



# Full wwPDB X-ray Structure Validation Report ⓘ

May 19, 2022 – 03:11 pm BST

PDB ID : 7Z43  
Title : Influenza B polymerase with Pol II pSer5 CTD peptide mimic bound in site 1B and 2B  
Authors : Cusack, S.; Drncova, P.  
Deposited on : 2022-03-03  
Resolution : 3.12 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.28.1  
buster-report : 1.1.7 (2018)  
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) : 2.28.1

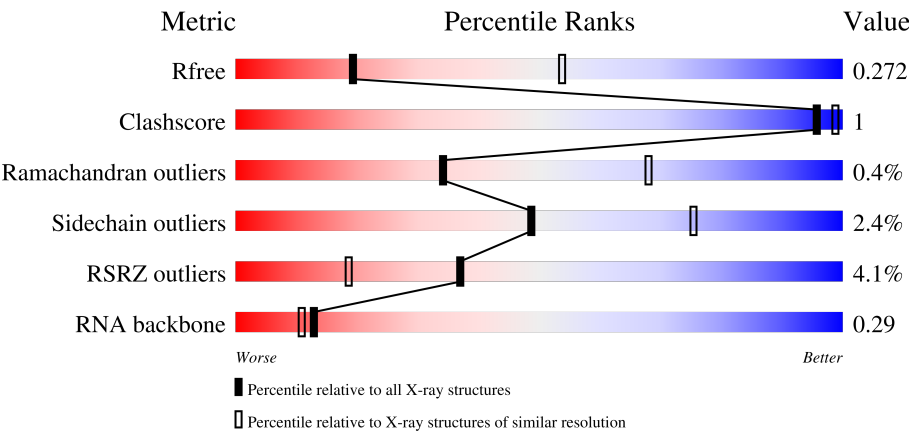
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.12 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 <sub>free</sub>	130704	1292 (3.14-3.10)
Clashscore	141614	1389 (3.14-3.10)
Ramachandran outliers	138981	1337 (3.14-3.10)
Sidechain outliers	138945	1337 (3.14-3.10)
RSRZ outliers	127900	1260 (3.14-3.10)
RNA backbone	3102	1134 (3.44-2.80)

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	AAA	751	<div><div>7%</div><div><div></div><div>91%</div><div>• 5%</div></div></div>
2	BBB	772	<div><div></div><div><div>92%</div><div>5% •</div></div></div>
3	CCC	798	<div><div>4%</div><div><div></div><div>86%</div><div>5% 8%</div></div></div>
4	RRR	18	<div><div></div><div><div>28%</div><div>61%</div><div>11%</div></div></div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
5	VVV	14	 29% 50% 21%
6	MaM	13	 23% 15% 62%
7	XXX	28	 11% 25% 75%
7	YYY	28	 57% • 39%

## 2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 18464 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 Polymerase acidic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	717	Total	C	N	O	S	0	0	0
			5755	3657	963	1095	40			

There are 25 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	-13	GLY	-	expression tag	UNP Q5V8Z9
AAA	-12	SER	-	expression tag	UNP Q5V8Z9
AAA	-11	HIS	-	expression tag	UNP Q5V8Z9
AAA	-10	HIS	-	expression tag	UNP Q5V8Z9
AAA	-9	HIS	-	expression tag	UNP Q5V8Z9
AAA	-8	HIS	-	expression tag	UNP Q5V8Z9
AAA	-7	HIS	-	expression tag	UNP Q5V8Z9
AAA	-6	HIS	-	expression tag	UNP Q5V8Z9
AAA	-5	HIS	-	expression tag	UNP Q5V8Z9
AAA	-4	HIS	-	expression tag	UNP Q5V8Z9
AAA	-3	GLY	-	expression tag	UNP Q5V8Z9
AAA	-2	SER	-	expression tag	UNP Q5V8Z9
AAA	-1	GLY	-	expression tag	UNP Q5V8Z9
AAA	0	SER	-	expression tag	UNP Q5V8Z9
AAA	727	GLY	-	expression tag	UNP Q5V8Z9
AAA	728	SER	-	expression tag	UNP Q5V8Z9
AAA	729	GLY	-	expression tag	UNP Q5V8Z9
AAA	730	SER	-	expression tag	UNP Q5V8Z9
AAA	731	GLY	-	expression tag	UNP Q5V8Z9
AAA	732	GLU	-	expression tag	UNP Q5V8Z9
AAA	733	ASN	-	expression tag	UNP Q5V8Z9
AAA	734	LEU	-	expression tag	UNP Q5V8Z9
AAA	735	TYR	-	expression tag	UNP Q5V8Z9
AAA	736	PHE	-	expression tag	UNP Q5V8Z9
AAA	737	GLN	-	expression tag	UNP Q5V8Z9

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	BBB	750	Total	C	N	O	S	0	0	0
			5881	3711	1020	1098	52			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
BBB	-8	GLY	-	expression tag	UNP Q5V8Y6
BBB	-7	SER	-	expression tag	UNP Q5V8Y6
BBB	-6	GLY	-	expression tag	UNP Q5V8Y6
BBB	-5	SER	-	expression tag	UNP Q5V8Y6
BBB	-4	GLY	-	expression tag	UNP Q5V8Y6
BBB	-3	SER	-	expression tag	UNP Q5V8Y6
BBB	-2	GLY	-	expression tag	UNP Q5V8Y6
BBB	-1	SER	-	expression tag	UNP Q5V8Y6
BBB	0	GLY	-	expression tag	UNP Q5V8Y6
BBB	753	GLY	-	expression tag	UNP Q5V8Y6
BBB	754	SER	-	expression tag	UNP Q5V8Y6
BBB	755	GLY	-	expression tag	UNP Q5V8Y6
BBB	756	SER	-	expression tag	UNP Q5V8Y6
BBB	757	GLY	-	expression tag	UNP Q5V8Y6
BBB	758	GLU	-	expression tag	UNP Q5V8Y6
BBB	759	ASN	-	expression tag	UNP Q5V8Y6
BBB	760	LEU	-	expression tag	UNP Q5V8Y6
BBB	761	TYR	-	expression tag	UNP Q5V8Y6
BBB	762	PHE	-	expression tag	UNP Q5V8Y6
BBB	763	GLN	-	expression tag	UNP Q5V8Y6

- Molecule 3 is a protein called Polymerase basic protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	CCC	732	Total	C	N	O	S	0	0	0
			5858	3723	1028	1067	40			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
CCC	-8	GLY	-	expression tag	UNP Q5V8X3
CCC	-7	SER	-	expression tag	UNP Q5V8X3
CCC	-6	GLY	-	expression tag	UNP Q5V8X3
CCC	-5	SER	-	expression tag	UNP Q5V8X3
CCC	-4	GLY	-	expression tag	UNP Q5V8X3
CCC	-3	SER	-	expression tag	UNP Q5V8X3
CCC	-2	GLY	-	expression tag	UNP Q5V8X3

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
CCC	-1	SER	-	expression tag	UNP Q5V8X3
CCC	0	GLY	-	expression tag	UNP Q5V8X3
CCC	771	GLY	-	expression tag	UNP Q5V8X3
CCC	772	TRP	-	expression tag	UNP Q5V8X3
CCC	773	SER	-	expression tag	UNP Q5V8X3
CCC	774	HIS	-	expression tag	UNP Q5V8X3
CCC	775	PRO	-	expression tag	UNP Q5V8X3
CCC	776	GLN	-	expression tag	UNP Q5V8X3
CCC	777	PHE	-	expression tag	UNP Q5V8X3
CCC	778	GLU	-	expression tag	UNP Q5V8X3
CCC	779	LYS	-	expression tag	UNP Q5V8X3
CCC	780	GLY	-	expression tag	UNP Q5V8X3
CCC	781	SER	-	expression tag	UNP Q5V8X3
CCC	782	GLY	-	expression tag	UNP Q5V8X3
CCC	783	SER	-	expression tag	UNP Q5V8X3
CCC	784	GLU	-	expression tag	UNP Q5V8X3
CCC	785	ASN	-	expression tag	UNP Q5V8X3
CCC	786	LEU	-	expression tag	UNP Q5V8X3
CCC	787	TYR	-	expression tag	UNP Q5V8X3
CCC	788	PHE	-	expression tag	UNP Q5V8X3
CCC	789	GLN	-	expression tag	UNP Q5V8X3

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	RRR	18	Total	C	N	O	P	0	0	0
			349	156	52	124	17			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	VVV	14	Total	C	N	O	P	0	0	0
			307	137	62	94	14			

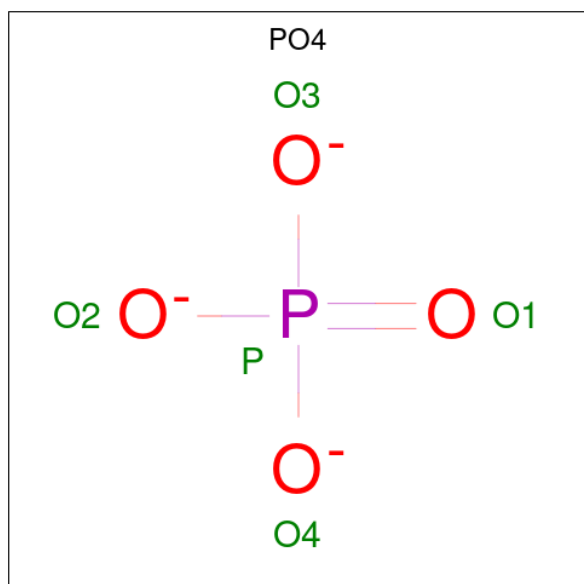
- Molecule 6 is a RNA chain called RNA (5'-R(P\*AP\*AP\*UP\*CP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	MaM	5	Total	C	N	O	P	0	0	0
			89	39	15	30	5			

- Molecule 7 is a protein called SER-TYR-SER-PRO-THR-SEP-PRO.

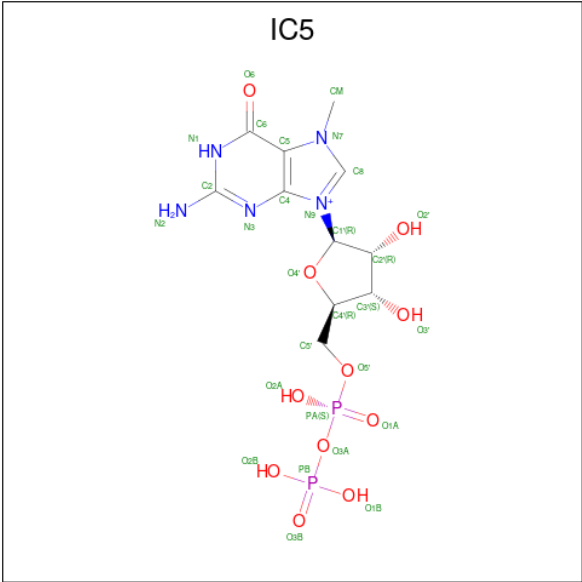
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	XXX	7	Total	C	N	O	P	0	0	0
			55	32	7	15	1			
7	YYY	17	Total	C	N	O	P	0	0	0
			135	81	17	35	2			

- Molecule 8 is PHOSPHATE ION (three-letter code: PO4) (formula: O<sub>4</sub>P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	BBB	1	Total	O	P	0	0
			5	4	1		

- Molecule 9 is [(2 {R},3 {S},4 {R},5 {R})-5-(2-azanyl-7-methyl-6-oxidanylidene-1 {H}-purin-9-yl)-3,4-bis(oxidanyl)oxolan-2-yl]methyl phosphono hydrogen phosphate (three-letter code: IC5) (formula: C<sub>11</sub>H<sub>18</sub>N<sub>5</sub>O<sub>11</sub>P<sub>2</sub>) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
9	MaM	1	Total	C	N	O	P	0	0
			29	11	5	11	2		

- Molecule 10 is water.

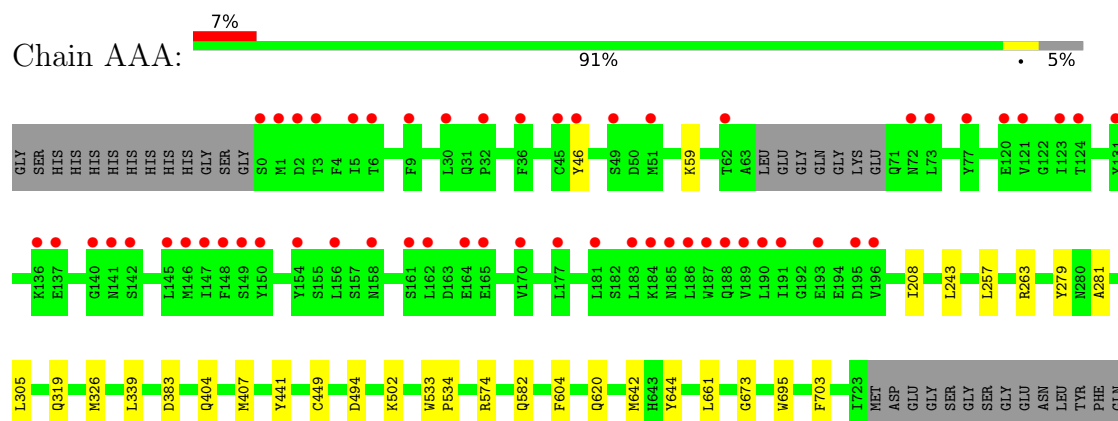
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	YYY	1	Total	O	0	0
			1	1		



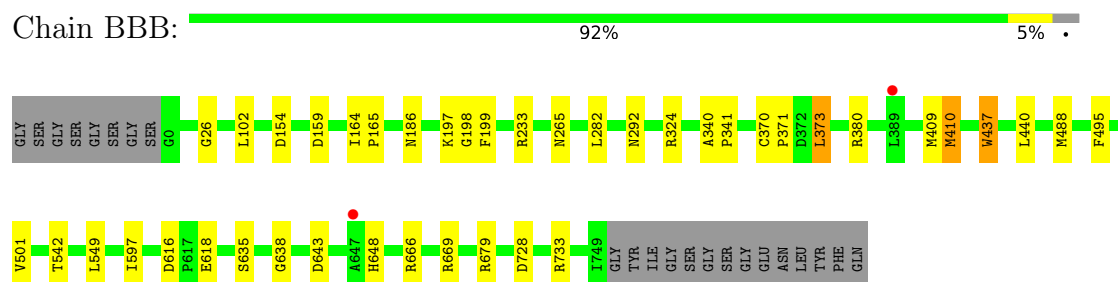
### 3 Residue-property plots [i](#)

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 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.

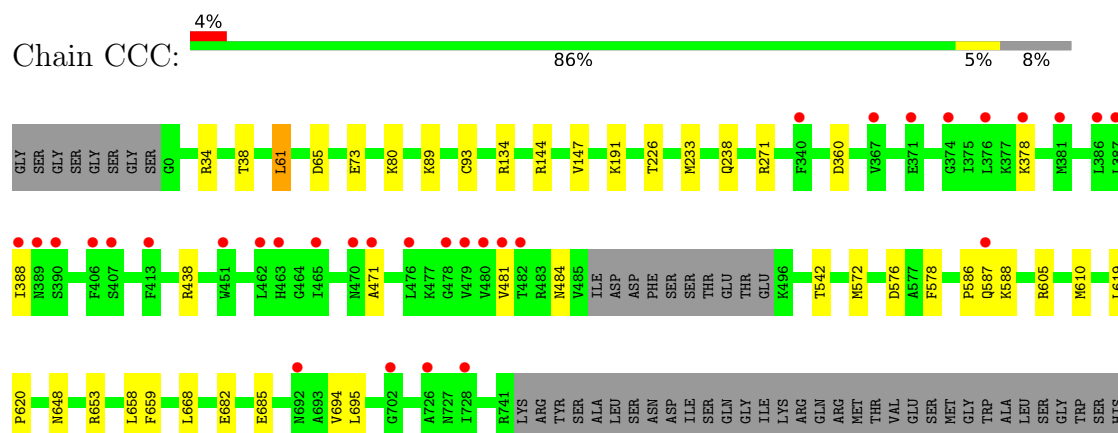
- Molecule 1: Polymerase acidic protein



- Molecule 2: RNA-directed RNA polymerase catalytic subunit

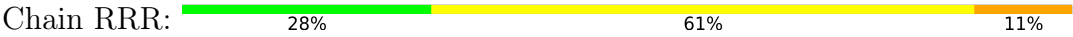


- Molecule 3: Polymerase basic protein 2



PRO GLN PHE GLU LYS SER GLY SER GLN ASN LEU TYR PHE GLN

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



U1 A2 U3 A4 C5 U6 U7 C8 U9 G10 C11 U12 U13 C14 U15 G16 C17 U18

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



A1 G2 U3 A4 G5 U6 A7 A8 A11 G12 A13 G14

• Molecule 6: RNA (5'-R(P\*AP\*AP\*UP\*CP\*A)-3')



A1 A2 U3 C4 U5 A U A A U A G C

• Molecule 7: SER-TYR-SER-PRO-THR-SEP-PRO



TYR SER PRO THR SEP PRO S21 Y22 S23 P24 T25 S26 P27 SER TYR SER PRO THR SEP PRO TYR SER PRO THR SEP PRO TYR SER PRO THR SEP PRO

• Molecule 7: SER-TYR-SER-PRO-THR-SEP-PRO



Y31 T41 F47 THR SEP PRO SER TYR SER PRO THR SEP PRO SER

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	200.82Å 200.82Å 257.02Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	76.97 – 3.12 76.85 – 3.12	Depositor EDS
% Data completeness (in resolution range)	73.1 (76.97-3.12) 73.1 (76.85-3.12)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.85 (at 3.13Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.240 , 0.272 0.240 , 0.272	Depositor DCC
$R_{free}$ test set	3951 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	98.3	Xtriage
Anisotropy	0.003	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.007 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	18464	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	116.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.24% 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 [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: PO4, IC5, SEP

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	AAA	0.67	0/5871	0.69	0/7918
2	BBB	0.67	0/5997	0.70	0/8085
3	CCC	0.67	0/5958	0.70	0/8006
4	RRR	0.93	0/386	1.31	7/598 (1.2%)
5	VVV	0.98	1/345 (0.3%)	1.23	5/535 (0.9%)
6	MaM	0.97	0/98	1.24	0/150
7	XXX	0.63	0/46	0.63	0/61
7	YYY	0.60	0/120	0.62	0/163
All	All	0.68	1/18821 (0.0%)	0.74	12/25516 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	VVV	1	A	OP3-P	-7.32	1.52	1.61

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	VVV	5	G	P-O3'-C3'	-7.91	110.21	119.70
4	RRR	17	C	P-O3'-C3'	-6.76	111.59	119.70
4	RRR	8	C	P-O3'-C3'	-6.40	112.02	119.70
4	RRR	3	U	P-O3'-C3'	-6.24	112.21	119.70
4	RRR	6	C	P-O3'-C3'	-6.24	112.22	119.70
4	RRR	12	U	P-O3'-C3'	-5.87	112.66	119.70
5	VVV	6	U	P-O3'-C3'	-5.78	112.76	119.70
4	RRR	11	C	P-O3'-C3'	-5.52	113.07	119.70
5	VVV	1	A	P-O3'-C3'	-5.35	113.28	119.70
4	RRR	16	G	P-O3'-C3'	-5.26	113.39	119.70
5	VVV	13	A	P-O3'-C3'	-5.19	113.47	119.70
5	VVV	12	G	P-O3'-C3'	-5.07	113.61	119.70

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	AAA	5755	0	5730	11	0
2	BBB	5881	0	5904	18	0
3	CCC	5858	0	6031	14	0
4	RRR	349	0	177	2	0
5	VVV	307	0	153	1	0
6	MaM	89	0	43	0	0
7	XXX	55	0	42	0	0
7	YYY	135	0	108	0	0
8	BBB	5	0	0	0	0
9	MaM	29	0	0	0	0
10	YYY	1	0	0	0	0
All	All	18464	0	18188	40	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 1.

All (40) 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:BBB:340:ALA:HB3	2:BBB:341:PRO:HD3	1.74	0.70
2:BBB:282:LEU:HD22	2:BBB:440:LEU:HD13	1.76	0.67
2:BBB:370:CYS:HA	2:BBB:373:LEU:HD13	1.81	0.63
2:BBB:410:MET:O	4:RRR:16:G:N2	2.33	0.61
1:AAA:305:LEU:HD11	1:AAA:494:ASP:HB3	1.91	0.53
2:BBB:635:SER:O	2:BBB:638:GLY:N	2.44	0.51
1:AAA:661:LEU:HD23	1:AAA:703:PHE:CZ	2.45	0.51
4:RRR:10:G:N3	4:RRR:10:G:H2'	2.27	0.50
3:CCC:61:LEU:HD13	3:CCC:93:CYS:SG	2.53	0.49
1:AAA:533:TRP:N	1:AAA:534:PRO:HD3	2.28	0.48
1:AAA:383:ASP:O	2:BBB:380:ARG:NH1	2.47	0.47
2:BBB:102:LEU:HD11	2:BBB:265:ASN:HD22	1.79	0.47

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:BBB:728:ASP:O	2:BBB:733:ARG:N	2.46	0.47
3:CCC:147:VAL:HG12	3:CCC:226:THR:HG23	1.97	0.47
2:BBB:495:PHE:O	2:BBB:501:VAL:HG22	2.16	0.46
1:AAA:574:ARG:HB3	2:BBB:549:LEU:HD22	1.98	0.46
1:AAA:441:TYR:HB3	1:AAA:604:PHE:CE2	2.51	0.46
1:AAA:533:TRP:N	1:AAA:534:PRO:CD	2.80	0.45
3:CCC:80:LYS:HA	3:CCC:93:CYS:HA	2.00	0.44
3:CCC:578:PHE:CD1	3:CCC:620:PRO:HA	2.52	0.44
2:BBB:370:CYS:N	2:BBB:371:PRO:CD	2.81	0.44
2:BBB:197:LYS:O	2:BBB:199:PHE:N	2.50	0.43
2:BBB:669:ARG:NH1	3:CCC:38:THR:OG1	2.51	0.43
3:CCC:648:ASN:HA	3:CCC:658:LEU:HD11	2.00	0.43
1:AAA:644:TYR:HA	2:BBB:26:GLY:HA2	2.00	0.43
2:BBB:437:TRP:CD1	2:BBB:437:TRP:C	2.91	0.43
1:AAA:279:TYR:CE1	1:AAA:281:ALA:HB3	2.54	0.43
5:VVV:3:U:H6	5:VVV:3:U:O5'	2.00	0.43
3:CCC:619:LEU:N	3:CCC:620:PRO:CD	2.82	0.43
3:CCC:388:ILE:HD11	3:CCC:481:VAL:HG22	2.00	0.42
3:CCC:586:PRO:O	3:CCC:588:LYS:N	2.50	0.42
2:BBB:164:ILE:N	2:BBB:165:PRO:HD2	2.35	0.42
1:AAA:582:GLN:HB2	2:BBB:542:THR:HG21	2.01	0.42
2:BBB:597:ILE:C	2:BBB:597:ILE:HD12	2.41	0.41
3:CCC:653:ARG:HD3	3:CCC:694:VAL:HG21	2.03	0.41
3:CCC:682:GLU:HB2	3:CCC:685:GLU:HB2	2.03	0.41
1:AAA:642:MET:HG3	1:AAA:695:TRP:CZ3	2.56	0.41
3:CCC:659:PHE:CZ	3:CCC:668:LEU:HD11	2.54	0.41
3:CCC:233:MET:HG3	3:CCC:233:MET:O	2.20	0.41
3:CCC:694:VAL:HG13	3:CCC:695:LEU:HG	2.03	0.40

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	AAA	713/751 (95%)	659 (92%)	53 (7%)	1 (0%)	51	83
2	BBB	748/772 (97%)	707 (94%)	37 (5%)	4 (0%)	29	63
3	CCC	728/798 (91%)	663 (91%)	63 (9%)	2 (0%)	41	73
7	XXX	4/28 (14%)	4 (100%)	0	0	100	100
7	YYY	13/28 (46%)	11 (85%)	1 (8%)	1 (8%)	1	5
All	All	2206/2377 (93%)	2044 (93%)	154 (7%)	8 (0%)	34	68

All (8) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	BBB	198	GLY
1	AAA	673	GLY
2	BBB	648	HIS
3	CCC	471	ALA
7	YYY	41	THR
2	BBB	679	ARG
2	BBB	154	ASP
3	CCC	587	GLN

### 5.3.2 Protein sidechains ⓘ

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	AAA	638/664 (96%)	624 (98%)	14 (2%)	52	77
2	BBB	643/657 (98%)	629 (98%)	14 (2%)	52	77
3	CCC	639/694 (92%)	620 (97%)	19 (3%)	41	70
7	XXX	6/24 (25%)	6 (100%)	0	100	100
7	YYY	15/24 (62%)	15 (100%)	0	100	100
All	All	1941/2063 (94%)	1894 (98%)	47 (2%)	49	75

All (47) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	46	TYR
1	AAA	59	LYS
1	AAA	208	ILE
1	AAA	243	LEU
1	AAA	257	LEU
1	AAA	263	ARG
1	AAA	319	GLN
1	AAA	326	MET
1	AAA	339	LEU
1	AAA	404	GLN
1	AAA	407	MET
1	AAA	449	CYS
1	AAA	502	LYS
1	AAA	620	GLN
2	BBB	159	ASP
2	BBB	186	ASN
2	BBB	233	ARG
2	BBB	292	ASN
2	BBB	324	ARG
2	BBB	373	LEU
2	BBB	409	MET
2	BBB	410	MET
2	BBB	437	TRP
2	BBB	488	MET
2	BBB	616	ASP
2	BBB	618	GLU
2	BBB	643	ASP
2	BBB	666	ARG
3	CCC	34	ARG
3	CCC	61	LEU
3	CCC	65	ASP
3	CCC	73	GLU
3	CCC	89	LYS
3	CCC	134	ARG
3	CCC	144	ARG
3	CCC	191	LYS
3	CCC	238	GLN
3	CCC	271	ARG
3	CCC	360	ASP
3	CCC	378	LYS
3	CCC	438	ARG
3	CCC	484	ASN
3	CCC	542	THR

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
3	CCC	572	MET
3	CCC	576	ASP
3	CCC	605	ARG
3	CCC	610	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	RRR	15/18 (83%)	6 (40%)	1 (6%)
5	VVV	13/14 (92%)	6 (46%)	1 (7%)
6	MaM	3/13 (23%)	2 (66%)	0
All	All	31/45 (68%)	14 (45%)	2 (6%)

All (14) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	RRR	4	A
4	RRR	5	C
4	RRR	6	C
4	RRR	13	U
4	RRR	14	C
4	RRR	15	U
5	VVV	3	U
5	VVV	4	A
5	VVV	7	A
5	VVV	8	A
5	VVV	11	A
5	VVV	13	A
6	MaM	2	A
6	MaM	4	C

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
4	RRR	4	A
5	VVV	5	G

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

3 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	$\# Z  > 2$
7	SEP	YYY	35	7	8,9,10	0.60	0	8,12,14	0.60	0
7	SEP	YYY	42	7	8,9,10	0.62	0	8,12,14	0.68	0
7	SEP	XXX	26	7	8,9,10	0.58	0	8,12,14	0.59	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	SEP	YYY	35	7	-	1/5/8/10	-
7	SEP	YYY	42	7	-	2/5/8/10	-
7	SEP	XXX	26	7	-	4/5/8/10	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (7) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	YYY	35	SEP	N-CA-CB-OG
7	XXX	26	SEP	N-CA-CB-OG
7	XXX	26	SEP	CB-OG-P-O1P
7	XXX	26	SEP	CB-OG-P-O2P
7	XXX	26	SEP	CB-OG-P-O3P
7	YYY	42	SEP	N-CA-CB-OG
7	YYY	42	SEP	CA-CB-OG-P

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](#)

2 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$
9	IC5	MaM	101	6	26,31,31	1.16	3 (11%)	31,49,49	2.05	3 (9%)
8	PO4	BBB	801	-	4,4,4	0.65	0	6,6,6	0.42	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
9	IC5	MaM	101	6	-	5/12/32/32	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
9	MaM	101	IC5	PB-O3B	3.57	1.62	1.50
9	MaM	101	IC5	C6-N1	3.05	1.38	1.33
9	MaM	101	IC5	C2'-C1'	-2.43	1.50	1.53

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	MaM	101	IC5	C5-C6-N1	-8.57	111.71	123.43
9	MaM	101	IC5	C2-N1-C6	5.92	125.33	115.93

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	MaM	101	IC5	N3-C2-N1	-3.05	123.16	127.22

There are no chirality outliers.

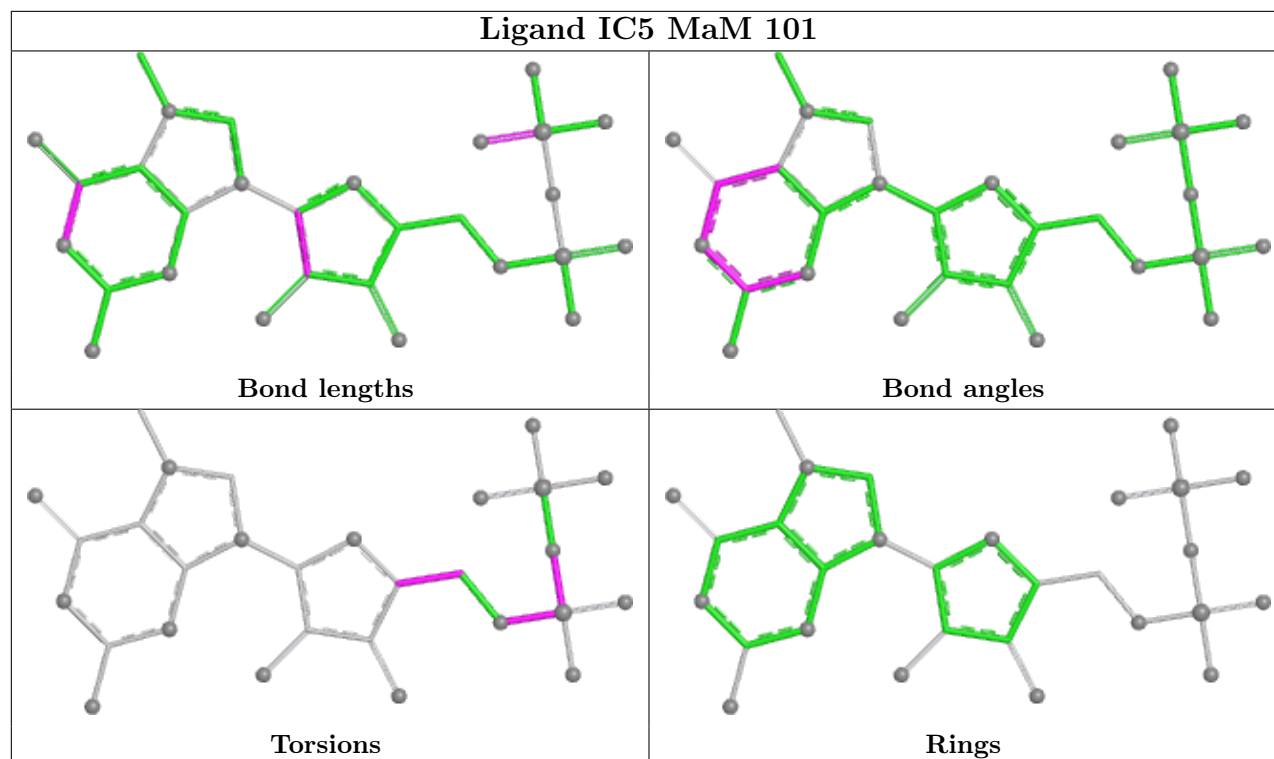
All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
9	MaM	101	IC5	C5'-O5'-PA-O1A
9	MaM	101	IC5	PB-O3A-PA-O5'
9	MaM	101	IC5	O4'-C4'-C5'-O5'
9	MaM	101	IC5	C3'-C4'-C5'-O5'
9	MaM	101	IC5	PB-O3A-PA-O1A

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 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 ⓘ

### 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	AAA	717/751 (95%)	0.32	56 (7%) 13 5	59, 104, 208, 247	0
2	BBB	750/772 (97%)	0.03	2 (0%) 94 89	62, 106, 146, 186	0
3	CCC	732/798 (91%)	0.22	32 (4%) 34 17	71, 111, 187, 223	0
4	RRR	18/18 (100%)	0.13	0 100 100	66, 94, 173, 175	0
5	VVV	14/14 (100%)	-0.36	0 100 100	72, 78, 81, 81	0
6	MaM	5/13 (38%)	0.75	0 100 100	143, 146, 192, 212	0
7	XXX	6/28 (21%)	1.66	3 (50%) 0 0	137, 139, 150, 152	0
7	YYY	15/28 (53%)	-0.05	0 100 100	106, 126, 141, 142	0
All	All	2257/2422 (93%)	0.19	93 (4%) 37 18	59, 107, 192, 247	0

All (93) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	189	VAL	7.1
1	AAA	150	TYR	6.9
1	AAA	190	LEU	6.5
1	AAA	73	LEU	4.8
1	AAA	147	ILE	4.7
1	AAA	183	LEU	4.7
1	AAA	141	ASN	4.5
3	CCC	374	GLY	4.2
3	CCC	471	ALA	3.9
1	AAA	148	PHE	3.8
3	CCC	451	TRP	3.7
3	CCC	465	ILE	3.6
1	AAA	181	LEU	3.6
1	AAA	186	LEU	3.5
3	CCC	389	ASN	3.5
1	AAA	145	LEU	3.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	AAA	187	TRP	3.4
1	AAA	162	LEU	3.3
3	CCC	479	VAL	3.2
3	CCC	386	LEU	3.2
3	CCC	388	ILE	3.1
1	AAA	191	ILE	3.1
1	AAA	184	LYS	3.1
3	CCC	476	LEU	3.1
1	AAA	193	GLU	3.1
1	AAA	32	PRO	3.0
3	CCC	390	SER	3.0
1	AAA	136	LYS	3.0
1	AAA	195	ASP	2.9
3	CCC	470	ASN	2.9
1	AAA	170	VAL	2.9
3	CCC	702	GLY	2.9
1	AAA	137	GLU	2.8
3	CCC	406	PHE	2.8
1	AAA	188	GLN	2.8
1	AAA	156	LEU	2.8
3	CCC	387	LEU	2.8
1	AAA	77	TYR	2.7
1	AAA	124	THR	2.7
1	AAA	30	LEU	2.7
3	CCC	340	PHE	2.7
1	AAA	149	SER	2.7
3	CCC	462	LEU	2.7
7	XXX	24	PRO	2.6
1	AAA	5	ILE	2.6
1	AAA	177	LEU	2.6
1	AAA	72	ASN	2.6
1	AAA	154	TYR	2.5
3	CCC	407	SER	2.5
3	CCC	480	VAL	2.5
3	CCC	482	THR	2.5
3	CCC	376	LEU	2.5
1	AAA	185	ASN	2.5
7	XXX	25	THR	2.4
1	AAA	0	SER	2.4
1	AAA	2	ASP	2.4
3	CCC	381	MET	2.4
1	AAA	164	GLU	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	AAA	196	VAL	2.4
1	AAA	123	ILE	2.4
1	AAA	6	THR	2.4
1	AAA	121	VAL	2.4
3	CCC	481	VAL	2.3
1	AAA	146	MET	2.3
3	CCC	367	VAL	2.3
1	AAA	142	SER	2.3
1	AAA	1	MET	2.3
7	XXX	22	TYR	2.3
2	BBB	389	LEU	2.3
1	AAA	140	GLY	2.3
3	CCC	692	ASN	2.3
1	AAA	165	GLU	2.2
3	CCC	478	GLY	2.2
1	AAA	3	THR	2.2
3	CCC	463	HIS	2.2
1	AAA	46	TYR	2.2
1	AAA	62	THR	2.2
1	AAA	158	ASN	2.2
2	BBB	647	ALA	2.2
1	AAA	51	MET	2.2
3	CCC	728	ILE	2.2
1	AAA	49	SER	2.1
3	CCC	587	GLN	2.1
1	AAA	45	CYS	2.1
3	CCC	413	PHE	2.1
1	AAA	131	TYR	2.1
3	CCC	371	GLU	2.1
1	AAA	161	SER	2.1
3	CCC	726	ALA	2.1
1	AAA	9	PHE	2.0
1	AAA	120	GLU	2.0
3	CCC	378	LYS	2.0
1	AAA	36	PHE	2.0

## 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( $\text{\AA}^2$ )	Q<0.9
7	SEP	XXX	26	10/11	0.84	0.16	132,136,147,149	0
7	SEP	YYY	42	10/11	0.89	0.11	141,149,154,155	0
7	SEP	YYY	35	10/11	0.91	0.14	121,123,125,128	0

### 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

### 6.4 Ligands [i](#)

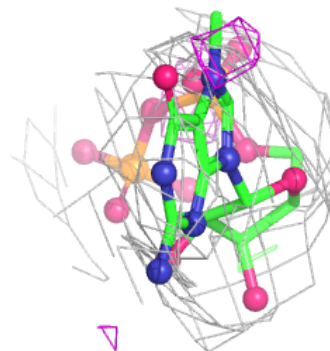
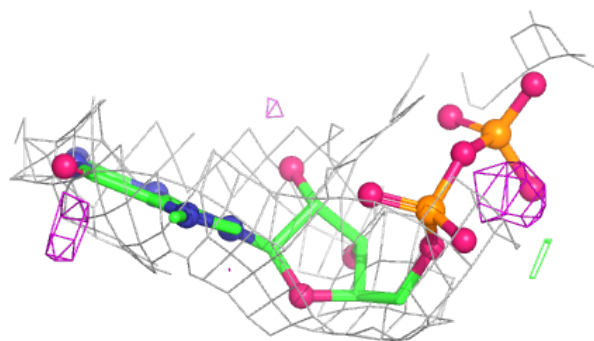
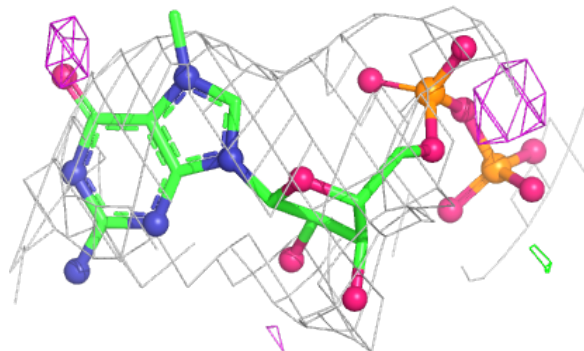
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 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
9	IC5	MaM	101	29/29	0.85	0.23	156,181,183,186	0
8	PO4	BBB	801	5/5	0.96	0.37	136,137,138,141	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 IC5 MaM 101:**

$2mF_o - DF_c$  (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.