



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 05:44 pm BST

PDB ID : 8ICE  
Title : DNA POLYMERASE BETA (POL B) (E.C.2.7.7.7) COMPLEXED WITH SEVEN BASE PAIRS OF DNA; SOAKED IN THE PRESENCE OF DATP (1 MILLIMOLAR) AND CDCL2 (1 MILLIMOLAR)  
Authors : Pelletier, H.; Sawaya, M.R.  
Deposited on : 1996-04-19  
Resolution : 3.20 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 : 2.11  
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) : 2.11

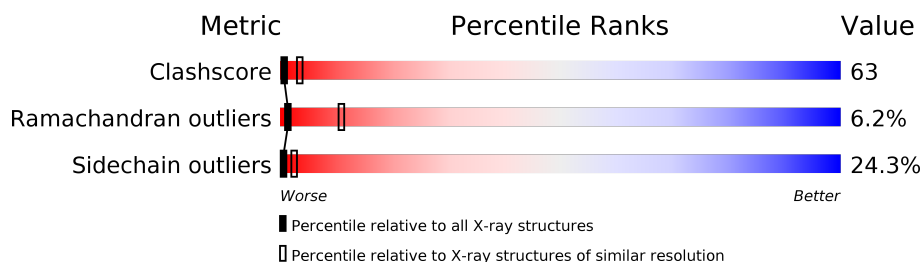
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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))
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	T	8	
2	P	8	
3	A	335	

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 3064 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 DNA chain called DNA (5'-D(\*CP\*AP\*TP\*TP\*AP\*GP\*AP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	T	8	Total	C	N	O	P	0	0	0
			145	69	27	42	7			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	8	Total	C	N	O	P	0	0	0
			145	69	24	45	7			

- Molecule 3 is a protein called PROTEIN (DNA POLYMERASE BETA (E.C.2.7.7.7)).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	327	Total	C	N	O	S	18	0	0
			2623	1657	458	499	9			

- Molecule 4 is CADMIUM ION (three-letter code: CD) (formula: Cd).

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

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Na	0	0
			2	2		

- Molecule 6 is 2'-DEOXYADENOSINE 5'-TRIPHOSPHATE (three-letter code: DTP) (formula: C<sub>10</sub>H<sub>16</sub>N<sub>5</sub>O<sub>12</sub>P<sub>3</sub>).



- Molecule 7 is water.



### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

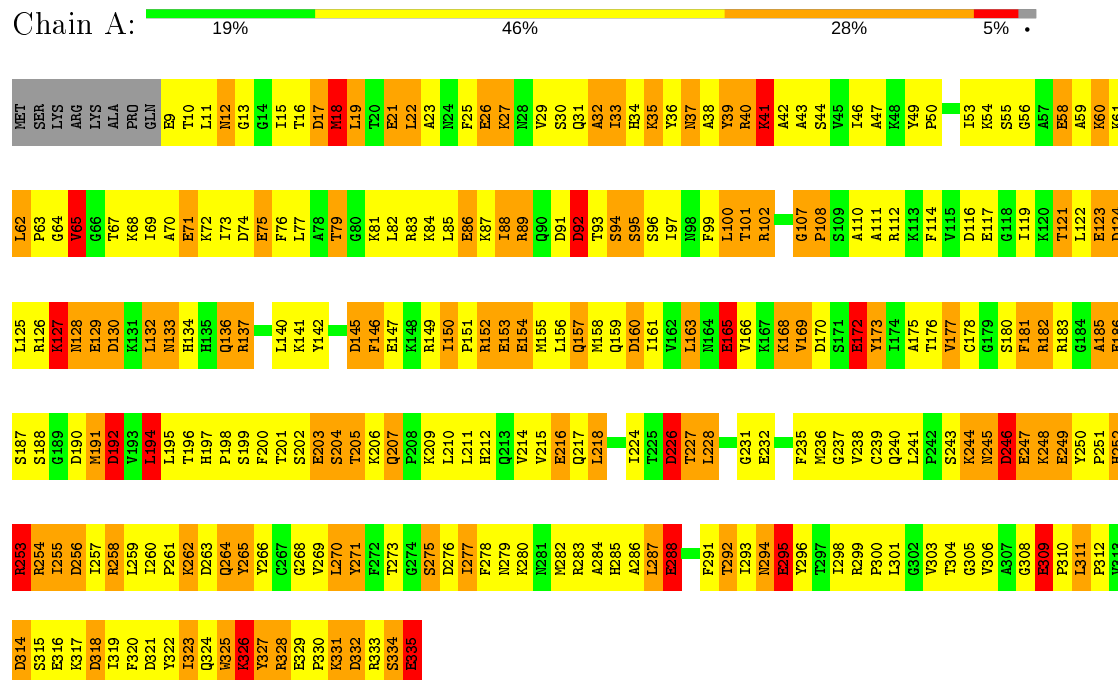
- Molecule 1: DNA (5'-D(\*CP\*AP\*TP\*TP\*AP\*GP\*AP\*A)-3')



- Molecule 2: DNA (5'-D(\*TP\*CP\*TP\*AP\*AP\*TP\*GP\*A)-3')



- Molecule 3: PROTEIN (DNA POLYMERASE BETA (E.C.2.7.7.7))



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	178.26 Å 57.76 Å 48.26 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 3.20 11.82 – 2.70	Depositor EDS
% Data completeness (in resolution range)	90.0 (20.00-3.20) 85.1 (11.82-2.70)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.10 (at 2.70 Å)	Xtriage
Refinement program	TNT 5-D	Depositor
R, $R_{free}$	0.171 , (Not available) 0.162 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	27.5	Xtriage
Anisotropy	0.222	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.15 , 105.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	3064	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.53% 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: NA, DTP, CD

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	T	2.11	5/162 (3.1%)	3.80	24/249 (9.6%)
2	P	2.16	6/160 (3.8%)	5.03	23/246 (9.3%)
3	A	1.31	28/2672 (1.0%)	1.79	63/3590 (1.8%)
All	All	1.42	39/2994 (1.3%)	2.29	110/4085 (2.7%)

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
3	A	3	0

The worst 5 of 39 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	T	3	DT	C1'-N1	8.20	1.59	1.49
3	A	203	GLU	CD-OE1	7.80	1.34	1.25
3	A	75	GLU	CD-OE1	7.72	1.34	1.25
3	A	172	GLU	CD-OE2	7.66	1.34	1.25
3	A	147	GLU	CD-OE2	7.52	1.33	1.25

The worst 5 of 110 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	P	7	DG	C8-N9-C1'	30.52	166.67	127.00
2	P	7	DG	C4-N9-C1'	-30.07	87.42	126.50
2	P	1	DT	C6-N1-C1'	-25.45	82.23	120.40
1	T	4	DT	C6-N1-C1'	-24.05	84.33	120.40
2	P	1	DT	C2-N1-C1'	23.26	155.42	118.20

All (3) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	17	ASP	CA
3	A	40	ARG	CA
3	A	314	ASP	CA

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	T	145	0	80	11	0
2	P	145	0	81	18	0
3	A	2623	0	2641	334	0
4	A	2	0	0	0	0
5	A	2	0	0	0	0
6	A	9	0	0	0	0
7	A	106	0	0	10	0
7	P	17	0	0	2	0
7	T	15	0	0	2	0
All	All	3064	0	2802	356	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 63.

The worst 5 of 356 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)
3:A:31:GLN:NE2	3:A:112:ARG:HH12	1.50	1.07
3:A:150:ILE:HD13	3:A:253:ARG:HG2	1.42	1.00
3:A:119:ILE:HG21	3:A:125:LEU:HD23	1.43	0.99
3:A:163:LEU:HD21	3:A:175:ALA:HB3	1.44	0.98
3:A:31:GLN:HE21	3:A:112:ARG:HH12	1.11	0.95

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
3	A	325/335 (97%)	261 (80%)	44 (14%)	20 (6%)	<b>1</b> <b>11</b>

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	204	SER
3	A	206	LYS
3	A	207	GLN
3	A	244	LYS
3	A	246	ASP

### 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
3	A	288/295 (98%)	218 (76%)	70 (24%)	<b>0</b> <b>2</b>

5 of 70 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	157	GLN
3	A	192	ASP
3	A	325	TRP
3	A	165	GLU
3	A	173	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such

sidechains are listed below:

Mol	Chain	Res	Type
3	A	133	ASN
3	A	136	GLN
3	A	264	GLN
3	A	128	ASN
3	A	279	ASN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 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	$\# Z  > 2$
6	DTP	A	338	4	6,8,32	1.39	1 (16%)	13,13,50	1.19	1 (7%)

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
6	DTP	A	338	4	-	3/6/6/34	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	A	338	DTP	PB-O3A	2.18	1.63	1.54

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	A	338	DTP	O2B-PB-O3B	2.12	111.76	104.64

There are no chirality outliers.

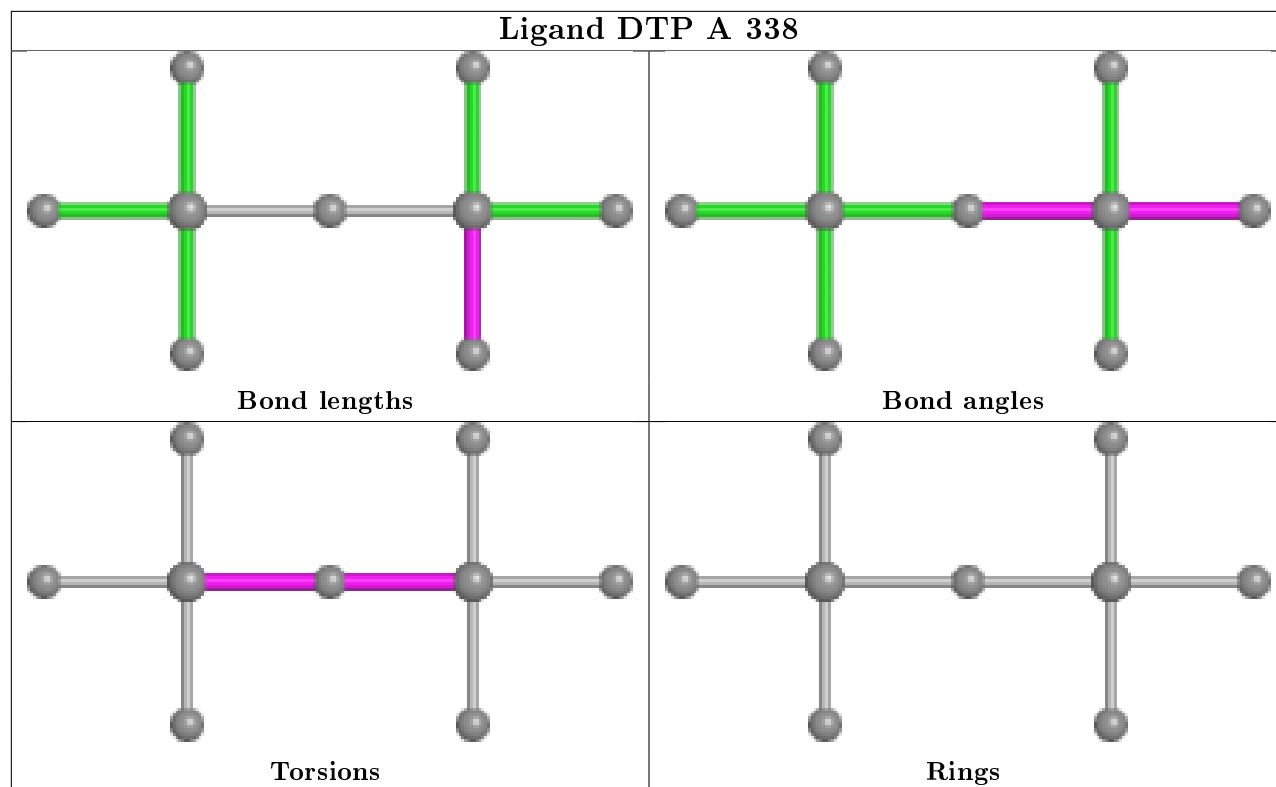
All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	338	DTP	PB-O3B-PG-O2G
6	A	338	DTP	PG-O3B-PB-O2B
6	A	338	DTP	PG-O3B-PB-O1B

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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

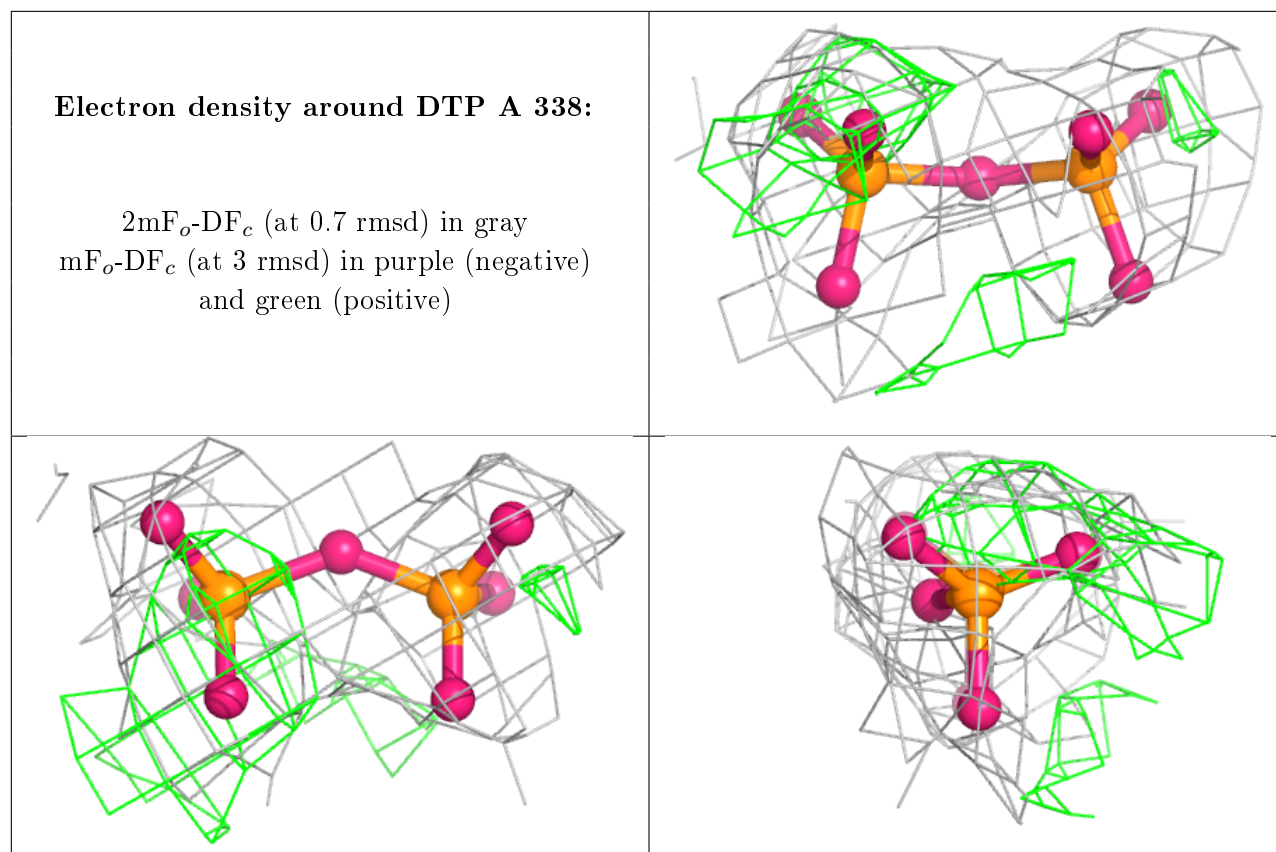
### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



## 6.5 Other polymers [i](#)

Unable to reproduce the depositor's R factor - this section is therefore empty.