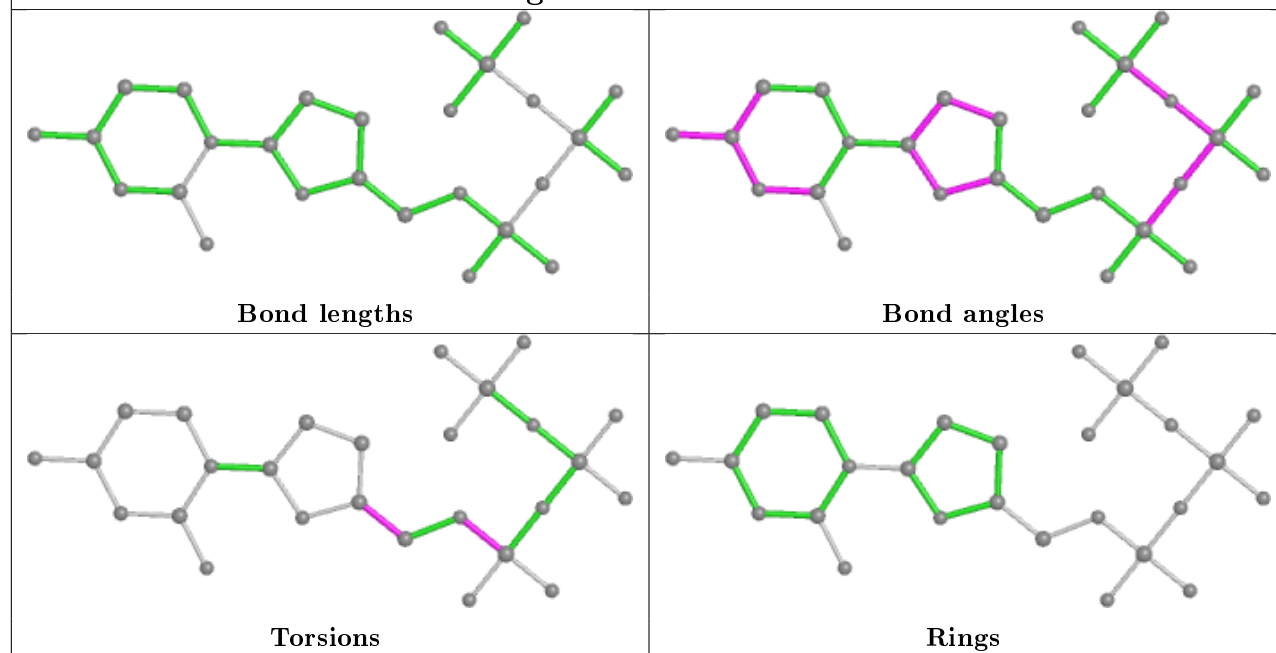
Mol	Chain	Res	Type	Atoms
7	M	502	GOL	O1-C1-C2-O2
7	M	502	GOL	O1-C1-C2-C3
7	M	502	GOL	C1-C2-C3-O3
5	I	501	DCT	C5'-O5'-PA-O1A
5	I	501	DCT	C5'-O5'-PA-O2A

There are no ring outliers.

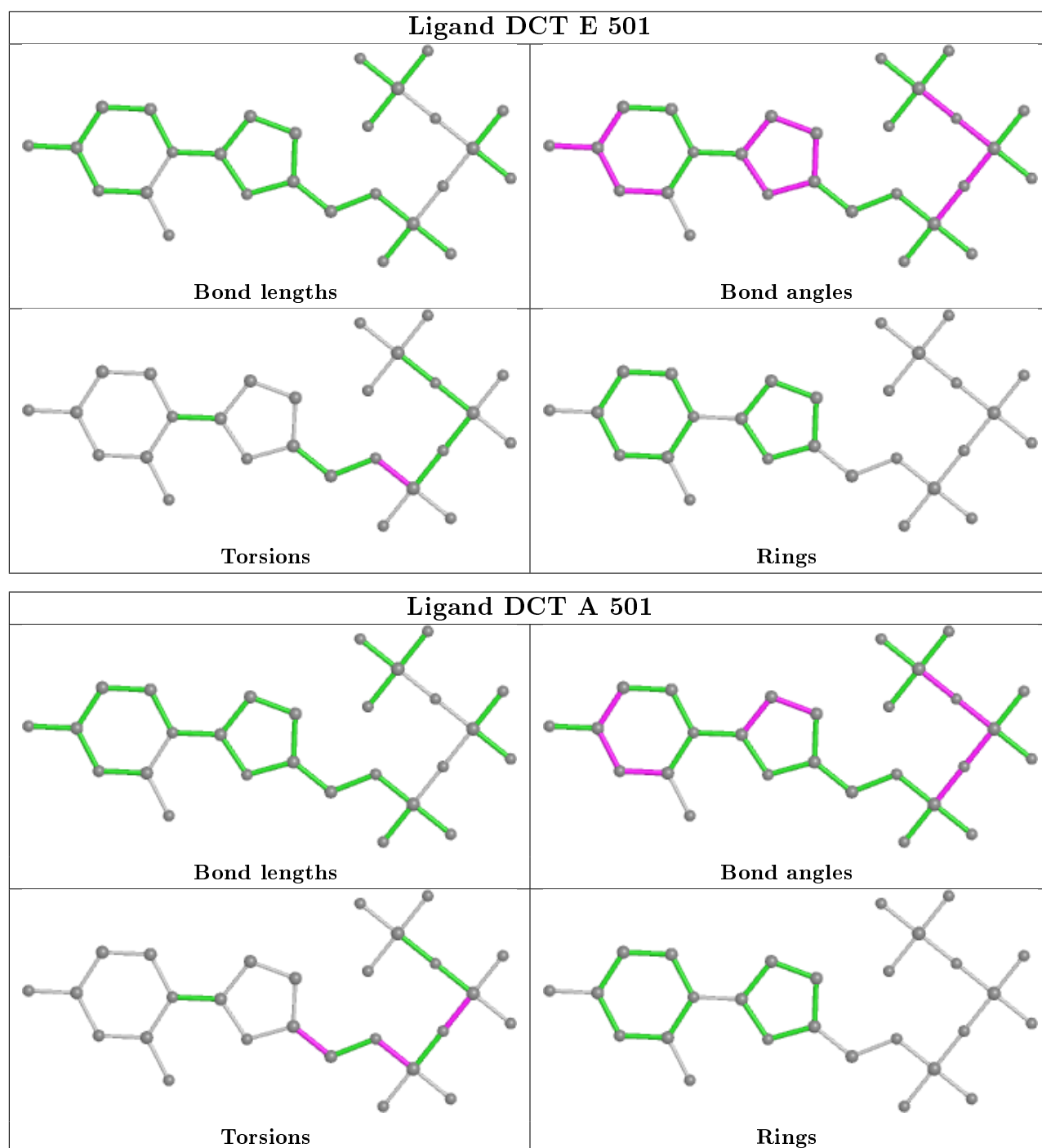
13 monomers are involved in 33 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	I	502	GOL	1	0
5	M	501	DCT	3	0
6	E	503	ACT	2	0
7	A	503	GOL	3	0
7	M	504	GOL	1	0
7	E	504	GOL	1	0
7	I	504	GOL	5	0
6	A	502	ACT	3	0
6	I	503	ACT	2	0
5	A	501	DCT	5	0
7	M	503	GOL	4	0
7	E	505	GOL	1	0
7	E	502	GOL	2	0

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 than 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 DCT I 501****Ligand DCT M 501**





## 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	463/472 (98%)	0.20	19 (4%) 37 36	30, 44, 63, 124	0
1	E	463/472 (98%)	0.25	21 (4%) 33 32	29, 43, 62, 123	0
1	I	463/472 (98%)	0.27	23 (4%) 28 28	29, 43, 62, 125	0
1	M	463/472 (98%)	0.24	17 (3%) 41 41	29, 43, 62, 125	0
2	B	18/24 (75%)	0.11	1 (5%) 24 23	38, 55, 129, 139	0
2	F	17/24 (70%)	-0.02	1 (5%) 22 21	37, 46, 116, 126	0
2	J	18/24 (75%)	0.03	0 100 100	35, 53, 128, 138	0
2	N	17/24 (70%)	0.24	2 (11%) 4 3	36, 47, 116, 126	0
3	C	11/15 (73%)	-0.25	0 100 100	38, 53, 115, 115	0
3	G	11/15 (73%)	-0.19	0 100 100	35, 52, 115, 116	0
3	K	11/15 (73%)	-0.16	0 100 100	35, 51, 114, 115	0
3	O	11/15 (73%)	-0.07	1 (9%) 9 7	36, 49, 114, 114	0
4	D	5/9 (55%)	1.60	2 (40%) 0 0	124, 127, 131, 136	0
4	H	5/9 (55%)	1.89	2 (40%) 0 0	123, 127, 128, 135	0
4	L	5/9 (55%)	3.00	2 (40%) 0 0	124, 128, 130, 133	0
4	P	5/9 (55%)	0.79	0 100 100	124, 127, 129, 136	0
All	All	1986/2080 (95%)	0.24	91 (4%) 32 31	29, 44, 73, 139	0

The worst 5 of 91 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	M	362	GLU	7.5
4	L	806	G	6.1
4	L	807	A	5.7
4	H	806	G	5.2
1	A	344	LEU	4.8

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

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
3	DOC	O	701	18/19	0.96	0.18	30,36,40,41	0
3	DOC	G	701	18/19	0.96	0.16	33,37,40,40	0
3	DOC	K	701	18/19	0.97	0.19	31,37,39,40	0
3	DOC	C	701	18/19	0.98	0.16	30,37,40,40	0

## 6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 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<sup>th</sup> 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.

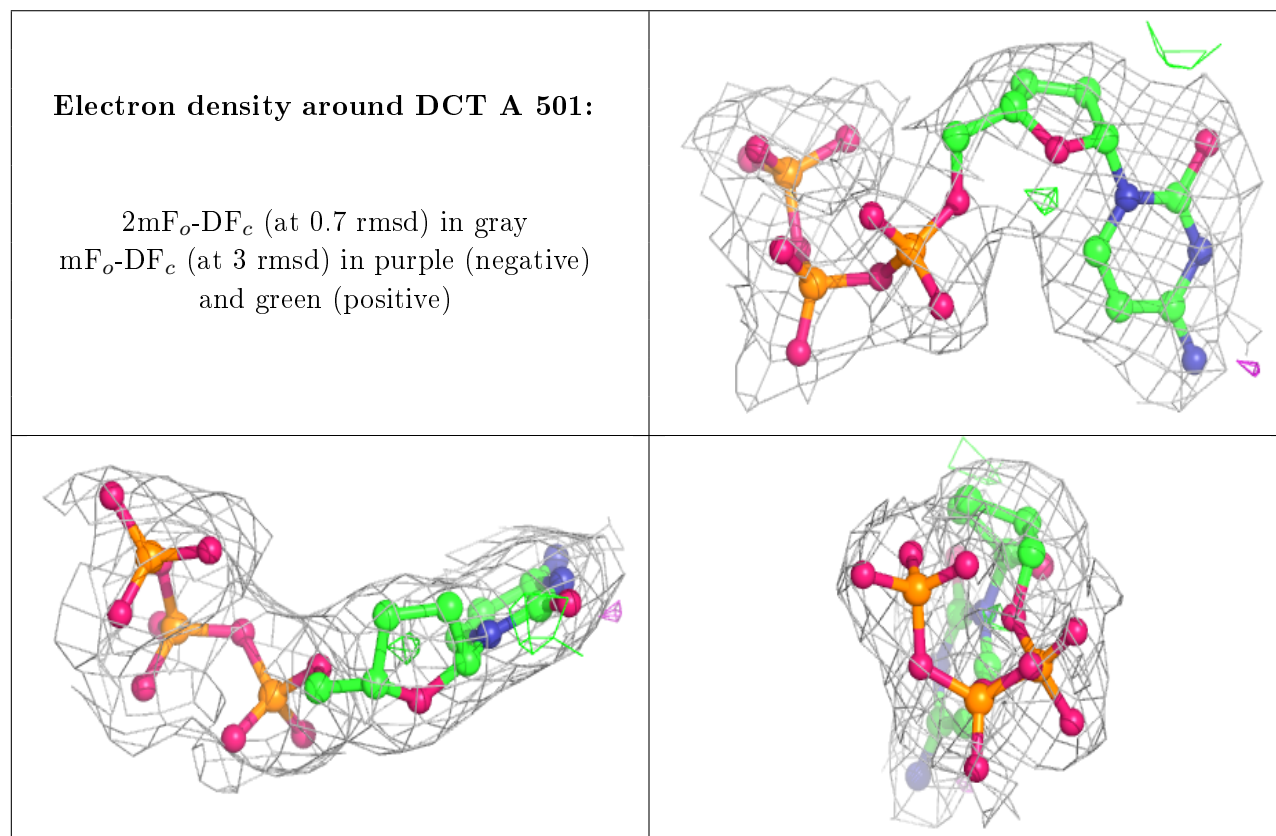
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
7	GOL	M	502	6/6	0.80	0.20	54,66,74,74	0
7	GOL	A	505	6/6	0.83	0.13	52,66,75,79	0
6	ACT	E	503	4/4	0.83	0.66	67,69,69,74	0
7	GOL	I	502	6/6	0.85	0.20	53,71,76,76	0
7	GOL	E	502	6/6	0.85	0.17	49,67,72,72	0
7	GOL	M	503	6/6	0.86	0.54	44,51,53,53	0
7	GOL	I	504	6/6	0.86	0.35	47,53,58,64	0
7	GOL	A	503	6/6	0.87	0.39	45,51,52,52	0
6	ACT	A	502	4/4	0.88	0.18	66,67,69,70	0
5	DCT	A	501	27/27	0.89	0.15	39,52,89,94	0
5	DCT	M	501	27/27	0.90	0.16	38,55,85,89	0
5	DCT	E	501	27/27	0.90	0.19	39,57,90,91	0
6	ACT	I	503	4/4	0.92	0.18	61,61,63,65	0
5	DCT	I	501	27/27	0.92	0.16	39,52,96,99	0
7	GOL	E	504	6/6	0.92	0.46	47,50,54,59	0
7	GOL	I	505	6/6	0.94	0.16	34,41,43,48	0
7	GOL	M	504	6/6	0.94	0.14	40,45,46,47	0
7	GOL	A	504	6/6	0.94	0.19	43,44,46,51	0

*Continued on next page...*

*Continued from previous page...*

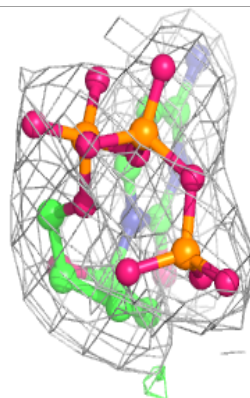
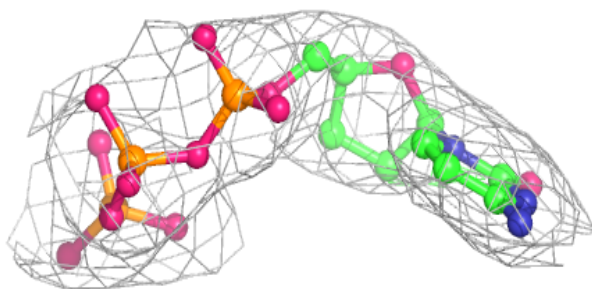
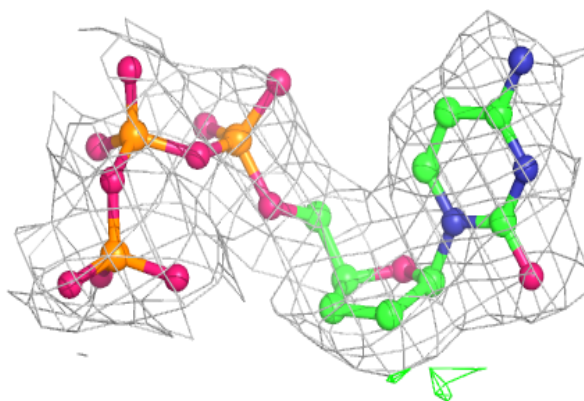
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
7	GOL	E	505	6/6	0.96	0.17	35,41,42,43	0

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.

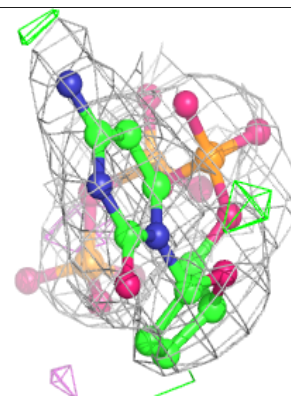
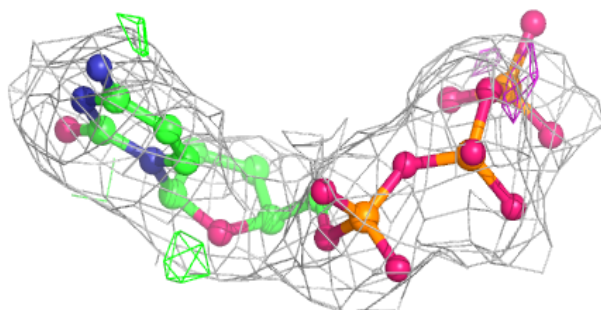
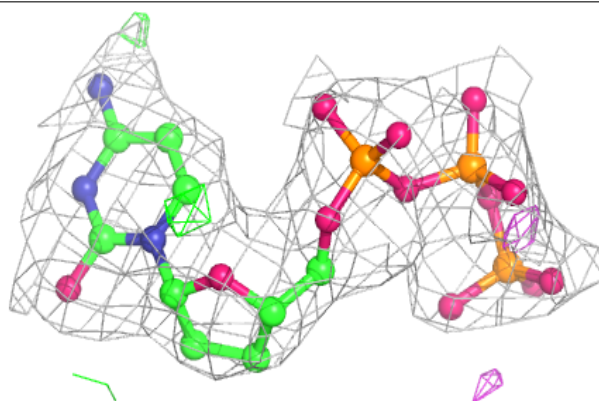


**Electron density around DCT M 501:**

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

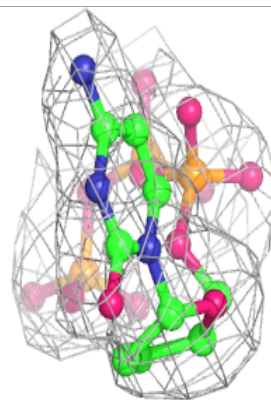
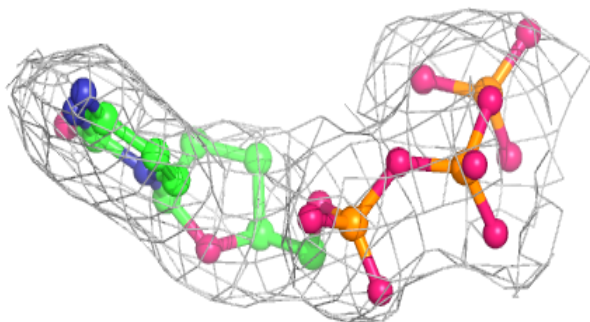
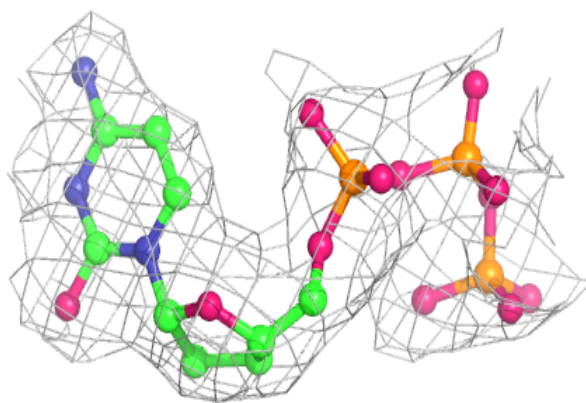
**Electron density around DCT E 501:**

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



**Electron density around DCT I 501:**

$2mF_o-DF_c$  (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.