



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 19, 2023 – 03:47 AM EDT

PDB ID : 5D4E  
Title : Crystal structure of Thermus thermophilus product complex for transcription initiation with 3'-dephosphate-CoA and CTP  
Authors : Zhang, Y.; Ebright, R.H.  
Deposited on : 2015-08-07  
Resolution : 3.08 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.35.1  
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) : 2.35.1

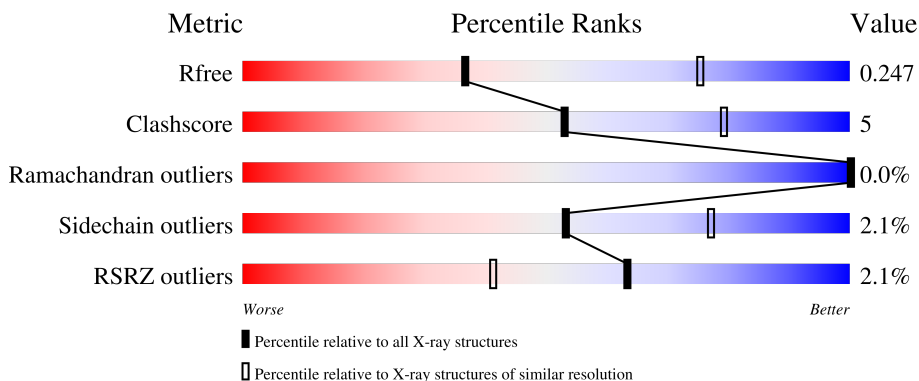
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.08 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.



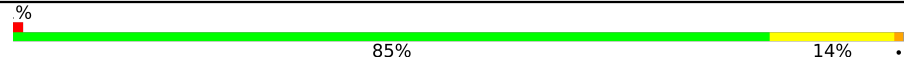

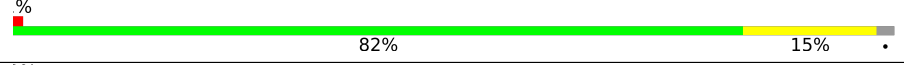


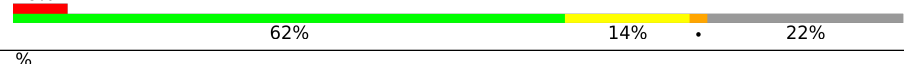
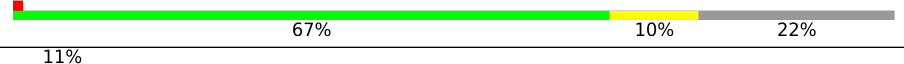

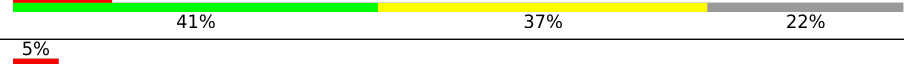

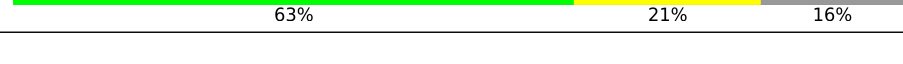
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1447 (3.10-3.06)
Clashscore	141614	1546 (3.10-3.06)
Ramachandran outliers	138981	1487 (3.10-3.06)
Sidechain outliers	138945	1486 (3.10-3.06)
RSRZ outliers	127900	1416 (3.10-3.06)

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  $\geq 3$ , 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	315	 2% 61% 12% 27%
1	B	315	 % 61% 10% 29%
1	K	315	 2% 61% 12% 27%
1	L	315	 60% 11% 29%
2	C	1119	 4% 81% 18% ..

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	M	1119	 % 85% 14%
3	D	1524	 % 81% 16%
3	N	1524	 % 82% 15%
4	E	99	 2% 82% 13% 5%
4	O	99	 % 82% 13% 5%
5	F	443	 6% 62% 14% 22%
5	P	443	 % 67% 10% 22%
6	H	27	 11% 59% 19% 22%
6	S	27	 11% 41% 37% 22%
7	G	19	 5% 74% 11% 16%
7	R	19	 11% 63% 21% 16%

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
12	COD	D	2008	-	-	-	X

## 2 Entry composition [i](#)

There are 13 unique types of molecules in this entry. The entry contains 57723 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 DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	231	Total	C	N	O	S	0	0	0
			1809	1155	315	337	2			
1	B	223	Total	C	N	O	S	0	0	0
			1758	1124	305	327	2			
1	K	231	Total	C	N	O	S	0	0	0
			1809	1155	315	337	2			
1	L	223	Total	C	N	O	S	0	0	0
			1758	1124	305	327	2			

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	1112	Total	C	N	O	S	0	0	0
			8774	5550	1565	1635	24			
2	M	1112	Total	C	N	O	S	0	1	0
			8783	5558	1565	1636	24			

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	1486	Total	C	N	O	S	0	0	0
			11738	7441	2067	2195	35			
3	N	1486	Total	C	N	O	S	0	0	0
			11738	7441	2067	2195	35			

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	94	Total	C	N	O	S	0	0	0
			758	483	132	139	4			
4	O	94	Total	C	N	O	S	0	0	0
			758	483	132	139	4			

- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	346	2807	1770	509	524	4	0	0	0
5	P	346	2807	1770	509	524	4	0	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-19	MET	-	initiating methionine	UNP Q72L95
F	-18	GLY	-	expression tag	UNP Q72L95
F	-17	SER	-	expression tag	UNP Q72L95
F	-16	SER	-	expression tag	UNP Q72L95
F	-15	HIS	-	expression tag	UNP Q72L95
F	-14	HIS	-	expression tag	UNP Q72L95
F	-13	HIS	-	expression tag	UNP Q72L95
F	-12	HIS	-	expression tag	UNP Q72L95
F	-11	HIS	-	expression tag	UNP Q72L95
F	-10	HIS	-	expression tag	UNP Q72L95
F	-9	SER	-	expression tag	UNP Q72L95
F	-8	SER	-	expression tag	UNP Q72L95
F	-7	GLY	-	expression tag	UNP Q72L95
F	-6	LEU	-	expression tag	UNP Q72L95
F	-5	VAL	-	expression tag	UNP Q72L95
F	-4	PRO	-	expression tag	UNP Q72L95
F	-3	ARG	-	expression tag	UNP Q72L95
F	-2	GLY	-	expression tag	UNP Q72L95
F	-1	SER	-	expression tag	UNP Q72L95
F	0	HIS	-	expression tag	UNP Q72L95
F	46	THR	ALA	conflict	UNP Q72L95
P	-19	MET	-	initiating methionine	UNP Q72L95
P	-18	GLY	-	expression tag	UNP Q72L95
P	-17	SER	-	expression tag	UNP Q72L95
P	-16	SER	-	expression tag	UNP Q72L95
P	-15	HIS	-	expression tag	UNP Q72L95
P	-14	HIS	-	expression tag	UNP Q72L95
P	-13	HIS	-	expression tag	UNP Q72L95
P	-12	HIS	-	expression tag	UNP Q72L95
P	-11	HIS	-	expression tag	UNP Q72L95
P	-10	HIS	-	expression tag	UNP Q72L95
P	-9	SER	-	expression tag	UNP Q72L95
P	-8	SER	-	expression tag	UNP Q72L95
P	-7	GLY	-	expression tag	UNP Q72L95

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
P	-6	LEU	-	expression tag	UNP Q72L95
P	-5	VAL	-	expression tag	UNP Q72L95
P	-4	PRO	-	expression tag	UNP Q72L95
P	-3	ARG	-	expression tag	UNP Q72L95
P	-2	GLY	-	expression tag	UNP Q72L95
P	-1	SER	-	expression tag	UNP Q72L95
P	0	HIS	-	expression tag	UNP Q72L95
P	46	THR	ALA	conflict	UNP Q72L95

- Molecule 6 is a DNA chain called DNA (27-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	H	21	Total	C	N	O	P	0	0	0
			435	207	87	121	20			
6	S	21	Total	C	N	O	P	0	0	0
			435	207	87	121	20			

- Molecule 7 is a DNA chain called DNA (5'-D(\*CP\*C\*TP\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*T  
P\*GP\*AP\*GP\*TP\*AP\*GP\*AP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	16	Total	C	N	O	P	0	0	0
			327	156	62	94	15			
7	R	16	Total	C	N	O	P	0	0	0
			328	157	62	94	15			

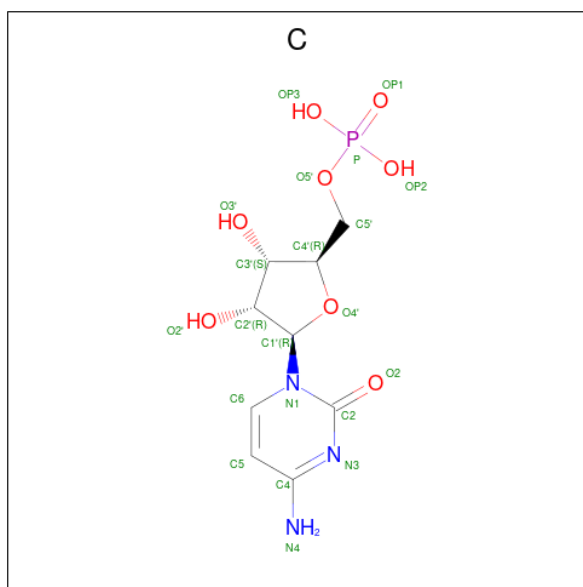
- Molecule 8 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	B	1	Total	Mg	0	0
			1	1		
8	D	3	Total	Mg	0	0
			3	3		
8	F	1	Total	Mg	0	0
			1	1		
8	N	3	Total	Mg	0	0
			3	3		
8	P	1	Total	Mg	0	0
			1	1		

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

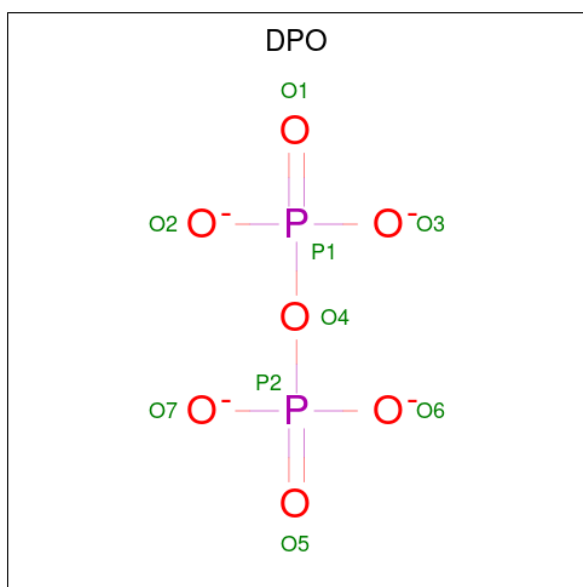
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	2	Total	Zn	0	0
			2	2		
9	N	2	Total	Zn	0	0
			2	2		

- Molecule 10 is CYTIDINE-5'-MONOPHOSPHATE (three-letter code: C) (formula:  $C_9H_{14}N_3O_8P$ ).



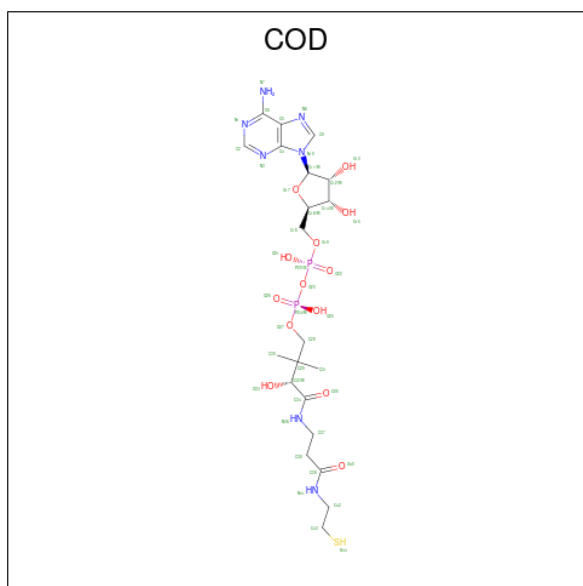
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
10	D	1	Total	C	N	O	P	0	0
			20	9	3	7	1		
10	N	1	Total	C	N	O	P	0	0
			20	9	3	7	1		

- Molecule 11 is DIPHOSPHATE (three-letter code: DPO) (formula:  $O_7P_2$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
11	D	1	Total	O	P			
			9	7	2	0	0	
11	N	1	Total	O	P			
			9	7	2	0	0	

- Molecule 12 is DEPHOSPHO COENZYME A (three-letter code: COD) (formula:  $C_{21}H_{35}N_7O_{13}P_2S$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
12	D	1	Total	C	N	O	P		
			28	11	5	10	2	0	0
12	N	1	Total	C	N	O	P		
			27	10	5	10	2	0	0



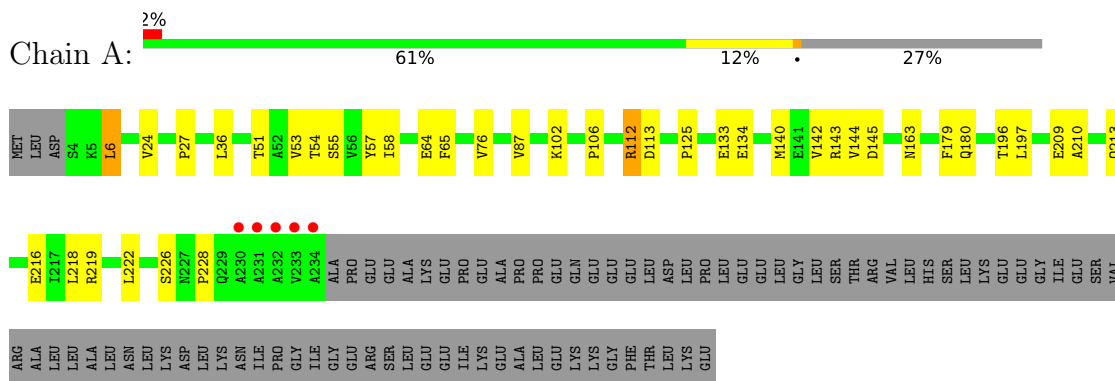
- Molecule 13 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
13	A	17	Total O 17 17	0	0
13	B	10	Total O 10 10	0	0
13	C	74	Total O 74 74	0	0
13	D	144	Total O 144 144	0	0
13	E	8	Total O 8 8	0	0
13	F	17	Total O 17 17	0	0
13	H	6	Total O 6 6	0	0
13	G	3	Total O 3 3	0	0
13	K	16	Total O 16 16	0	0
13	L	14	Total O 14 14	0	0
13	M	178	Total O 178 178	0	0
13	N	217	Total O 217 217	0	0
13	O	16	Total O 16 16	0	0
13	P	41	Total O 41 41	0	0
13	R	9	Total O 9 9	0	0
13	S	5	Total O 5 5	0	0

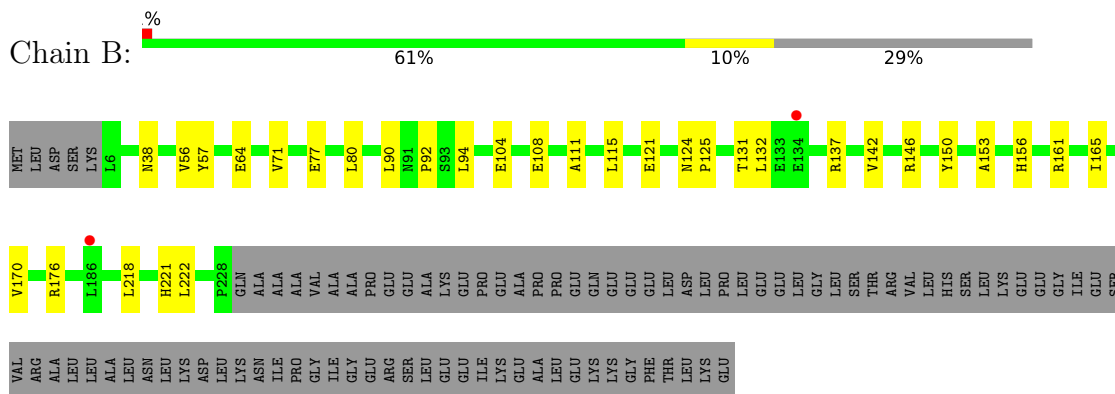
### 3 Residue-property plots [i](#)

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.

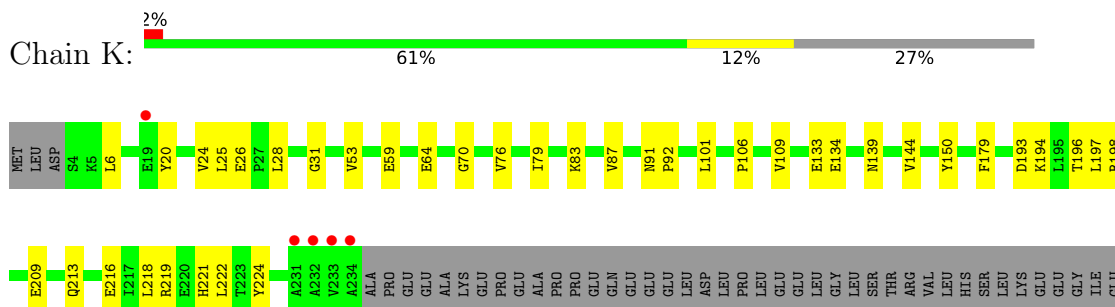
- Molecule 1: DNA-directed RNA polymerase subunit alpha



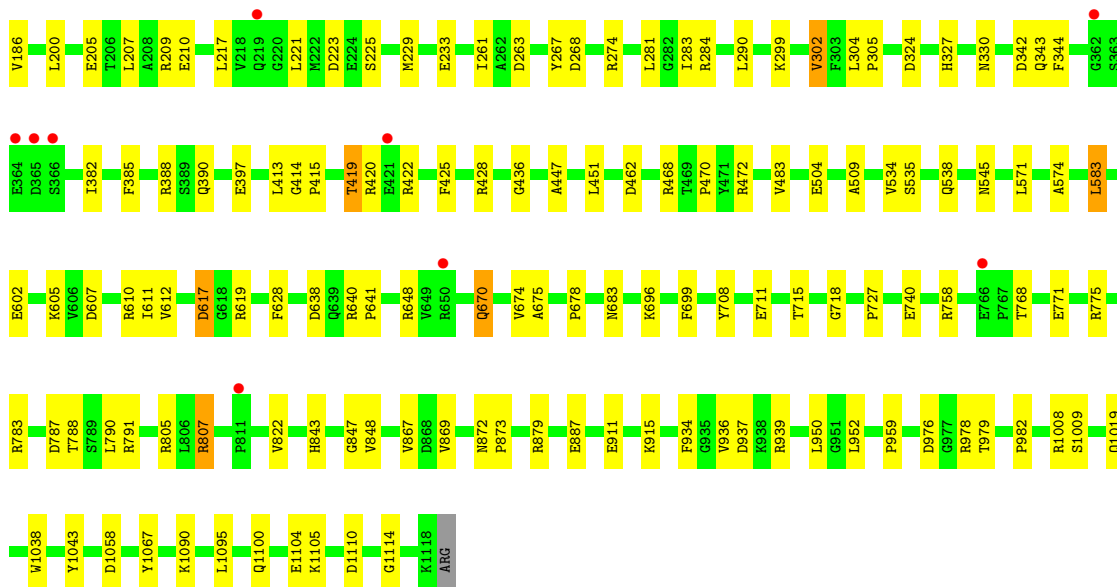
- Molecule 1: DNA-directed RNA polymerase subunit alpha



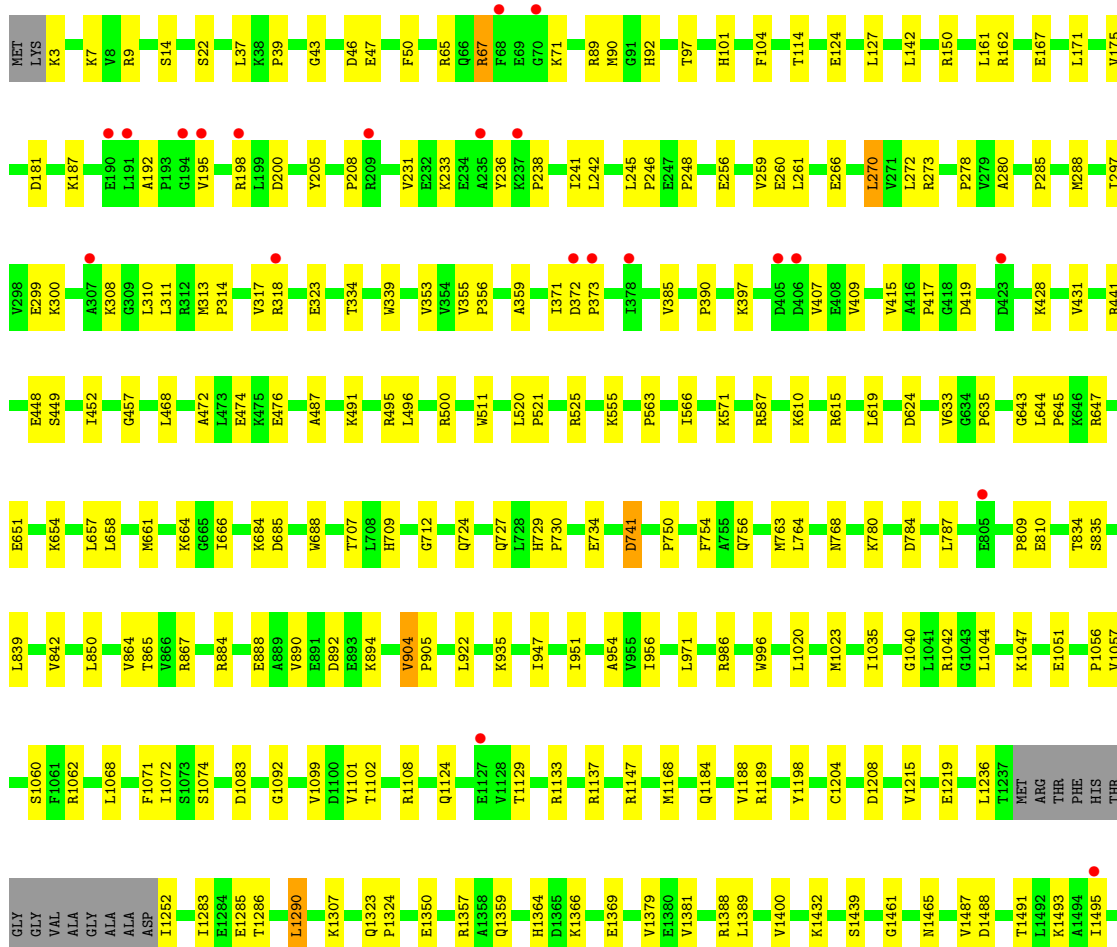
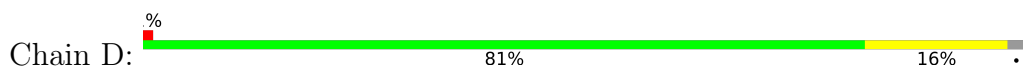
- Molecule 1: DNA-directed RNA polymerase subunit alpha

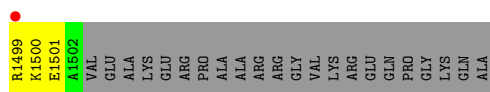




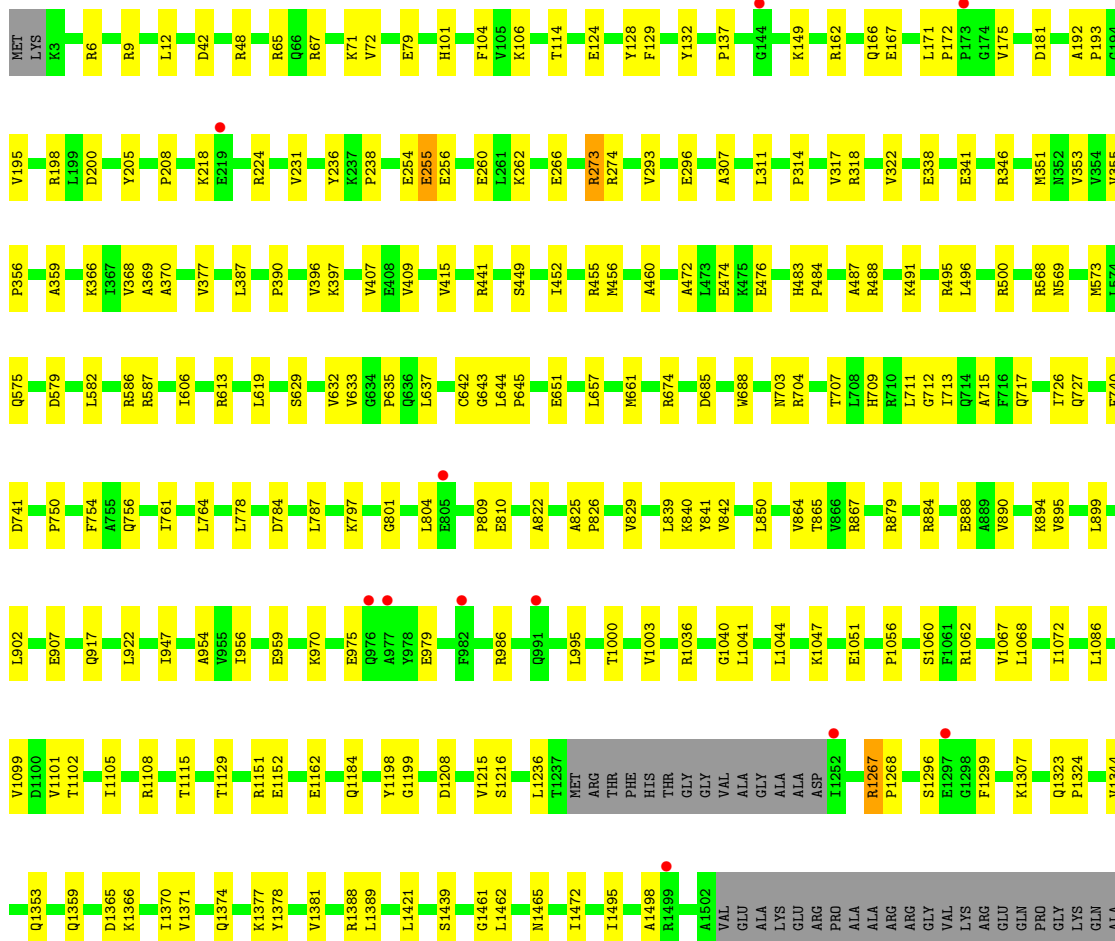
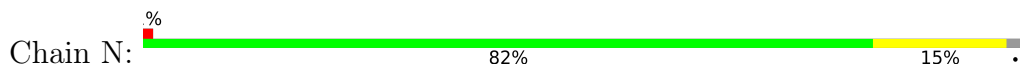


• Molecule 3: DNA-directed RNA polymerase subunit beta'

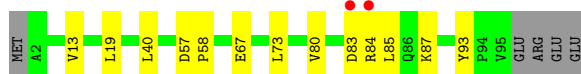
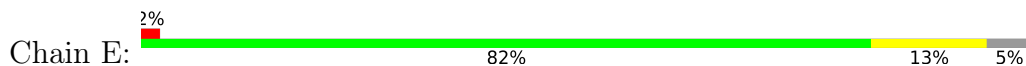




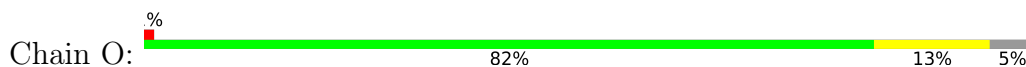
- Molecule 3: DNA-directed RNA polymerase subunit beta'



- Molecule 4: DNA-directed RNA polymerase subunit omega

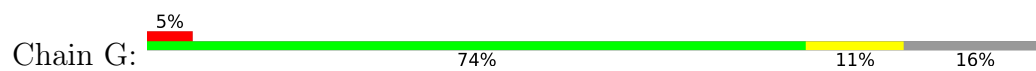


- Molecule 4: DNA-directed RNA polymerase subunit omega





- Molecule 7: DNA (5'-D(\*CP\*C\*TP\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*AP\*GP\*AP\*G)-3')



- Molecule 7: DNA (5'-D(\*CP\*C\*TP\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*AP\*GP\*AP\*G)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	185.81Å 103.95Å 297.14Å 90.00° 98.47° 90.00°	Depositor
Resolution (Å)	39.68 – 3.08 39.68 – 3.08	Depositor EDS
% Data completeness (in resolution range)	89.9 (39.68-3.08) 90.0 (39.68-3.08)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	0.12	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.26 (at 3.06Å)	Xtrriage
Refinement program	PHENIX 1.8_1069	Depositor
R, $R_{free}$	0.210 , 0.246 0.210 , 0.247	Depositor DCC
$R_{free}$ test set	9379 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	62.5	Xtrriage
Anisotropy	0.029	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 38.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	57723	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 46.93 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.0674e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: DPO, COD, ZN, MG

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	0.21	0/1841	0.41	0/2504
1	B	0.26	0/1790	0.45	0/2435
1	K	0.22	0/1841	0.42	0/2504
1	L	0.26	0/1790	0.46	1/2435 (0.0%)
2	C	0.22	0/8941	0.41	0/12092
2	M	0.23	0/8954	0.42	0/12110
3	D	0.21	0/11944	0.40	0/16149
3	N	0.22	0/11944	0.41	0/16149
4	E	0.21	0/772	0.37	0/1040
4	O	0.22	0/772	0.39	0/1040
5	F	0.22	0/2852	0.39	0/3837
5	P	0.21	0/2852	0.38	0/3837
6	H	0.45	0/489	0.99	0/752
6	S	0.45	0/489	1.02	1/752 (0.1%)
7	G	0.44	0/367	0.93	0/565
7	R	0.44	0/368	0.99	1/567 (0.2%)
All	All	0.23	0/58006	0.44	3/78768 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	S	1	DT	O4'-C1'-N1	5.44	111.81	108.00
7	R	16	DT	O4'-C4'-C3'	-5.28	102.39	104.50
1	L	91	ASN	C-N-CD	5.11	139.13	128.40

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1809	0	1863	24	0
1	B	1758	0	1808	22	0
1	K	1809	0	1863	25	0
1	L	1758	0	1808	21	0
2	C	8774	0	8877	121	0
2	M	8783	0	8886	94	0
3	D	11738	0	11971	136	0
3	N	11738	0	11971	133	0
4	E	758	0	770	6	0
4	O	758	0	770	9	0
5	F	2807	0	2882	42	0
5	P	2807	0	2882	29	0
6	H	435	0	238	4	0
6	S	435	0	238	6	0
7	G	327	0	179	2	0
7	R	328	0	182	2	0
8	B	1	0	0	0	0
8	D	3	0	0	0	0
8	F	1	0	0	0	0
8	N	3	0	0	0	0
8	P	1	0	0	0	0
9	D	2	0	0	0	0
9	N	2	0	0	0	0
10	D	20	0	11	0	0
10	N	20	0	11	2	0
11	D	9	0	0	0	0
11	N	9	0	0	0	0
12	D	28	0	11	0	0
12	N	27	0	11	0	0
13	A	17	0	0	0	0
13	B	10	0	0	0	0
13	C	74	0	0	2	0
13	D	144	0	0	0	0
13	E	8	0	0	0	0
13	F	17	0	0	0	0
13	G	3	0	0	0	0
13	H	6	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
13	K	16	0	0	0	0
13	L	14	0	0	0	0
13	M	178	0	0	3	0
13	N	217	0	0	2	0
13	O	16	0	0	0	0
13	P	41	0	0	0	0
13	R	9	0	0	0	0
13	S	5	0	0	0	0
All	All	57723	0	57232	614	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 5.

The worst 5 of 614 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:C:628:PHE:H	2:C:638:ASP:HB3	1.40	0.86
2:M:168:ARG:HD3	2:M:268:ASP:HB3	1.64	0.79
2:M:165:LEU:HB2	2:M:168:ARG:HG3	1.63	0.78
2:C:802:ARG:HB2	2:C:826:TYR:HB2	1.66	0.75
2:C:165:LEU:HB2	2:C:168:ARG:HG3	1.68	0.74

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	229/315 (73%)	224 (98%)	5 (2%)	0	100   100
1	B	221/315 (70%)	216 (98%)	5 (2%)	0	100   100
1	K	229/315 (73%)	225 (98%)	4 (2%)	0	100   100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	L	221/315 (70%)	217 (98%)	3 (1%)	1 (0%)	29	61
2	C	1108/1119 (99%)	1075 (97%)	33 (3%)	0	100	100
2	M	1109/1119 (99%)	1080 (97%)	29 (3%)	0	100	100
3	D	1482/1524 (97%)	1445 (98%)	37 (2%)	0	100	100
3	N	1482/1524 (97%)	1447 (98%)	35 (2%)	0	100	100
4	E	92/99 (93%)	90 (98%)	2 (2%)	0	100	100
4	O	92/99 (93%)	90 (98%)	2 (2%)	0	100	100
5	F	344/443 (78%)	334 (97%)	9 (3%)	1 (0%)	41	71
5	P	344/443 (78%)	339 (98%)	5 (2%)	0	100	100
All	All	6953/7630 (91%)	6782 (98%)	169 (2%)	2 (0%)	100	100

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	F	376	ILE
1	L	8	ALA

### 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	200/273 (73%)	196 (98%)	4 (2%)	55	78
1	B	196/273 (72%)	195 (100%)	1 (0%)	88	94
1	K	200/273 (73%)	199 (100%)	1 (0%)	88	94
1	L	196/273 (72%)	194 (99%)	2 (1%)	76	89
2	C	936/941 (100%)	919 (98%)	17 (2%)	59	80
2	M	937/941 (100%)	916 (98%)	21 (2%)	52	76
3	D	1253/1279 (98%)	1223 (98%)	30 (2%)	49	74
3	N	1253/1279 (98%)	1232 (98%)	21 (2%)	60	82
4	E	82/88 (93%)	79 (96%)	3 (4%)	34	65

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	O	82/88 (93%)	81 (99%)	1 (1%)	71	87
5	F	301/388 (78%)	285 (95%)	16 (5%)	22	53
5	P	301/388 (78%)	294 (98%)	7 (2%)	50	75
All	All	5937/6484 (92%)	5813 (98%)	124 (2%)	53	77

5 of 124 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	F	97	GLU
3	N	986	ARG
1	K	6	LEU
3	N	907	GLU
5	P	295	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1172	HIS
4	O	33	HIS
5	F	83	GLN
5	P	83	GLN
3	N	1254	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

### 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry

Of 19 ligands modelled in this entry, 13 are monoatomic - leaving 6 for Mogul analysis.

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
10	C	D	2005	8,12	18,21,22	0.34	0	26,30,33	0.38	0
10	C	N	1605	8,12	18,21,22	0.33	0	26,30,33	0.37	0
11	DPO	N	1606	8	6,8,8	0.98	0	13,13,13	1.07	1 (7%)
11	DPO	D	2006	8	6,8,8	0.96	0	13,13,13	1.16	1 (7%)
12	COD	N	1608	10	24,29,46	1.77	5 (20%)	29,45,68	1.68	6 (20%)
12	COD	D	2008	10	25,30,46	1.73	7 (28%)	28,46,68	1.80	6 (21%)

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
10	C	D	2005	8,12	-	2/7/25/26	0/2/2/2
10	C	N	1605	8,12	-	0/7/25/26	0/2/2/2
11	DPO	N	1606	8	-	0/6/6/6	-
11	DPO	D	2006	8	-	0/6/6/6	-
12	COD	N	1608	10	-	6/12/32/59	0/3/3/3
12	COD	D	2008	10	-	3/15/35/59	0/3/3/3

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
12	D	2008	COD	C2-N3	4.67	1.39	1.32
12	N	1608	COD	C2-N3	4.53	1.39	1.32
12	N	1608	COD	C18-C16	3.09	1.61	1.51
12	D	2008	COD	C18-C16	3.04	1.61	1.51
12	N	1608	COD	C2-N1	2.70	1.38	1.33

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
12	N	1608	COD	N3-C2-N1	-5.19	120.57	128.68
12	D	2008	COD	N3-C2-N1	-5.16	120.61	128.68
12	D	2008	COD	C14-C12-C11	4.58	107.87	100.98
12	N	1608	COD	C14-C12-C11	3.74	106.61	100.98
12	N	1608	COD	P20-O23-P24	-3.68	120.20	132.83

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

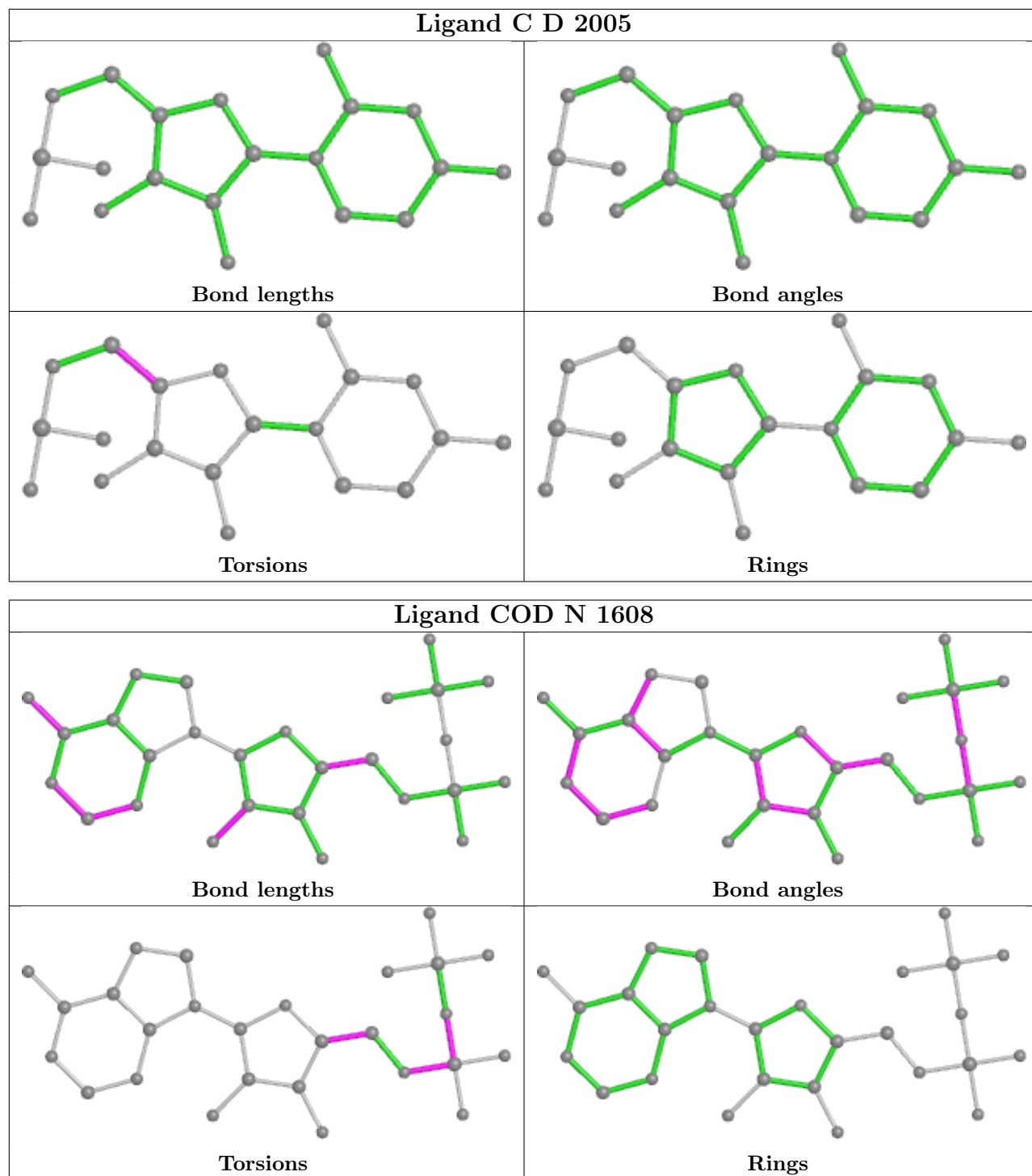
Mol	Chain	Res	Type	Atoms
10	D	2005	C	O4'-C4'-C5'-O5'
10	D	2005	C	C3'-C4'-C5'-O5'
12	N	1608	COD	C18-O19-P20-O21
12	N	1608	COD	C18-O19-P20-O22
12	N	1608	COD	C18-O19-P20-O23

There are no ring outliers.

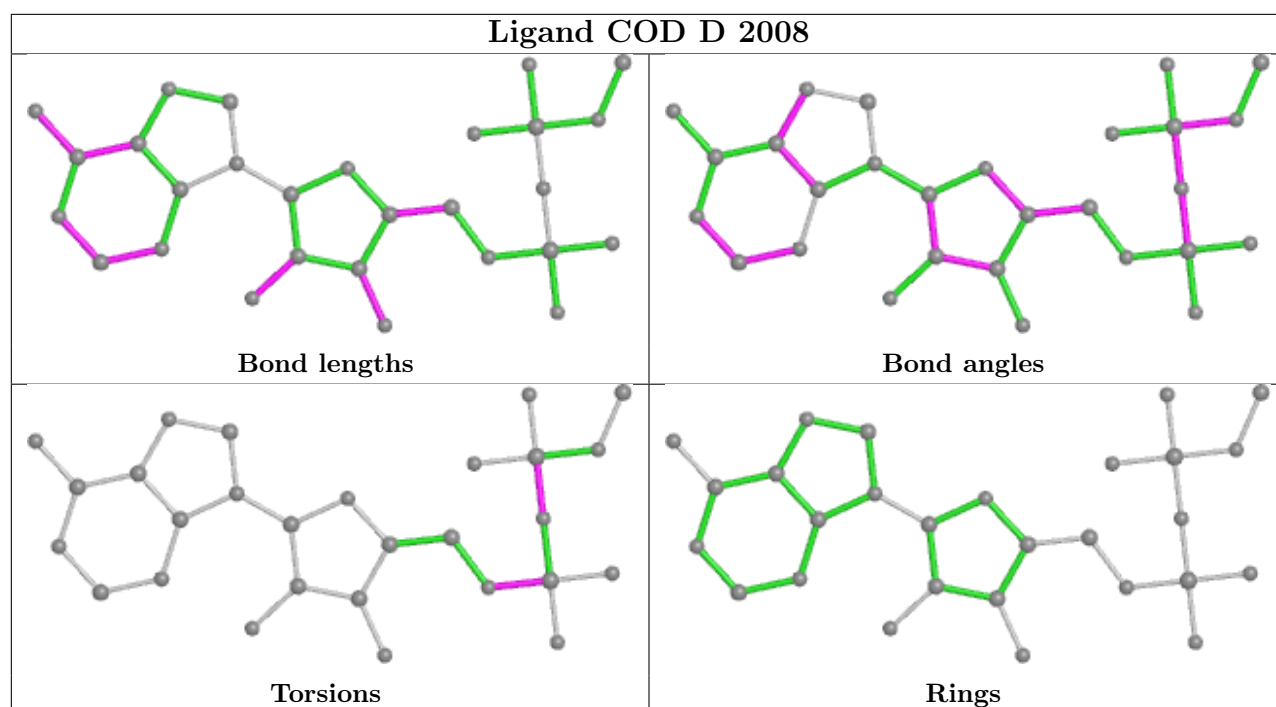
1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
10	N	1605	C	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.







## 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, 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	231/315 (73%)	-0.17	5 (2%) 62 39	29, 51, 78, 126	1 (0%)
1	B	223/315 (70%)	-0.28	2 (0%) 84 68	26, 53, 93, 122	0
1	K	231/315 (73%)	-0.39	5 (2%) 62 39	22, 38, 65, 128	1 (0%)
1	L	223/315 (70%)	-0.33	0 100 100	21, 49, 82, 102	0
2	C	1112/1119 (99%)	0.07	49 (4%) 34 17	12, 58, 115, 132	3 (0%)
2	M	1112/1119 (99%)	-0.35	11 (0%) 82 66	5, 32, 82, 118	3 (0%)
3	D	1486/1524 (97%)	-0.17	22 (1%) 73 53	7, 43, 96, 131	4 (0%)
3	N	1486/1524 (97%)	-0.29	11 (0%) 87 74	4, 34, 85, 125	4 (0%)
4	E	94/99 (94%)	-0.24	2 (2%) 63 41	22, 44, 88, 102	0
4	O	94/99 (94%)	-0.44	1 (1%) 80 62	14, 31, 67, 85	0
5	F	346/443 (78%)	0.26	27 (7%) 13 4	33, 70, 134, 152	0
5	P	346/443 (78%)	-0.20	5 (1%) 75 55	14, 43, 93, 112	0
6	H	21/27 (77%)	0.56	3 (14%) 2 1	47, 92, 134, 155	0
6	S	21/27 (77%)	0.30	3 (14%) 2 1	35, 73, 118, 146	0
7	G	16/19 (84%)	0.55	1 (6%) 20 8	49, 77, 129, 134	0
7	R	16/19 (84%)	0.54	2 (12%) 3 1	31, 60, 131, 137	0
All	All	7058/7722 (91%)	-0.18	149 (2%) 63 41	4, 45, 99, 155	16 (0%)

The worst 5 of 149 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	63	GLY	9.3
5	F	391	GLY	7.3
1	A	233	VAL	7.2
5	F	390	PHE	5.8
1	K	233	VAL	5.8

## 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](#)

There are no monosaccharides in this entry.

## 6.4 Ligands [i](#)

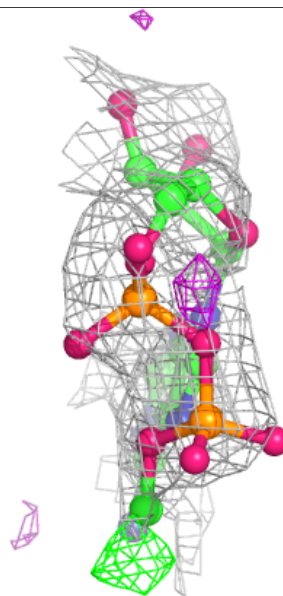
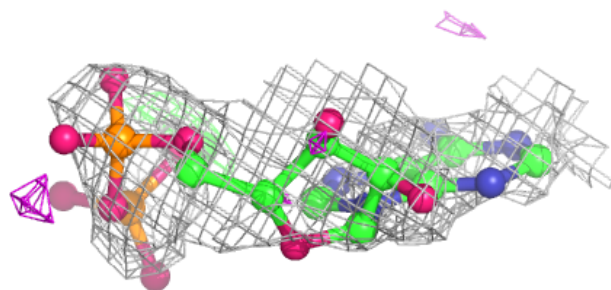
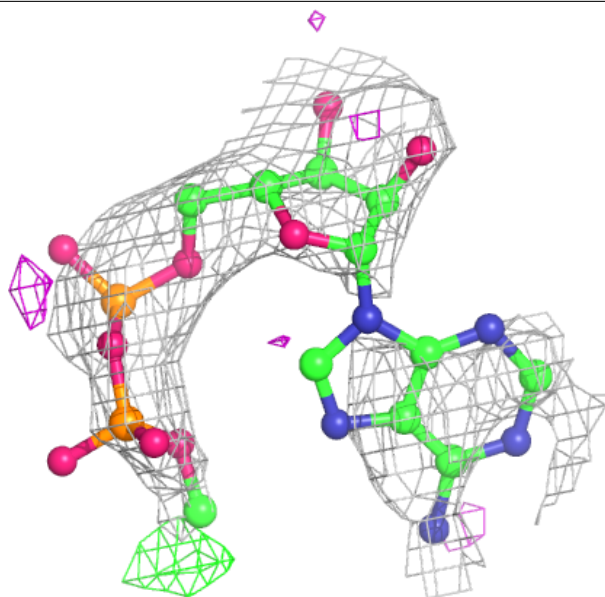
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
8	MG	B	2001	1/1	0.74	0.29	71,71,71,71	0
12	COD	D	2008	28/44	0.77	0.40	50,73,102,116	28
8	MG	F	2001	1/1	0.81	0.18	92,92,92,92	0
8	MG	P	2001	1/1	0.87	0.13	38,38,38,38	0
10	C	D	2005	20/21	0.88	0.25	43,61,71,80	20
12	COD	N	1608	27/44	0.89	0.23	35,43,62,65	27
10	C	N	1605	20/21	0.90	0.20	24,34,49,57	20
11	DPO	N	1606	9/9	0.90	0.19	26,51,74,84	0
8	MG	D	2007	1/1	0.92	0.17	49,49,49,49	0
11	DPO	D	2006	9/9	0.93	0.21	45,68,84,98	0
8	MG	N	1607	1/1	0.95	0.17	36,36,36,36	0
8	MG	D	2003	1/1	0.96	0.15	22,22,22,22	0
8	MG	D	2004	1/1	0.97	0.43	27,27,27,27	0
8	MG	N	1601	1/1	0.97	0.44	25,25,25,25	0
9	ZN	D	2002	1/1	0.97	0.04	92,92,92,92	0
8	MG	N	1604	1/1	0.99	0.19	13,13,13,13	0
9	ZN	N	1602	1/1	0.99	0.13	16,16,16,16	0
9	ZN	N	1603	1/1	0.99	0.05	50,50,50,50	0
9	ZN	D	2001	1/1	1.00	0.13	20,20,20,20	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.

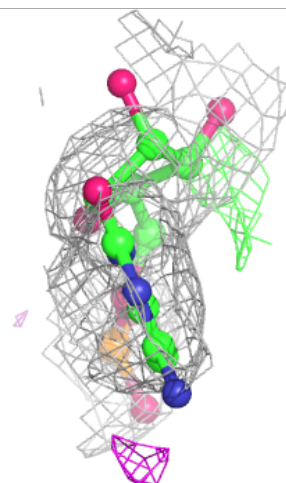
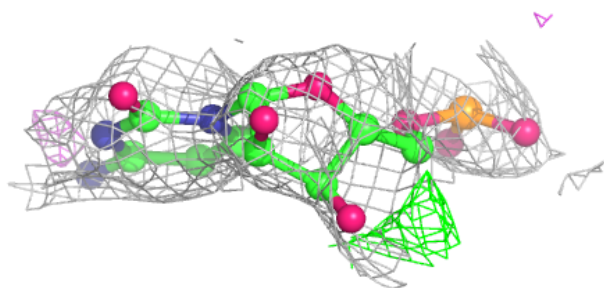
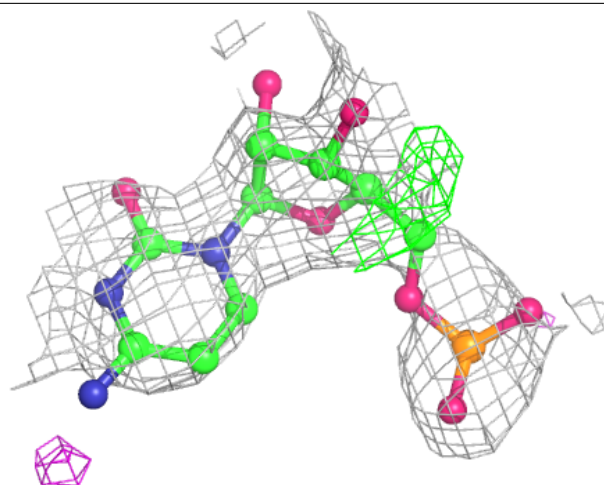
**Electron density around COD D 2008:**

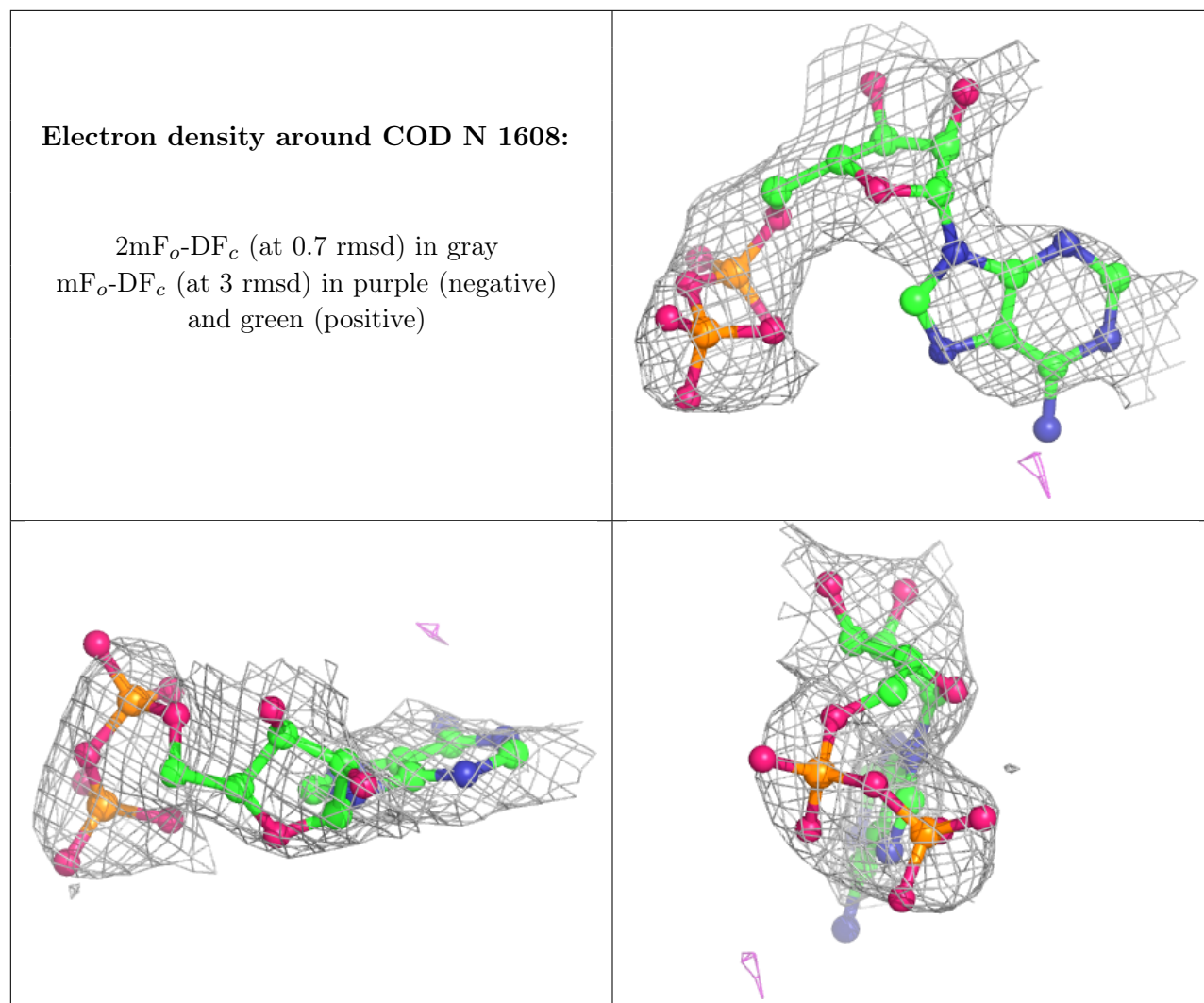
$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)



**Electron density around C D 2005:**

$2mF_o-DF_c$  (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.