



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 31, 2023 – 06:39 PM JST

PDB ID : 5GTL  
Title : NADPH complex structure of Aldehyde Dehydrogenase from Bacillus cereus  
Authors : Ngo, H.P.T.; Hong, S.H.; Ho, T.H.; Oh, D.K.; Kang, L.W.  
Deposited on : 2016-08-21  
Resolution : 2.00 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 : 2.36  
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) : 2.36

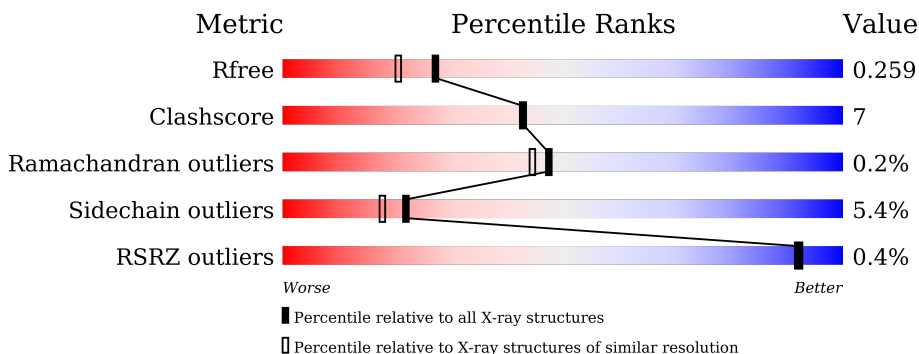
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	494	86% 12% ..
1	B	494	86% 13% ..
1	C	494	79% 18% ..
1	D	494	82% 15% ..

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 16824 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.

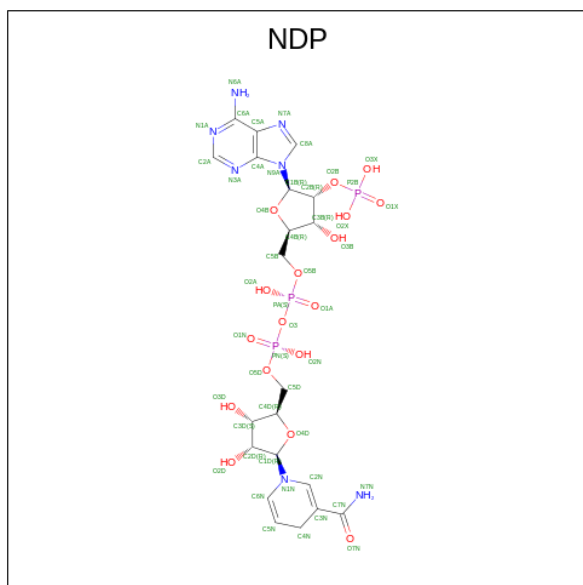
- Molecule 1 is a protein called Betaine-aldehyde dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	491	3791	2414	630	733	14	0	0	0
1	B	491	3791	2414	630	733	14	0	0	0
1	C	488	3768	2400	626	728	14	0	0	0
1	D	489	3776	2406	627	729	14	0	0	0

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Na	0	0
			2	2		
2	B	2	Total	Na	0	0
			2	2		
2	C	2	Total	Na	0	0
			2	2		
2	D	1	Total	Na	0	0
			1	1		

- Molecule 3 is NADPH DIHYDRO-NICOTINAMIDE-ADENINE-DINUCLEOTIDE PHOSPHATE (three-letter code: NDP) (formula: C<sub>21</sub>H<sub>30</sub>N<sub>7</sub>O<sub>17</sub>P<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
3	A	1	48	21	7	17	3	0	0
3	B	1	48	21	7	17	3	0	0
3	C	1	48	21	7	17	3	0	0
3	D	1	48	21	7	17	3	0	0

- Molecule 4 is water.

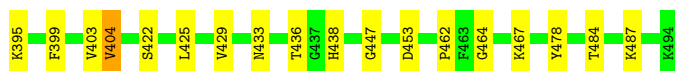
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	454	Total	O	0	0
			454	454		
4	B	421	Total	O	0	0
			421	421		
4	C	291	Total	O	0	0
			291	291		
4	D	333	Total	O	0	0
			333	333		





- Molecule 1: Betaine-aldehyde dehydrogenase

Chain D: 82% 15% ..



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	83.89Å 93.94Å 144.93Å 90.00° 97.67° 90.00°	Depositor
Resolution (Å)	47.88 – 2.00 47.88 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (47.88-2.00) 99.9 (47.88-2.00)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.91 (at 2.00Å)	Xtrriage
Refinement program	REFMAC 5.8.0135	Depositor
R, $R_{free}$	0.189 , 0.252 0.203 , 0.259	Depositor DCC
$R_{free}$ test set	7525 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	28.2	Xtrriage
Anisotropy	0.067	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 45.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	16824	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.75% 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: NDP, NA

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	1.23	1/3870 (0.0%)	0.98	4/5250 (0.1%)
1	B	1.22	2/3870 (0.1%)	0.99	4/5250 (0.1%)
1	C	1.08	0/3847	0.96	2/5218 (0.0%)
1	D	1.12	0/3855	0.97	2/5229 (0.0%)
All	All	1.16	3/15442 (0.0%)	0.97	12/20947 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	156	GLU	CD-OE2	-6.02	1.19	1.25
1	A	338	GLU	CD-OE1	-5.83	1.19	1.25
1	B	199	SER	CB-OG	-5.24	1.35	1.42

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	305	ARG	NE-CZ-NH1	8.94	124.77	120.30
1	B	83	ARG	NE-CZ-NH2	-7.94	116.33	120.30
1	B	305	ARG	NE-CZ-NH2	-7.29	116.66	120.30
1	B	254	ARG	NE-CZ-NH1	7.01	123.81	120.30
1	A	254	ARG	NE-CZ-NH1	6.71	123.65	120.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen



atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3791	0	3740	51	0
1	B	3791	0	3740	44	0
1	C	3768	0	3716	58	0
1	D	3776	0	3727	66	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	1	0	0	1	0
3	A	48	0	26	1	0
3	B	48	0	26	0	0
3	C	48	0	26	0	0
3	D	48	0	26	4	0
4	A	454	0	0	6	0
4	B	421	0	0	6	0
4	C	291	0	0	10	0
4	D	333	0	0	8	0
All	All	16824	0	15027	207	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 7.

The worst 5 of 207 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:C:351:VAL:O	1:C:355:ILE:HD12	1.32	1.28
1:C:391:MET:SD	4:C:806:HOH:O	2.00	1.17
1:D:347:GLN:O	1:D:351:VAL:HG12	1.66	0.95
1:A:195:GLN:H	1:A:195:GLN:HE21	1.22	0.86
1:D:385:THR:HG21	4:D:883:HOH:O	1.76	0.85

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	489/494 (99%)	474 (97%)	15 (3%)	0	100	100
1	B	489/494 (99%)	470 (96%)	19 (4%)	0	100	100
1	C	486/494 (98%)	463 (95%)	21 (4%)	2 (0%)	34	30
1	D	487/494 (99%)	466 (96%)	20 (4%)	1 (0%)	47	44
All	All	1951/1976 (99%)	1873 (96%)	75 (4%)	3 (0%)	47	44

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	373	LEU
1	C	424	GLY
1	D	166	PRO

### 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	398/401 (99%)	382 (96%)	16 (4%)	31	29
1	B	398/401 (99%)	380 (96%)	18 (4%)	27	24
1	C	395/401 (98%)	367 (93%)	28 (7%)	14	10
1	D	396/401 (99%)	372 (94%)	24 (6%)	18	14
All	All	1587/1604 (99%)	1501 (95%)	86 (5%)	22	18

5 of 86 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	387	VAL
1	D	195	GLN
1	C	406	LEU
1	D	109	ASN
1	D	311	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 48 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	325	ASN
1	C	482	ASN
1	C	326	ASN
1	C	433	ASN
1	D	114	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 [i](#)

Of 11 ligands modelled in this entry, 7 are monoatomic - leaving 4 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	$\# Z  > 2$
3	NDP	D	502	-	45,52,52	1.24	7 (15%)	53,80,80	1.33	5 (9%)
3	NDP	A	503	-	45,52,52	1.13	5 (11%)	53,80,80	1.35	8 (15%)
3	NDP	C	503	-	45,52,52	1.23	5 (11%)	53,80,80	1.40	8 (15%)
3	NDP	B	503	-	45,52,52	1.10	5 (11%)	53,80,80	1.34	8 (15%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NDP	D	502	-	-	3/30/77/77	0/5/5/5
3	NDP	A	503	-	-	4/30/77/77	0/5/5/5
3	NDP	C	503	-	-	3/30/77/77	0/5/5/5
3	NDP	B	503	-	-	4/30/77/77	0/5/5/5

The worst 5 of 22 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	503	NDP	C6N-C5N	3.97	1.40	1.33
3	D	502	NDP	C6N-C5N	3.32	1.39	1.33
3	D	502	NDP	C2A-N3A	3.22	1.37	1.32
3	B	503	NDP	C5A-C4A	3.10	1.49	1.40
3	A	503	NDP	C6N-C5N	3.03	1.38	1.33

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	503	NDP	O4D-C1D-N1N	4.42	116.70	108.06
3	A	503	NDP	O4D-C1D-N1N	4.06	116.00	108.06
3	D	502	NDP	O4D-C1D-N1N	4.01	115.90	108.06
3	C	503	NDP	N3A-C2A-N1A	-3.63	123.01	128.68
3	C	503	NDP	C2D-C1D-N1N	-3.49	104.56	113.30

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

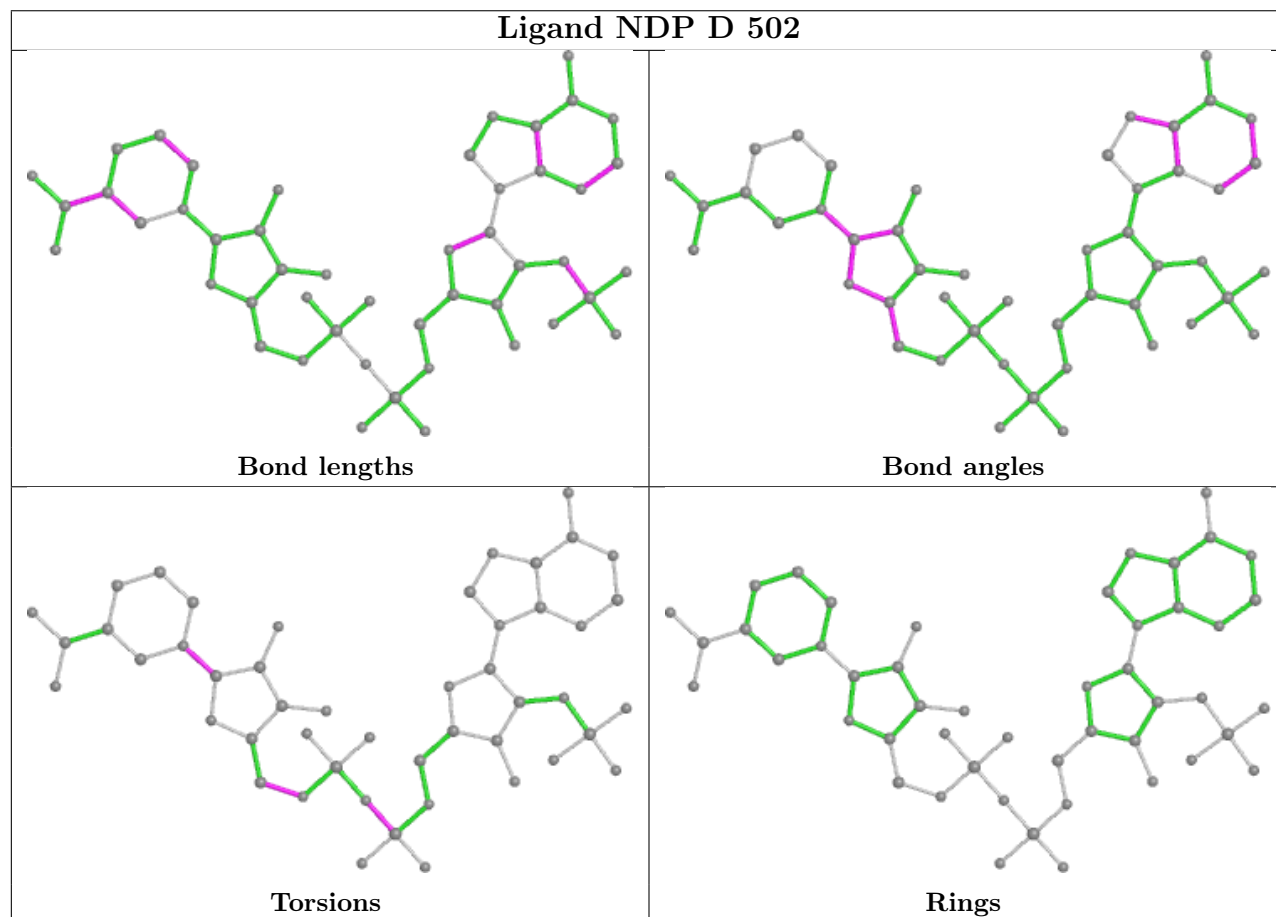
Mol	Chain	Res	Type	Atoms
3	A	503	NDP	O4D-C1D-N1N-C2N
3	A	503	NDP	C2N-C3N-C7N-N7N
3	B	503	NDP	O4D-C1D-N1N-C2N
3	B	503	NDP	C2N-C3N-C7N-N7N
3	C	503	NDP	O4D-C1D-N1N-C2N

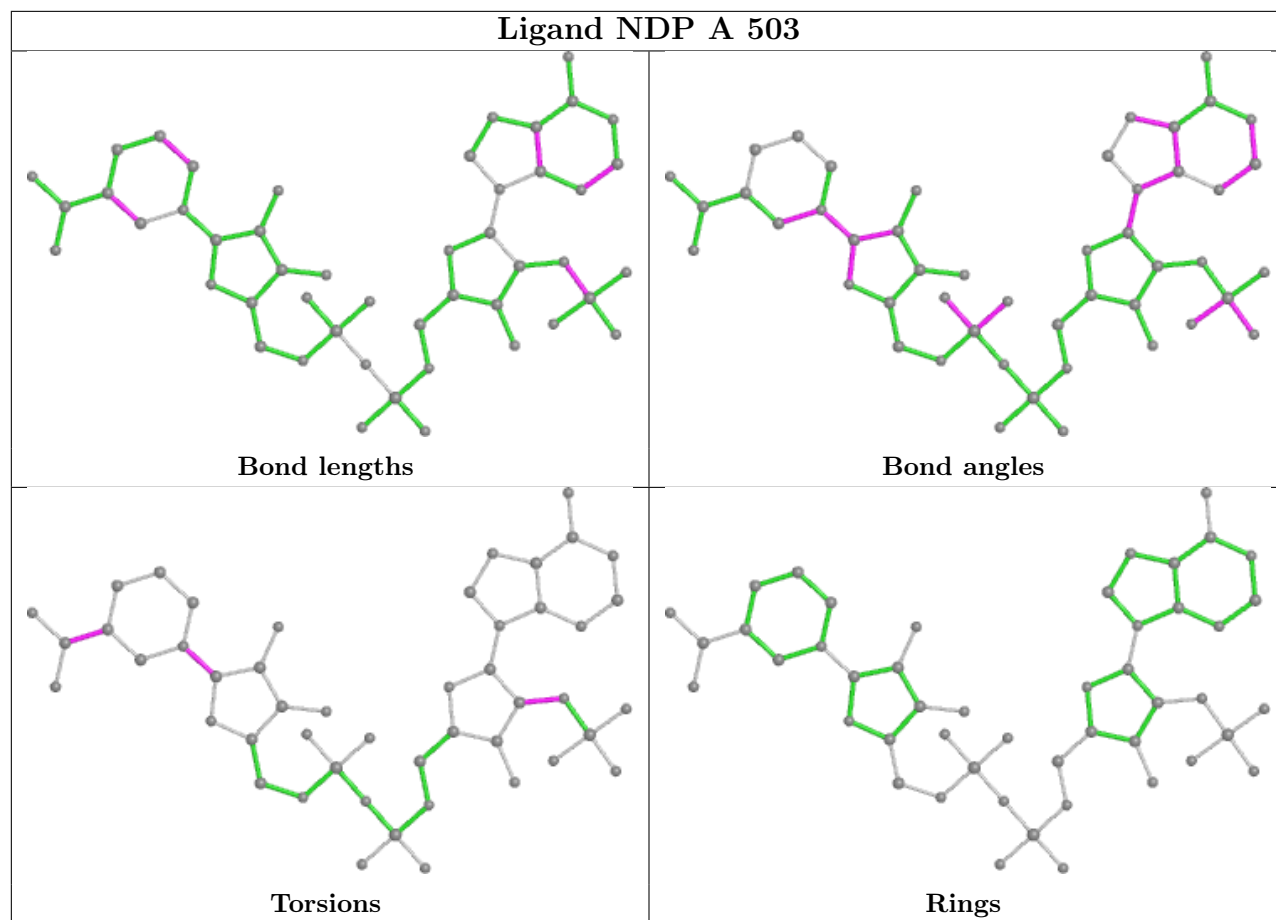
There are no ring outliers.

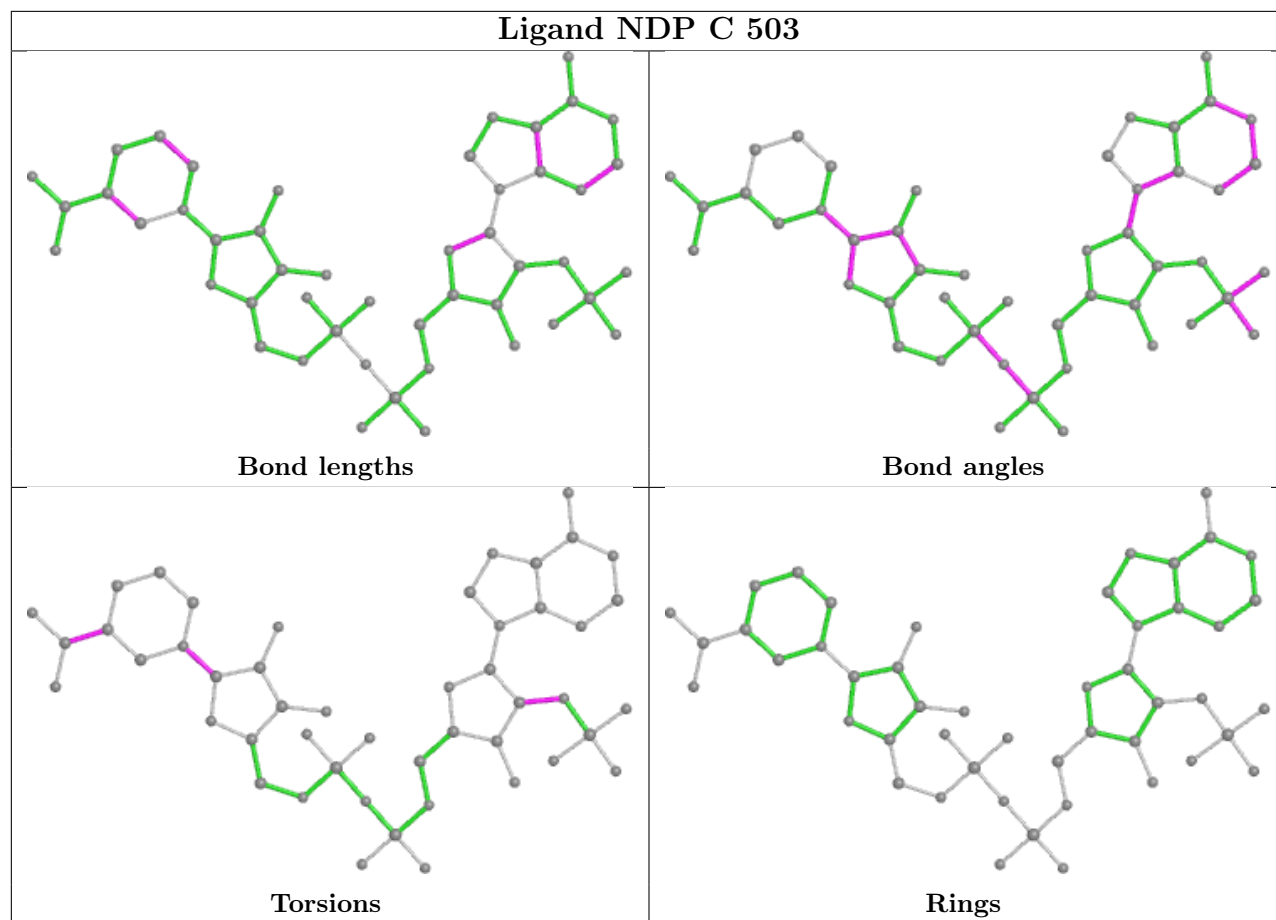
2 monomers are involved in 5 short contacts:

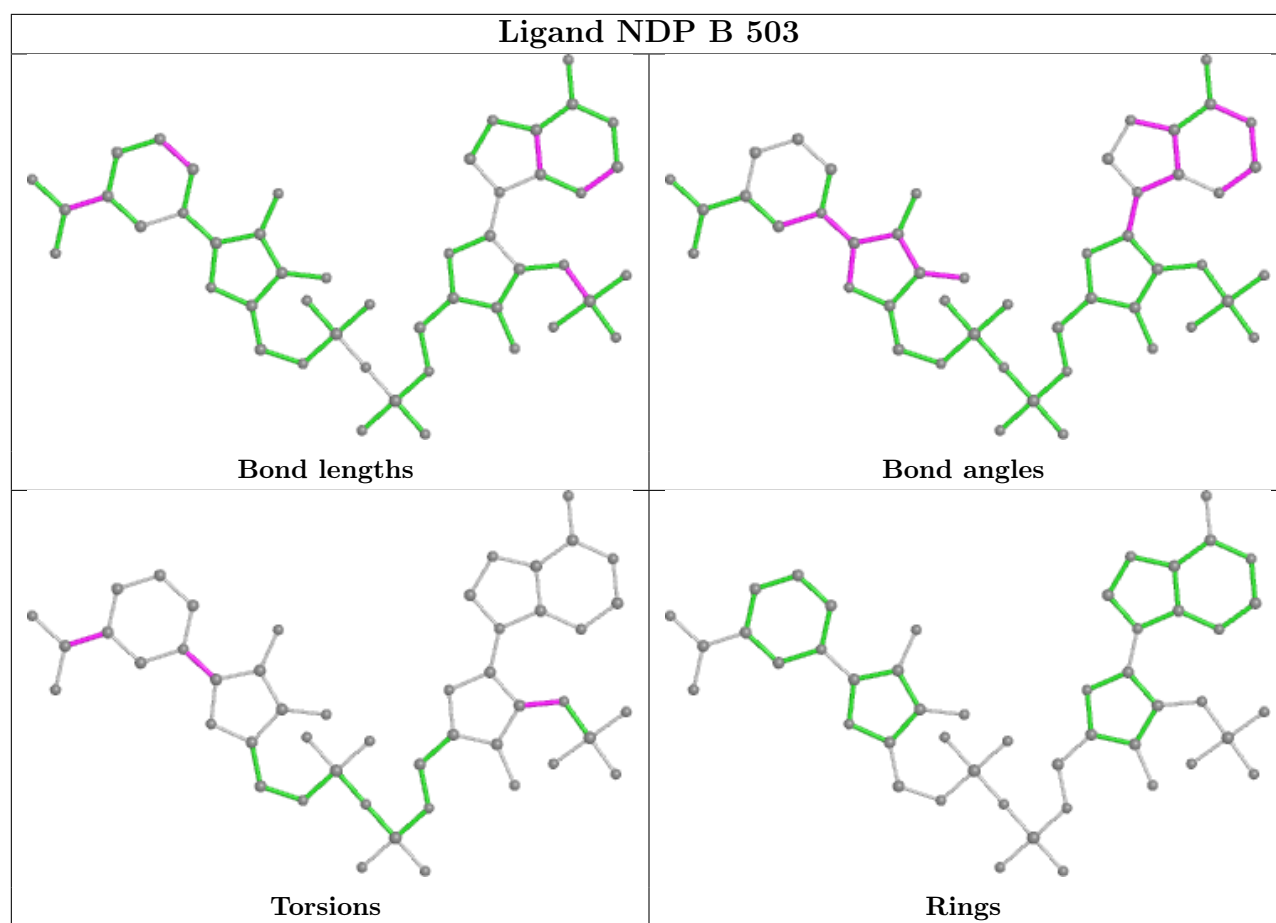
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	502	NDP	4	0
3	A	503	NDP	1	0

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	491/494 (99%)	-0.37	2 (0%) 92 92	13, 27, 42, 60	0
1	B	491/494 (99%)	-0.25	2 (0%) 92 92	15, 28, 43, 60	0
1	C	488/494 (98%)	-0.06	2 (0%) 92 92	19, 35, 58, 76	0
1	D	489/494 (98%)	-0.17	1 (0%) 95 94	18, 33, 50, 72	0
All	All	1959/1976 (99%)	-0.22	7 (0%) 92 92	13, 30, 50, 76	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	360	LYS	3.0
1	B	176	TRP	2.7
1	C	353	ASN	2.4
1	A	176	TRP	2.4
1	B	463	PHE	2.2

### 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](#)

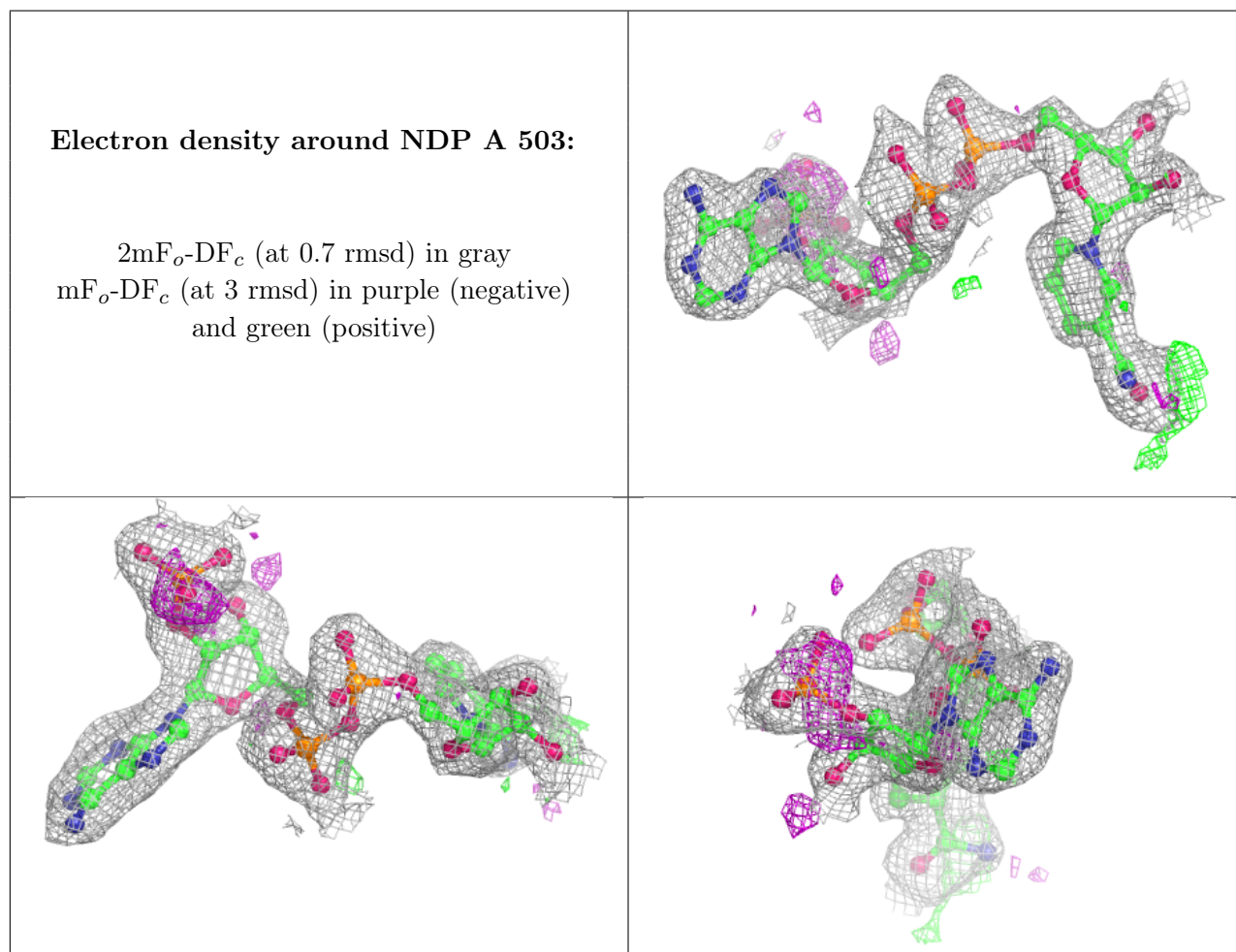
There are no monosaccharides in this entry.

### 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.

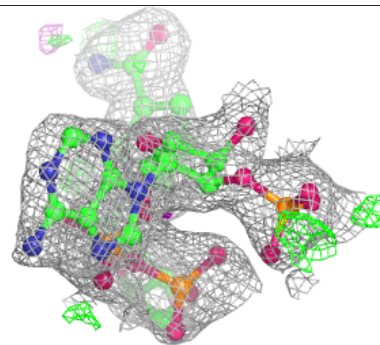
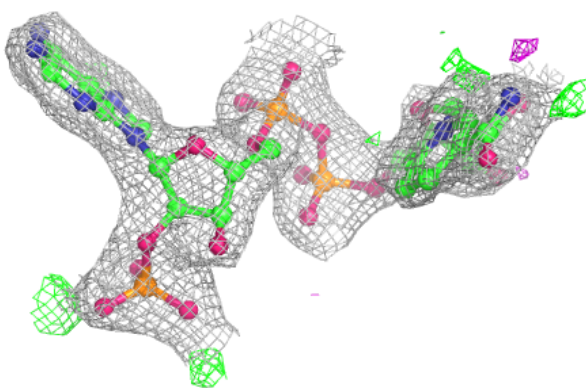
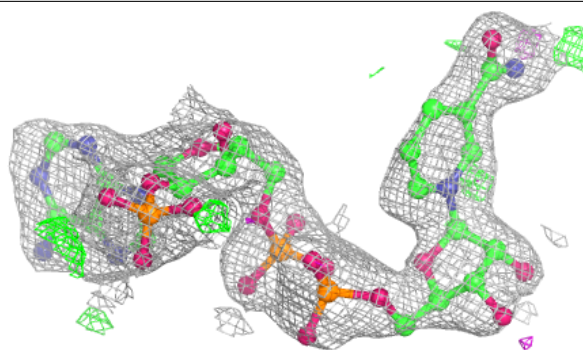
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	NA	C	502	1/1	0.85	0.09	40,40,40,40	0
3	NDP	A	503	48/48	0.91	0.12	36,45,51,52	0
3	NDP	D	502	48/48	0.92	0.12	43,54,69,73	0
3	NDP	B	503	48/48	0.94	0.11	32,39,44,46	0
3	NDP	C	503	48/48	0.94	0.12	41,49,56,60	0
2	NA	D	501	1/1	0.94	0.12	35,35,35,35	0
2	NA	B	502	1/1	0.95	0.07	32,32,32,32	0
2	NA	B	501	1/1	0.96	0.10	27,27,27,27	0
2	NA	A	502	1/1	0.96	0.05	28,28,28,28	0
2	NA	C	501	1/1	0.96	0.08	42,42,42,42	0
2	NA	A	501	1/1	0.98	0.07	23,23,23,23	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.



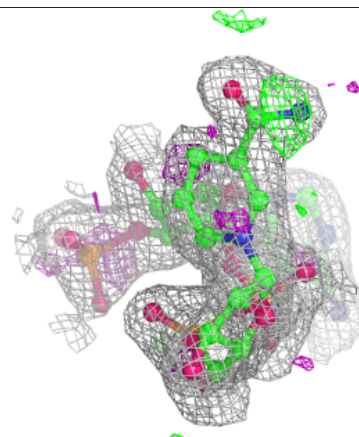
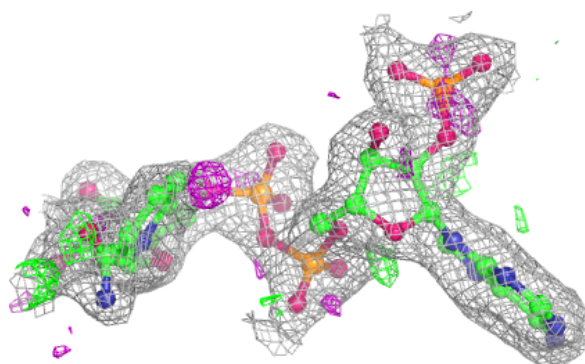
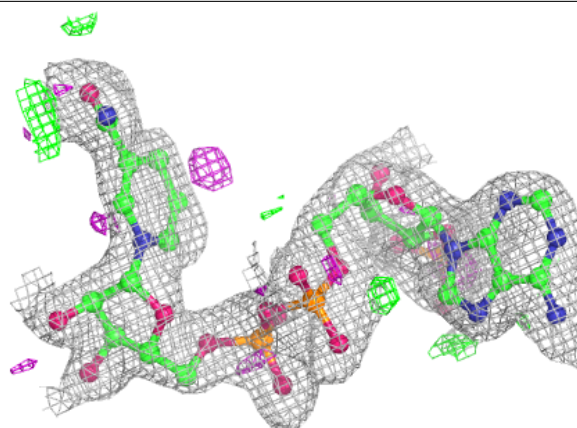
**Electron density around NDP D 502:**

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

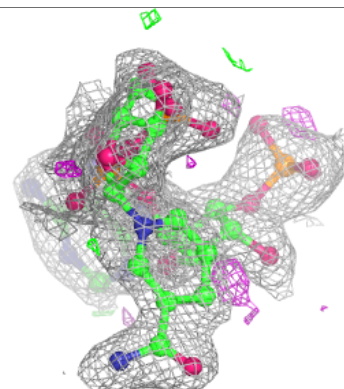
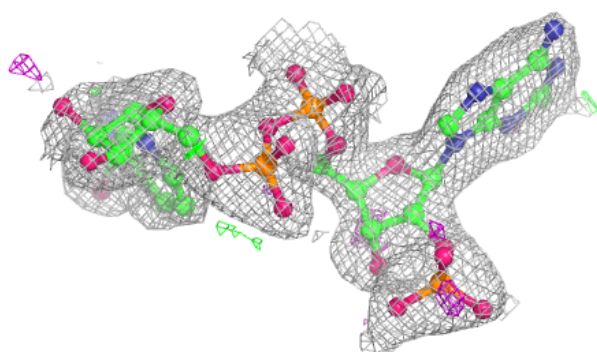
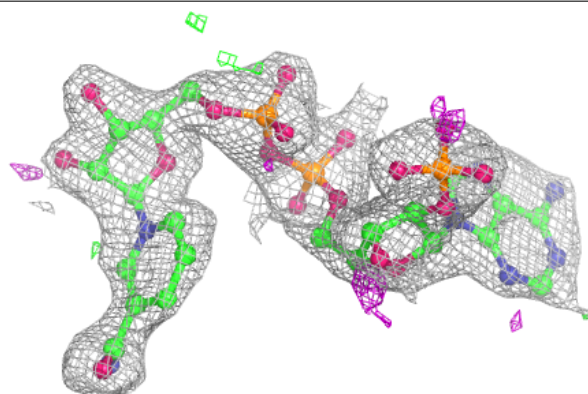


**Electron density around NDP B 503:**

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

**Electron density around NDP C 503:**

$2mF_o-DF_c$  (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.