



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 1, 2026 – 10:09 AM UTC

PDB ID : 5NKL / pdb_00005nkl
Title : Crystal structure of the large fragment of DNA polymerase I from *Thermus Aquaticus* in a closed ternary complex with the artificial base pair dDs-dPxTP
Authors : Betz, K.; Marx, A.; Diederichs, K.; Hirao, I.; Kimoto, M.
Deposited on : 2017-03-31
Resolution : 1.70 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The 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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Buster-report	:	wwPDB partial adaption of 1.1.7 (2018)
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

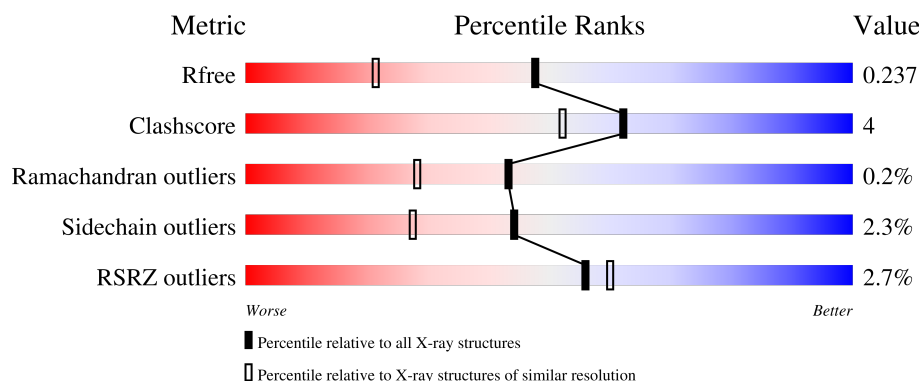
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.70 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}	180053	5551 (1.70-1.70)
Clashscore	190562	5924 (1.70-1.70)
Ramachandran outliers	187476	5846 (1.70-1.70)
Sidechain outliers	187428	5846 (1.70-1.70)
RSRZ outliers	180081	5554 (1.70-1.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. 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	540	3% 90% 9%
2	B	12	58% 42%
3	C	16	75% 25%

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 10035 atoms, of which 4752 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 polymerase I, thermostable.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	540	Total	C	H	N	O	S	0	6	0
			8694	2741	4375	781	784	13			

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*AP*CP*CP*AP*CP*GP*GP*CP*GP*CP*(DOC))-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	12	Total	C	H	N	O	P	0	0	0
			374	114	134	48	67	11			

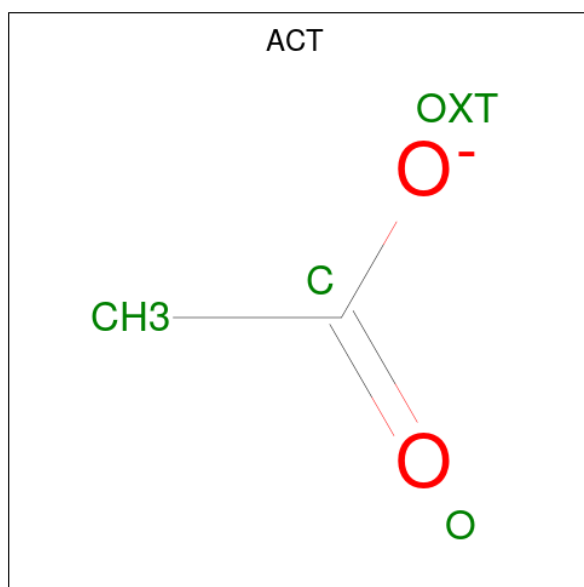
- Molecule 3 is a DNA chain called DNA (5'-D(*AP*AP*AP*(DNU)P*GP*GP*CP*GP*CP*CP*GP*TP*GP*GP*TP*C)-3').

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
3	C	16	Total	C	H	N	O	P	S	0	0	0
			515	161	182	64	92	15	1			

- Molecule 4 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

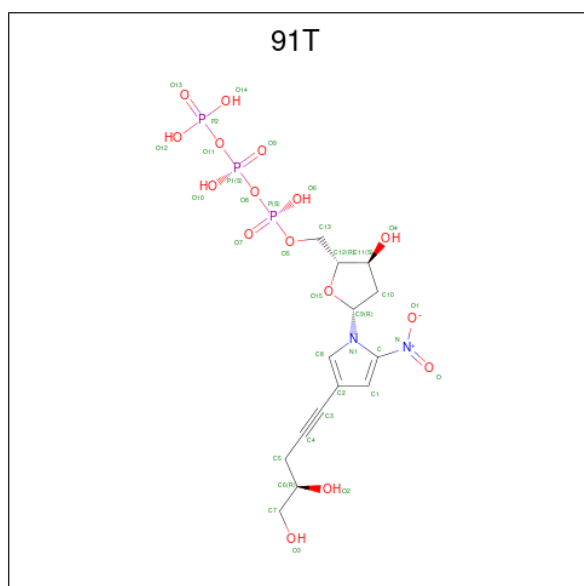
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	Mg	0	0
			2	2		

- Molecule 5 is ACETATE ION (CCD ID: ACT) (formula: C₂H₃O₂).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	H	O	0	0
			7	2	3	2		

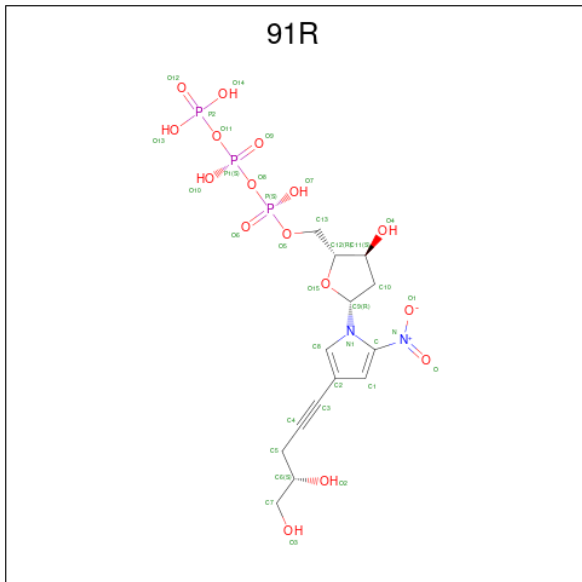
- Molecule 6 is [[(2 {R},3 {S},5 {R})-5-[4-[(4 {R})-4,5-bis(oxidanyl)pent-1-ynyl]-2-nitro-pyrro l-1-yl]-3-oxidanyl-oxolan-2-yl]methoxy-oxidanyl-phosphoryl] phosphono hydrogen phosphate (CCD ID: 91T) (formula: C₁₄H₂₁N₂O₁₆P₃).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
6	A	1	Total	C	H	N	O	P	0	1
			56	14	21	2	16	3		

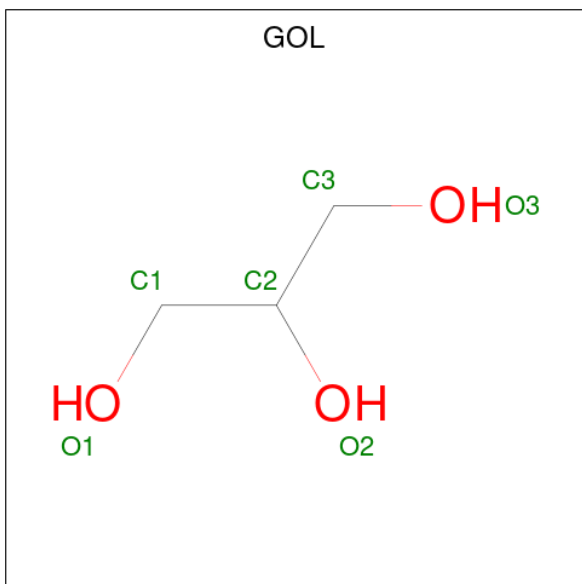
- Molecule 7 is [[(2 {R},3 {S},5 {R})-5-[4-[(4 {S})-4,5-bis(oxidanyl)pent-1-ynyl]-2-nitro-pyrro l-1-yl]-3-oxidanyl-oxolan-2-yl]methoxy-oxidanyl-phosphoryl] phosphono hydrogen phosphate (CCD ID: 91T) (formula: C₁₄H₂₁N₂O₁₆P₃).

-1-yl]-3-oxidanyl-oxolan-2-yl]methoxy-oxidanyl-phosphoryl] phosphono hydrogen phosphate (CCD ID: 91R) (formula: $C_{14}H_{21}N_2O_{16}P_3$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
7	A	1	Total	C	H	N	O	P	0	1
			56	14	21	2	16	3		

- Molecule 8 is GLYCEROL (CCD ID: GOL) (formula: $\text{C}_3\text{H}_8\text{O}_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	C	1	Total	C	H	O	0	0
			14	3	8	3		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	C	1	Total	C	H	O	0	0
			14	3	8	3		

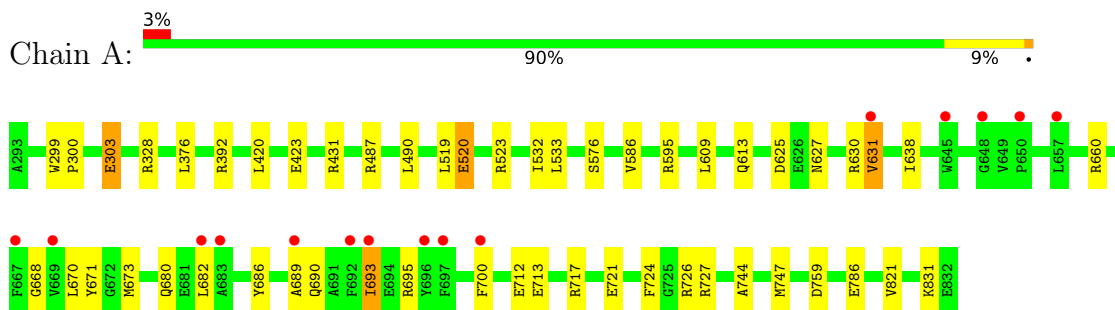
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	244	Total	O	0	0
			244	244		
9	B	26	Total	O	0	0
			26	26		
9	C	33	Total	O	0	0
			33	33		

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.

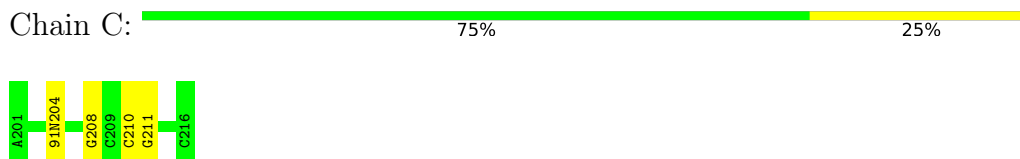
- Molecule 1: DNA polymerase I, thermostable



- Molecule 2: DNA (5'-D(*GP*AP*CP*CP*AP*CP*GP*GP*CP*GP*CP*(DOC))-3')



- Molecule 3: DNA (5'-D(*AP*AP*AP*(DNU)P*GP*GP*CP*GP*CP*CP*GP*TP*GP*GP*TP*P*C)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	108.78Å 108.78Å 90.57Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.10 – 1.70 47.10 – 1.70	Depositor EDS
% Data completeness (in resolution range)	98.5 (47.10-1.70) 98.3 (47.10-1.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.00 (at 1.70Å)	Xtriage
Refinement program	PHENIX (dev_2650: ???)	Depositor
R, R_{free}	0.181 , 0.237 0.182 , 0.237	Depositor DCC
R_{free} test set	3392 reflections (2.19%)	wwPDB-VP
Wilson B-factor (Å ²)	31.8	Xtriage
Anisotropy	0.271	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 46.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.036 for -h,-k,l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	10035	wwPDB-VP
Average B, all atoms (Å ²)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.74% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 91T, GOL, MG, 91R, DOC, ACT, 91N

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	0.50	0/4430	0.64	0/5999
2	B	0.66	0/249	0.80	0/382
3	C	0.66	0/345	0.79	0/529
All	All	0.52	0/5024	0.66	0/6910

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	4319	4375	4352	35	0
2	B	240	134	134	4	0
3	C	333	182	169	3	0
4	A	2	0	0	0	0
5	A	4	3	3	0	0
6	A	35	21	0	2	0
7	A	35	21	0	1	0
8	C	12	16	16	0	0
9	A	244	0	0	6	0
9	B	26	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
9	C	33	0	0	0	0
All	All	5283	4752	4674	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 4.

The worst 5 of 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)
1:A:625:ASP:OD2	9:A:1001:HOH:O	1.94	0.84
1:A:700:PHE:HA	9:A:1001:HOH:O	2.01	0.61
1:A:520:GLU:OE1	1:A:523:ARG:NH2	2.38	0.56
3:C:210:DC:H2'	3:C:211:DG:C8	2.42	0.55
1:A:631:VAL:HG11	1:A:638:ILE:HD11	1.89	0.55

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	544/540 (101%)	526 (97%)	17 (3%)	1 (0%)	43 28

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	586	VAL

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	446/441 (101%)	436 (98%)	10 (2%)	45 29

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	693	ILE
1	A	695	ARG
1	A	712	GLU
1	A	520	GLU
1	A	631	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	443	HIS
1	A	680	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	$\# Z > 2$
2	DOC	B	112	3,2	16,19,20	0.54	0	20,26,29	0.50	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
2	DOC	B	112	3,2	-	0/7/18/19	0/2/2/2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	$\# Z > 2$
5	ACT	A	903	-	3,3,3	1.62	1 (33%)	3,3,3	0.60	0
7	91R	A	905[B]	4	34,36,36	0.79	1 (2%)	42,54,54	0.63	1 (2%)
8	GOL	C	302	-	5,5,5	0.17	0	5,5,5	0.32	0
6	91T	A	904[A]	4	34,36,36	0.76	1 (2%)	42,54,54	0.67	0
8	GOL	C	301	-	5,5,5	0.16	0	5,5,5	0.40	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
8	GOL	C	301	-	-	0/4/4/4	-
8	GOL	C	302	-	-	4/4/4/4	-
6	91T	A	904[A]	4	-	7/29/46/46	0/2/2/2
7	91R	A	905[B]	4	-	5/29/46/46	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	A	904[A]	91T	C8-C2	-3.40	1.33	1.38
7	A	905[B]	91R	C8-C2	-3.36	1.33	1.38
5	A	903	ACT	CH3-C	2.45	1.58	1.49

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	905[B]	91R	C5-C4-C3	-2.01	171.51	176.43

There are no chirality outliers.

5 of 16 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	904[A]	91T	C4-C5-C6-C7
6	A	904[A]	91T	C4-C5-C6-O2
7	A	905[B]	91R	P1-O11-P2-O14
8	C	302	GOL	C1-C2-C3-O3
8	C	302	GOL	O1-C1-C2-C3

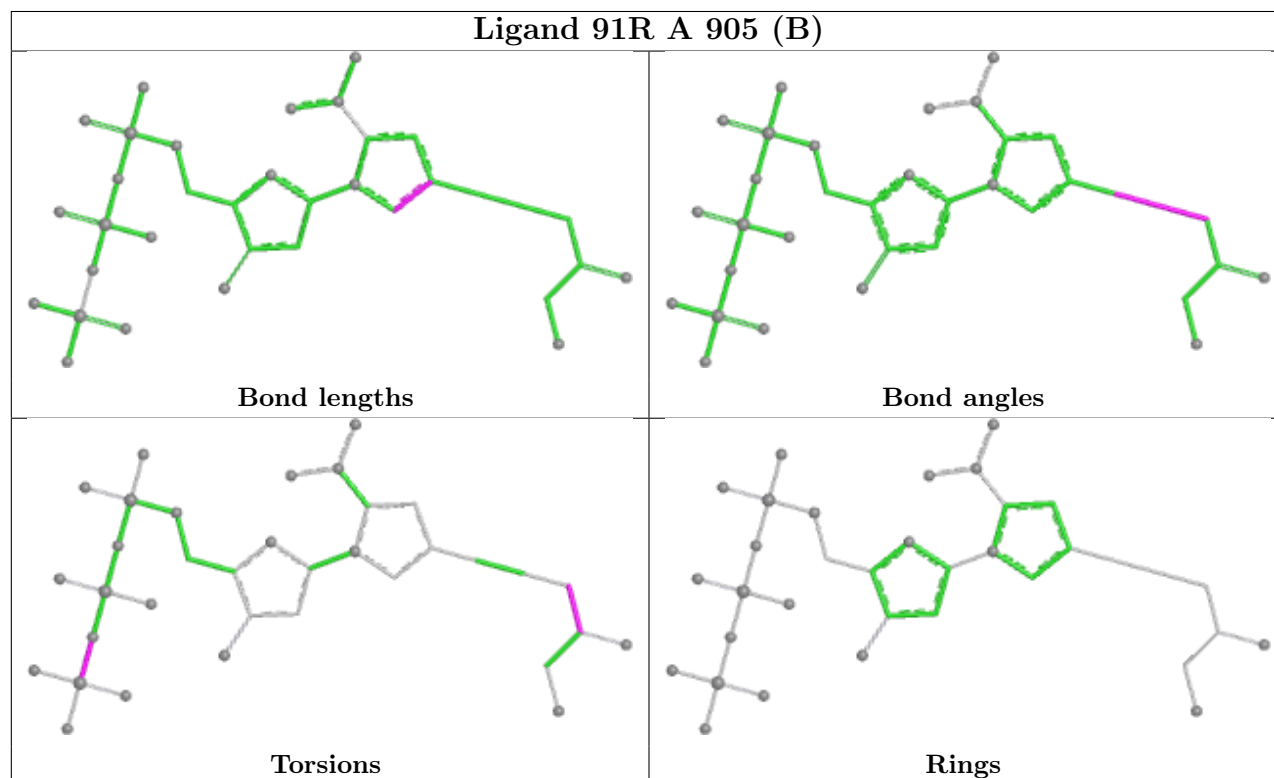
There are no ring outliers.

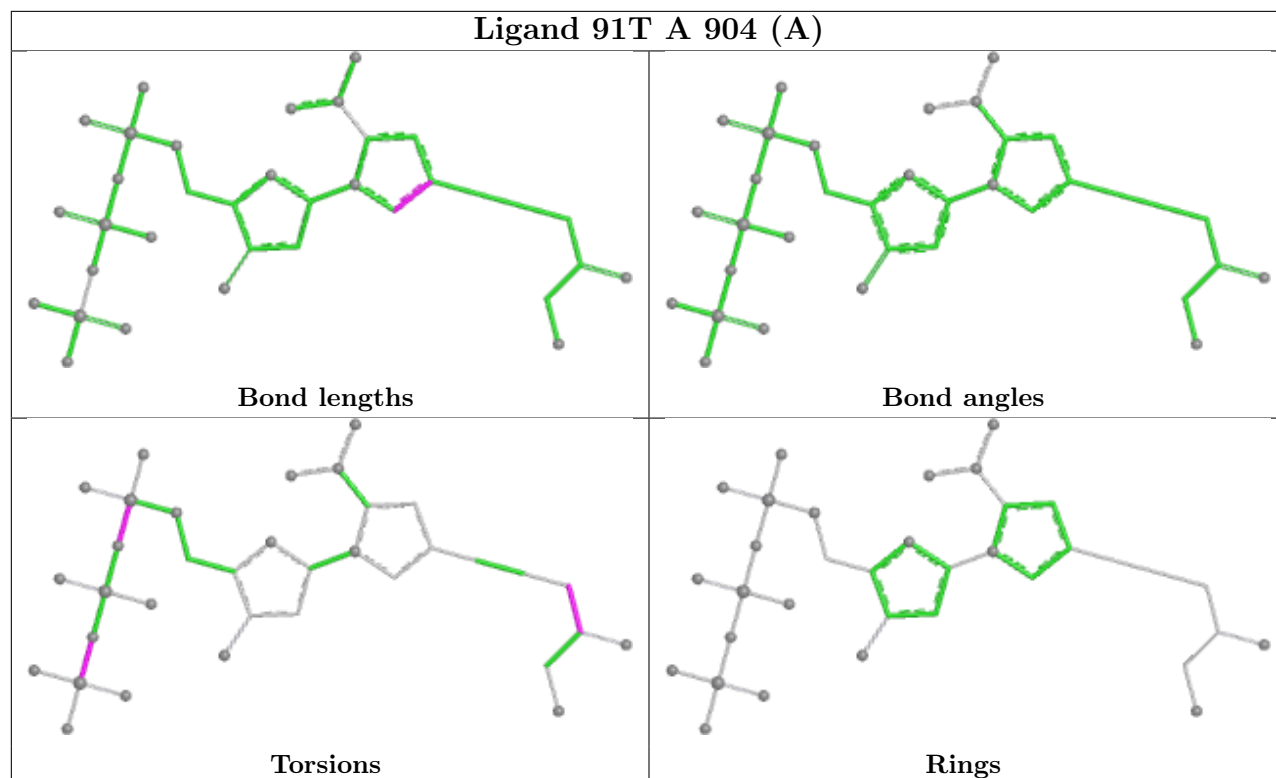
2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	A	905[B]	91R	1	0
6	A	904[A]	91T	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, 95th 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 ²)	Q<0.9
1	A	540/540 (100%)	-0.05	15 (2%) 55 59	18, 50, 109, 165	3 (0%)
2	B	11/12 (91%)	-0.67	0 100 100	32, 35, 69, 71	0
3	C	15/16 (93%)	-0.33	0 100 100	32, 38, 117, 146	0
All	All	566/568 (99%)	-0.07	15 (2%) 56 60	18, 49, 112, 165	3 (0%)

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	645	TRP	3.5
1	A	689	ALA	3.5
1	A	697	PHE	3.3
1	A	693	ILE	3.1
1	A	692	PHE	2.9

6.2 Non-standard residues in protein, DNA, RNA chains [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, 95th 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 ²)	Q<0.9
2	DOC	B	112	18/19	0.97	0.05	28,36,43,44	0

6.3 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

6.4 Ligands

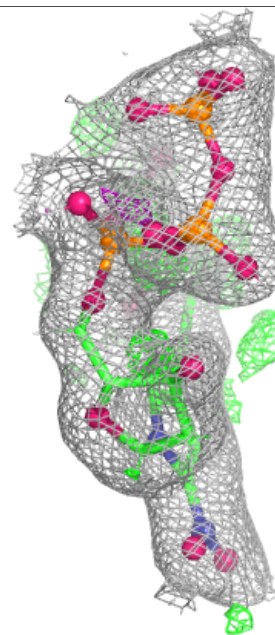
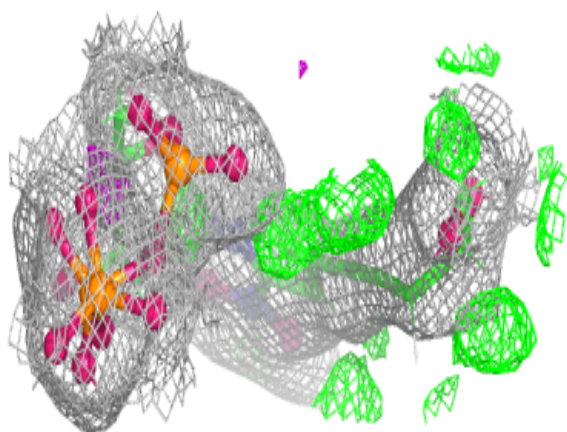
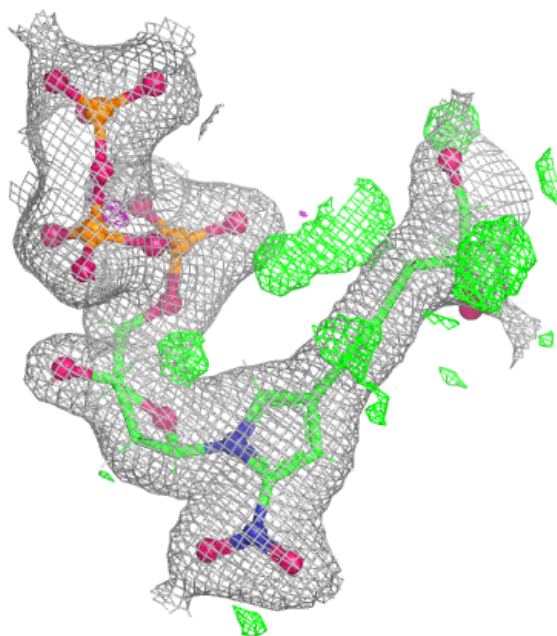
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
5	ACT	A	903	4/4	0.80	0.14	51,61,67,71	0
8	GOL	C	302	6/6	0.87	0.10	36,53,61,70	14
8	GOL	C	301	6/6	0.92	0.08	44,61,73,81	0
6	91T	A	904[A]	35/35	0.93	0.08	31,42,51,61	56
7	91R	A	905[B]	35/35	0.94	0.08	30,42,51,63	56
4	MG	A	902	1/1	0.98	0.09	45,45,45,45	0
4	MG	A	901	1/1	0.98	0.04	46,46,46,46	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.

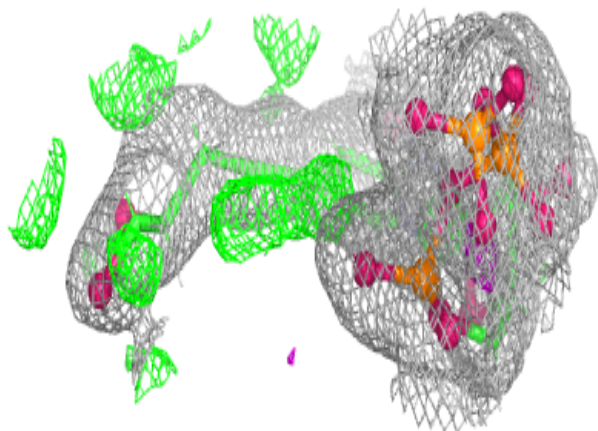
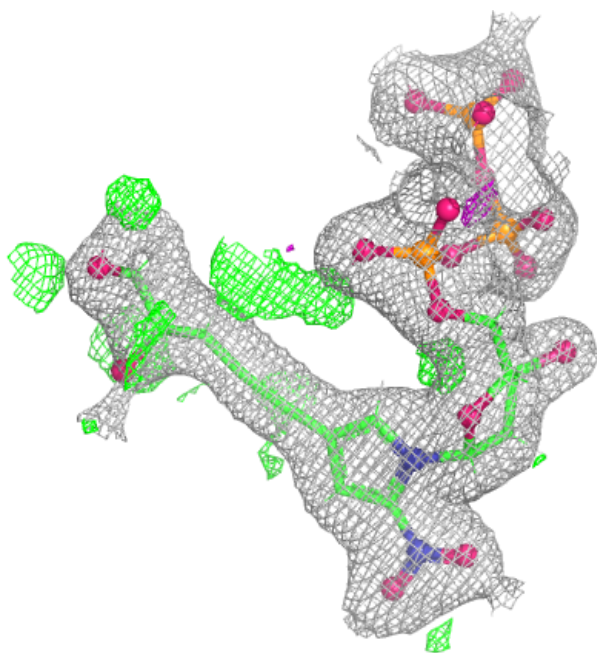
Electron density around 91T A 904 (A):

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



Electron density around 91R A 905 (B):

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



6.5 Other polymers ⓘ

There are no such residues in this entry.