



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 24, 2025 – 09:06 PM JST

PDB ID : 9J0S / pdb\_00009j0s  
Title : CRYSTAL STRUCTURE OF A NOVEL ALDEHYDE DEHYDROGENASE  
FROM KLEBSIELLA PNEUMONIAE WITH LIGAND  
Authors : Zhang, J.; Han, Y.; Liu, W.D.; Zhang, W.Y.  
Deposited on : 2024-08-02  
Resolution : 2.88 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 : 4-5-2 with Phenix2.0rc1  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 2.0rc1  
EDS : 3.0  
buster-report : 1.1.7 (2018)  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.006 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.44

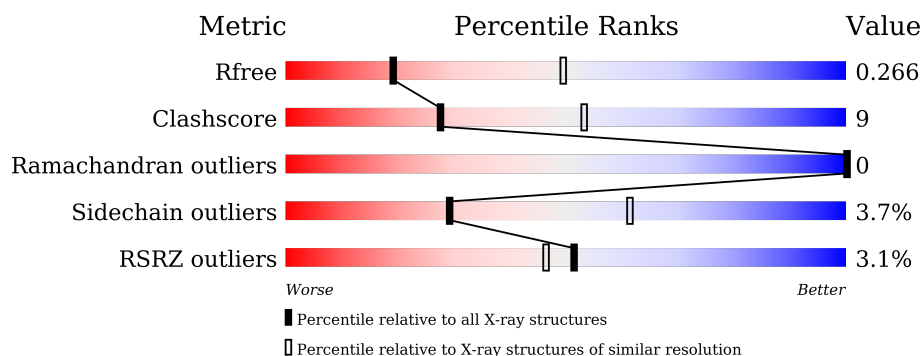
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.88 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))
$R_{free}$	164625	3316 (2.90-2.86)
Clashscore	180529	3609 (2.90-2.86)
Ramachandran outliers	177936	3529 (2.90-2.86)
Sidechain outliers	177891	3532 (2.90-2.86)
RSRZ outliers	164620	3319 (2.90-2.86)

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	489	<div> <div>5%</div> <div> <div>69%</div> <div>22%</div> <div>8%</div> </div> </div>
1	B	489	<div> <div>2%</div> <div> <div>74%</div> <div>19%</div> <div>7%</div> </div> </div>
1	C	489	<div> <div>3%</div> <div> <div>72%</div> <div>19%</div> <div>7%</div> </div> </div>
1	D	489	<div> <div>0%</div> <div> <div>76%</div> <div>16%</div> <div>8%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 13700 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.

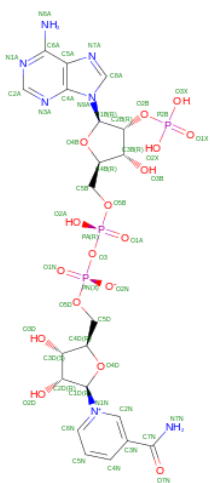
- Molecule 1 is a protein called Aldehyde dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	452	Total	C	N	O	S	0	0	0
			3239	2071	560	606	2			
1	B	457	Total	C	N	O	S	0	0	0
			3411	2162	608	638	3			
1	C	454	Total	C	N	O	S	0	0	0
			3402	2156	609	634	3			
1	D	452	Total	C	N	O	S	0	0	0
			3418	2168	613	634	3			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	277	SER	ASN	conflict	UNP A0A069Q1D5
B	277	SER	ASN	conflict	UNP A0A069Q1D5
C	277	SER	ASN	conflict	UNP A0A069Q1D5
D	277	SER	ASN	conflict	UNP A0A069Q1D5

- Molecule 2 is NADP NICOTINAMIDE-ADENINE-DINUCLEOTIDE PHOSPHATE (CCD ID: NAP) (formula: C<sub>21</sub>H<sub>28</sub>N<sub>7</sub>O<sub>17</sub>P<sub>3</sub>) (labeled as "Ligand of Interest" by depositor).



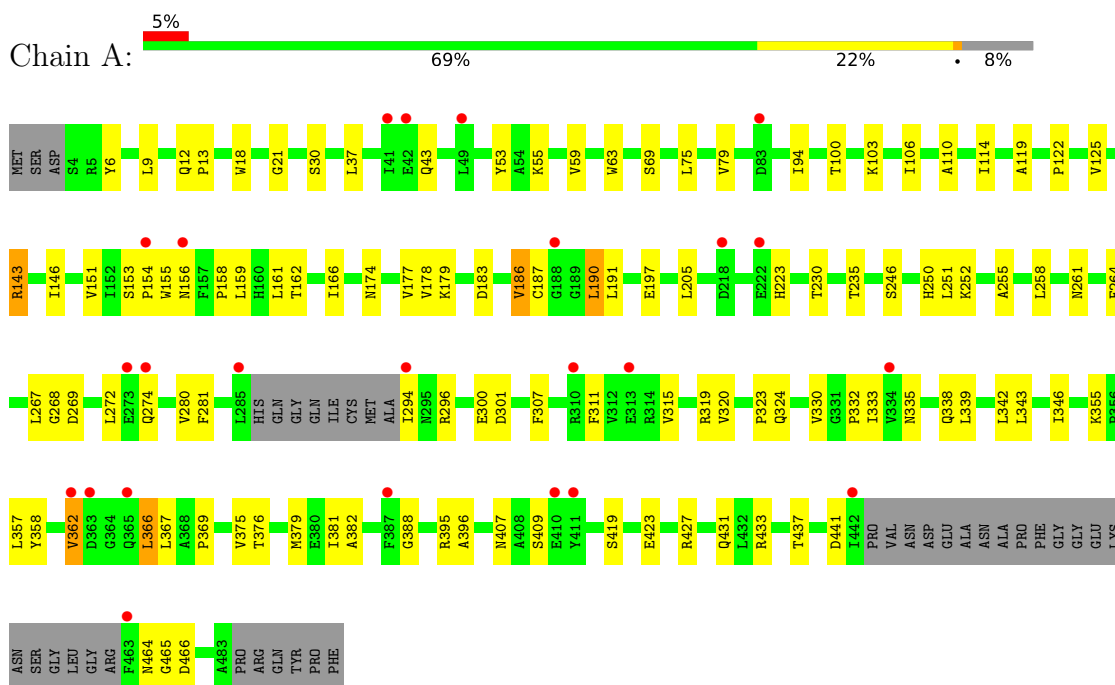
- Molecule 3 is water.

3	A	11	Total 11	O 11	0	0
3	B	22	Total 22	O 22	0	0
3	C	25	Total 25	O 25	0	0
3	D	14	Total 14	O 14	0	0

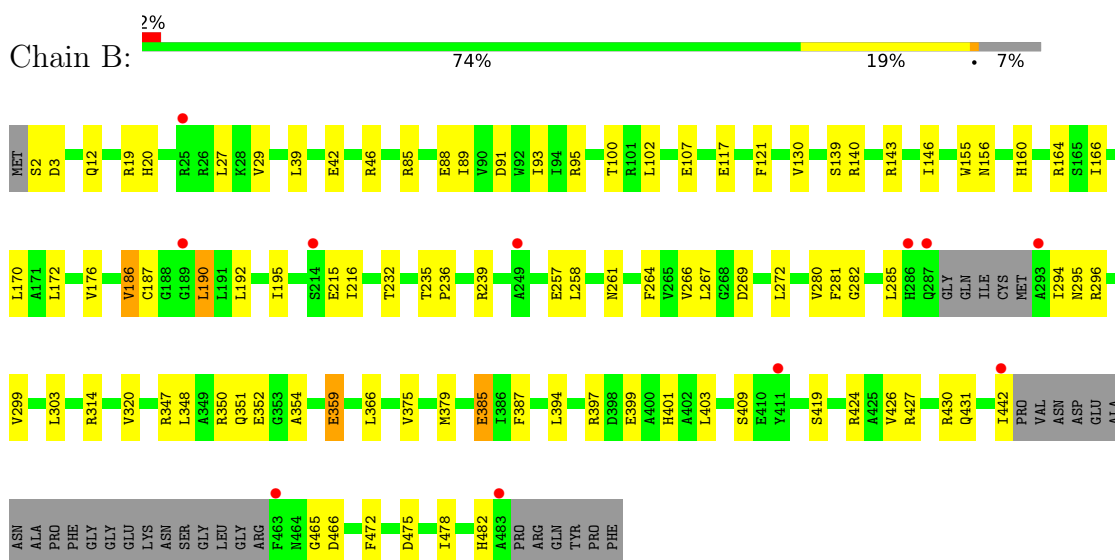
### 3 Residue-property plots [i](#)

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

#### • Molecule 1: Aldehyde dehydrogenase



#### • Molecule 1: Aldehyde dehydrogenase



Chain C:

Category	Value
Green	72%
Yellow	19%
Red	3%
Grey	7%

ALA	ASN	ASP	GLU	ALA	ASN	ALA	ALA	PRO	PHE	GLY	GLY	GLU	LYS	ASN	ASN	SER	GLY	L460	G461	R462	F463	G464	G465	D466	I469	E470	E471	F472	T473	T474	D475	A483	PRO	ARG	GLN	TYR	PRO	PHE							
I294	M156	F157	P158	L159	H160	L161	I166	L172	A175	V176	V177	V178	D183	V186	C187	G188	G189	L190	L191	G211	T216	N240	T244	H250	L251	K252	N261	V266	D269	A270	D271	L272	E273	F281	G282	K283	T284	L285	HIS	GLN	GLY	ILE	CYS	PRO	VAL

[illegible]

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	102.15Å 144.77Å 150.24Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.36 – 2.88 48.36 – 2.88	Depositor EDS
% Data completeness (in resolution range)	99.9 (48.36-2.88) 99.8 (48.36-2.88)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.54 (at 2.86Å)	Xtriage
Refinement program	PHENIX (1.21_5207: ???)	Depositor
R, $R_{free}$	0.207 , 0.260 0.214 , 0.266	Depositor DCC
$R_{free}$ test set	48579 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	55.4	Xtriage
Anisotropy	0.389	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 61.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.001 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	13700	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.57% 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: NAP

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	0.42	0/3309	0.65	0/4526
1	B	0.26	0/3482	0.48	2/4740 (0.0%)
1	C	0.33	0/3472	0.59	3/4721 (0.1%)
1	D	0.23	0/3490	0.45	0/4745
All	All	0.32	0/13753	0.55	5/18732 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	463	PHE	CA-CB-CG	7.48	121.28	113.80
1	C	463	PHE	N-CA-CB	6.05	119.02	110.12
1	C	410	GLU	N-CA-C	-5.27	102.58	110.23
1	B	187	CYS	CA-C-N	-5.25	117.61	122.29
1	B	187	CYS	C-N-CA	-5.25	117.61	122.29

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	3239	0	3089	76	0
1	B	3411	0	3338	59	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	3402	0	3349	58	0
1	D	3418	0	3381	53	0
2	A	31	0	11	0	0
2	B	31	0	11	0	0
2	C	48	0	25	2	0
2	D	48	0	25	1	0
3	A	11	0	0	0	0
3	B	22	0	0	1	0
3	C	25	0	0	1	0
3	D	14	0	0	0	0
All	All	13700	0	13229	240	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 9.

The worst 5 of 240 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:D:354:ALA:HB2	1:D:379:MET:HE1	1.48	0.95
1:A:186:VAL:HA	1:A:190:LEU:HB2	1.53	0.91
1:A:342:LEU:HD21	1:A:388:GLY:HA3	1.56	0.88
1:D:164:ARG:HH12	1:D:463:PHE:HB2	1.47	0.80
1:A:343:LEU:HD13	1:A:346:ILE:HD11	1.64	0.80

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	446/489 (91%)	436 (98%)	10 (2%)	0	100	100
1	B	451/489 (92%)	442 (98%)	9 (2%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	448/489 (92%)	437 (98%)	11 (2%)	0	100	100
1	D	446/489 (91%)	437 (98%)	9 (2%)	0	100	100
All	All	1791/1956 (92%)	1752 (98%)	39 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	A	306/377 (81%)	294 (96%)	12 (4%)	27	59
1	B	339/377 (90%)	328 (97%)	11 (3%)	34	66
1	C	339/377 (90%)	319 (94%)	20 (6%)	16	42
1	D	344/377 (91%)	338 (98%)	6 (2%)	56	81
All	All	1328/1508 (88%)	1279 (96%)	49 (4%)	29	61

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	144	SER
1	C	413	LEU
1	C	183	ASP
1	C	271	ASP
1	C	460	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	482	HIS
1	D	274	GLN
1	D	482	HIS
1	D	440	ASN
1	B	351	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

4 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	$\# Z  > 2$
2	NAP	A	501	-	27,33,52	2.97	5 (18%)	35,52,80	2.10	10 (28%)
2	NAP	D	501	-	45,52,52	2.70	11 (24%)	56,80,80	2.14	15 (26%)
2	NAP	B	501	-	27,33,52	3.05	5 (18%)	35,52,80	2.05	10 (28%)
2	NAP	C	501	-	45,52,52	2.79	12 (26%)	56,80,80	2.12	17 (30%)

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	NAP	A	501	-	-	4/17/37/67	0/3/3/5
2	NAP	D	501	-	-	19/31/67/67	0/5/5/5
2	NAP	B	501	-	-	5/17/37/67	0/3/3/5
2	NAP	C	501	-	-	14/31/67/67	0/5/5/5

The worst 5 of 33 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	501	NAP	P2B-O2B	13.54	1.84	1.59
2	C	501	NAP	P2B-O2B	13.49	1.84	1.59
2	A	501	NAP	P2B-O2B	12.95	1.83	1.59
2	D	501	NAP	P2B-O2B	12.90	1.83	1.59
2	C	501	NAP	C4N-C3N	5.90	1.49	1.39

The worst 5 of 52 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	501	NAP	C5N-C4N-C3N	-7.71	111.22	120.34
2	A	501	NAP	PA-O3-PN	-7.10	108.45	132.83
2	C	501	NAP	C5N-C4N-C3N	-6.97	112.10	120.34
2	C	501	NAP	PN-O3-PA	-6.88	109.23	132.83
2	B	501	NAP	PA-O3-PN	-6.84	109.35	132.83

There are no chirality outliers.

5 of 42 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	501	NAP	C5B-O5B-PA-O3
2	A	501	NAP	PN-O3-PA-O5B
2	B	501	NAP	C5B-O5B-PA-O3
2	B	501	NAP	PN-O3-PA-O5B
2	B	501	NAP	O4B-C4B-C5B-O5B

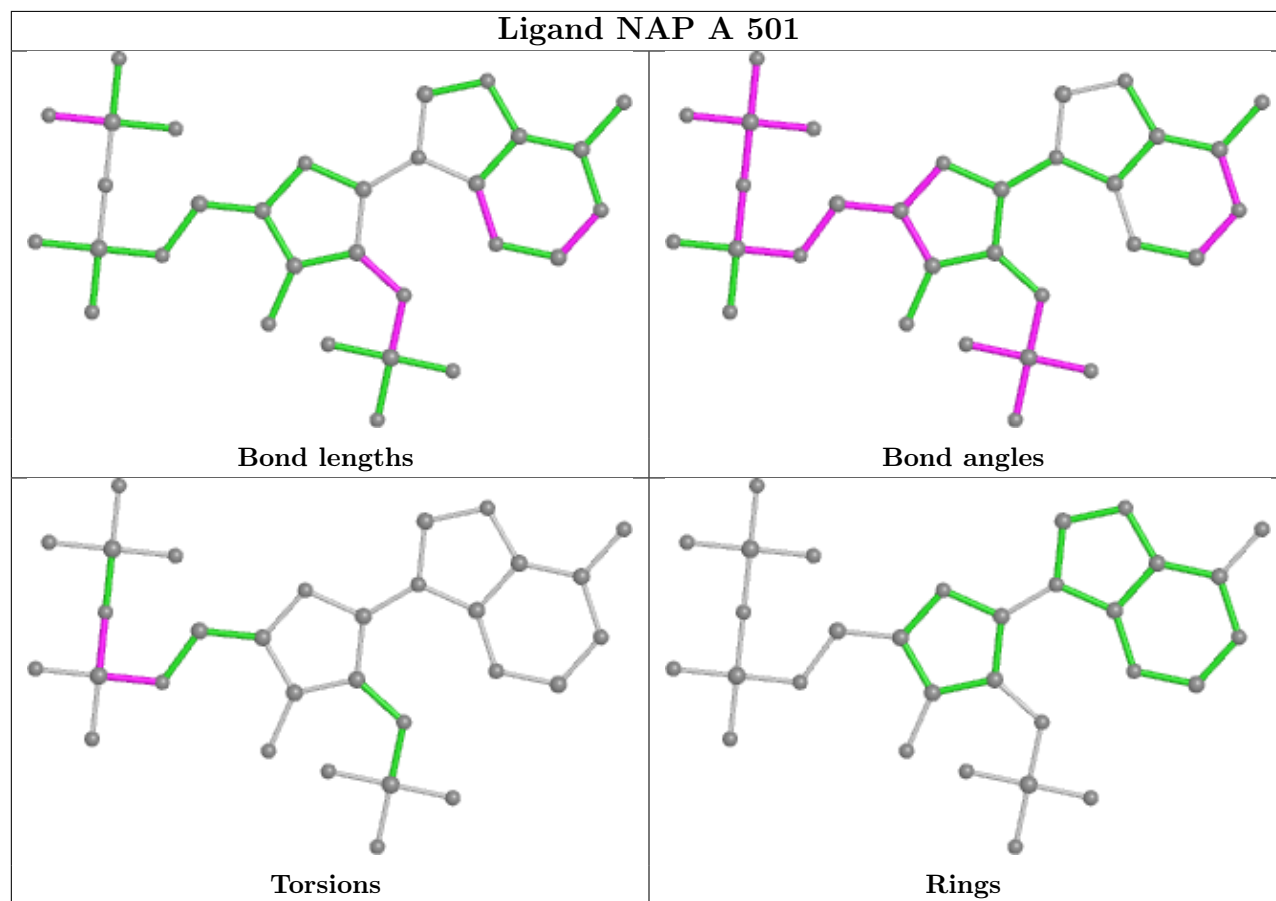
There are no ring outliers.

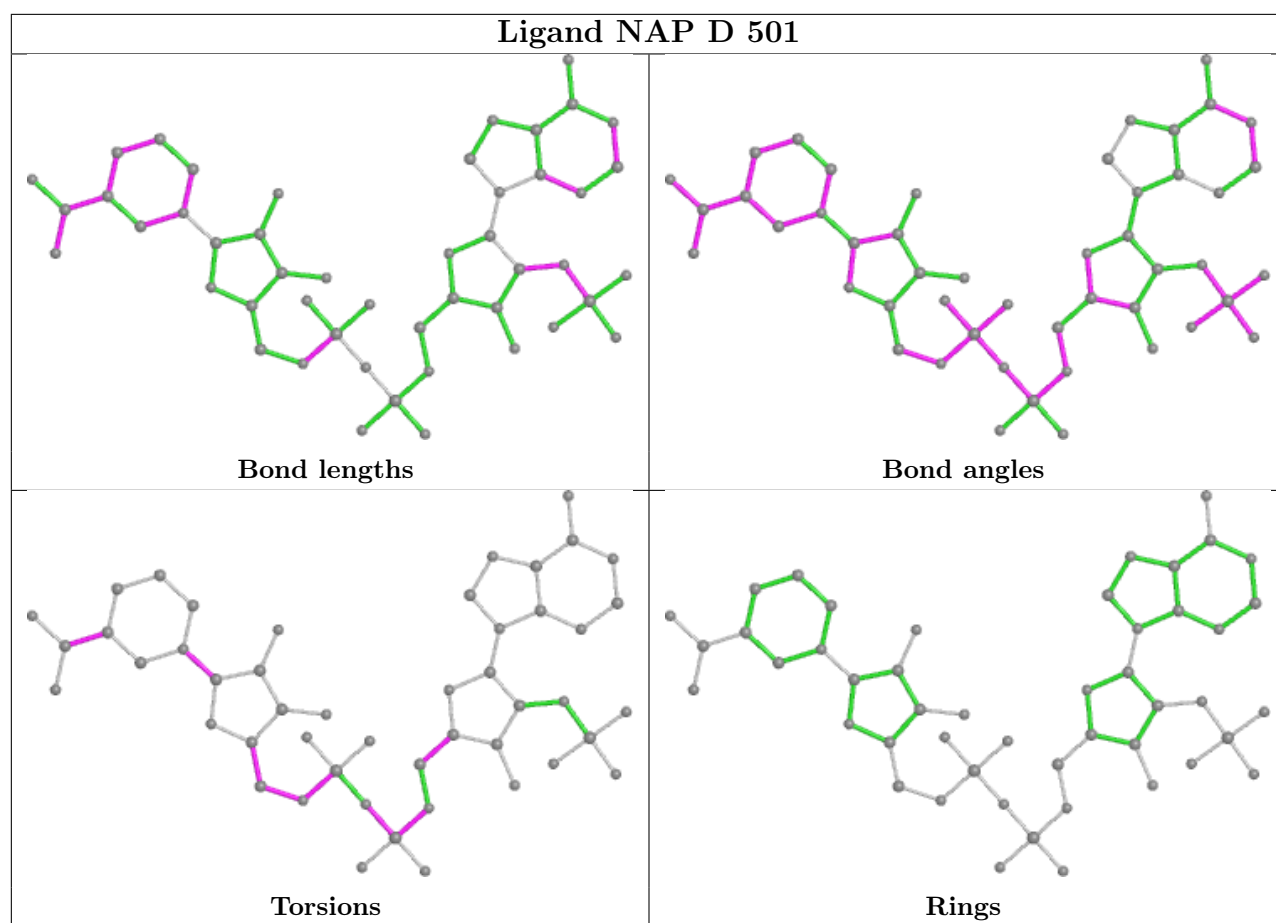
2 monomers are involved in 3 short contacts:

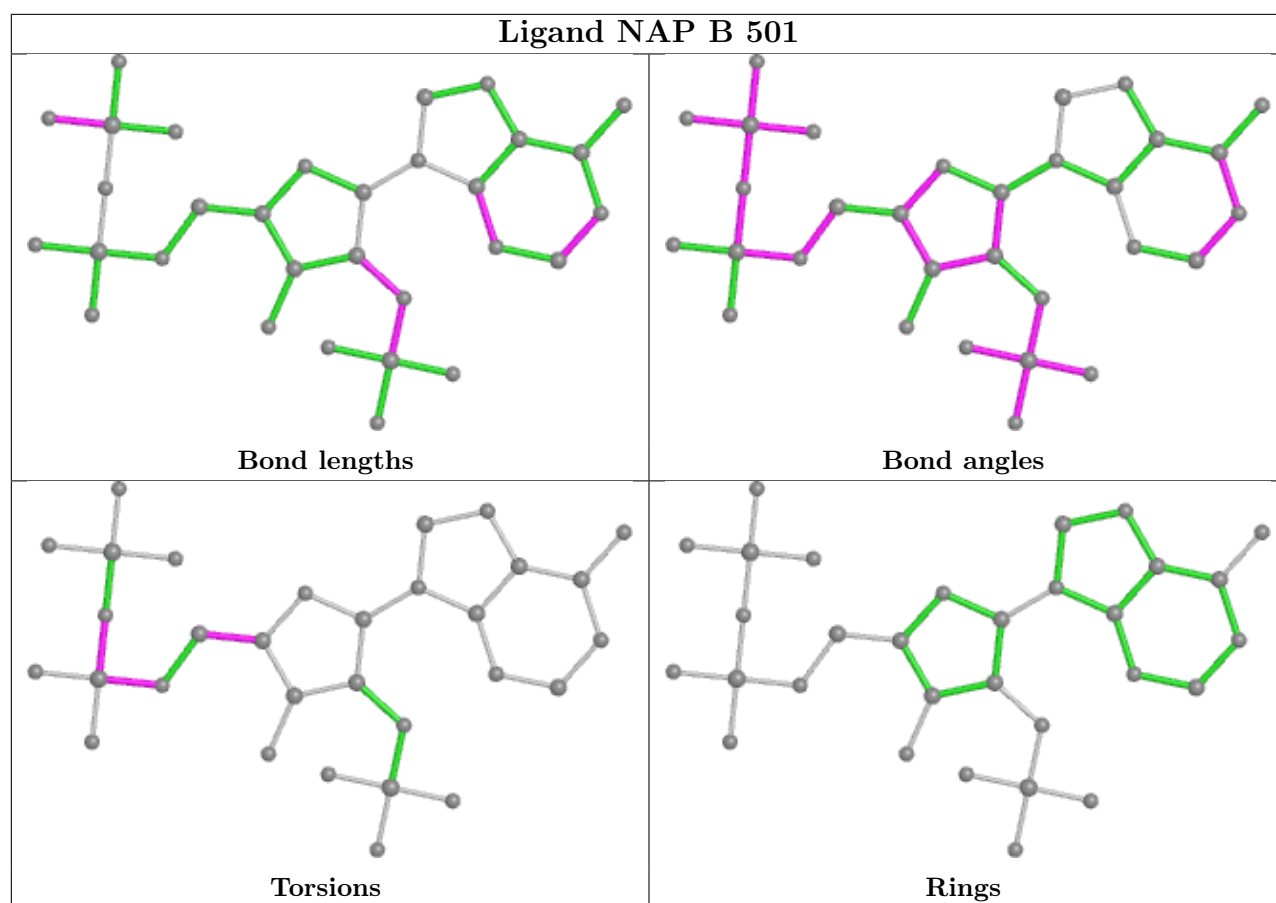
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	501	NAP	1	0
2	C	501	NAP	2	0

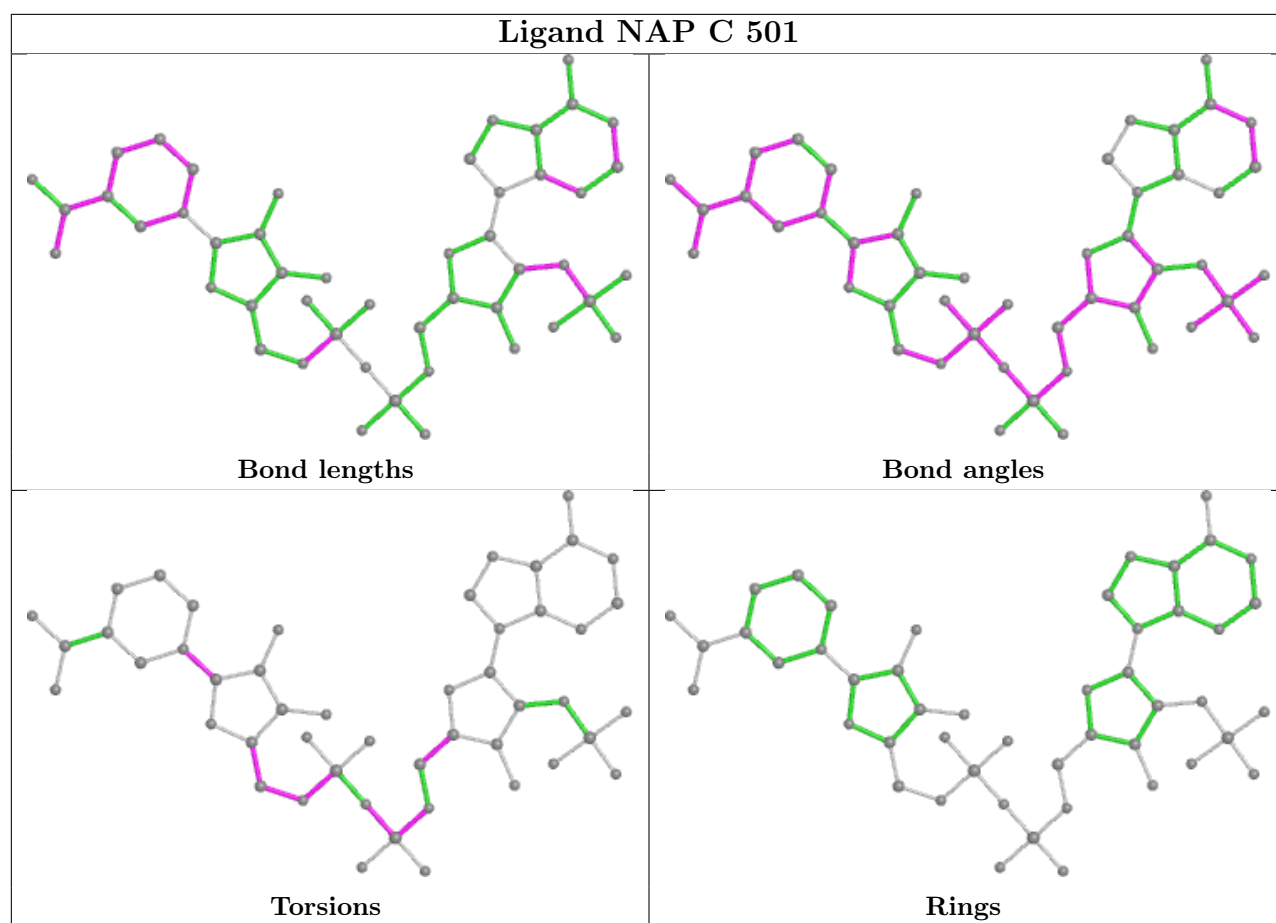
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and

any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.









## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.



## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ > 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q < 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	452/489 (92%)	0.45	24 (5%) 33 28	42, 75, 98, 114	0
1	B	457/489 (93%)	-0.14	11 (2%) 59 54	29, 45, 84, 117	0
1	C	454/489 (92%)	-0.05	16 (3%) 47 41	30, 49, 76, 111	0
1	D	452/489 (92%)	-0.17	6 (1%) 74 70	29, 52, 76, 125	0
All	All	1815/1956 (92%)	0.02	57 (3%) 51 46	29, 54, 93, 125	0

The worst 5 of 57 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	285	LEU	5.3
1	C	463	PHE	5.0
1	C	410	GLU	4.6
1	B	293	ALA	4.3
1	A	411	TYR	4.0

### 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](#)

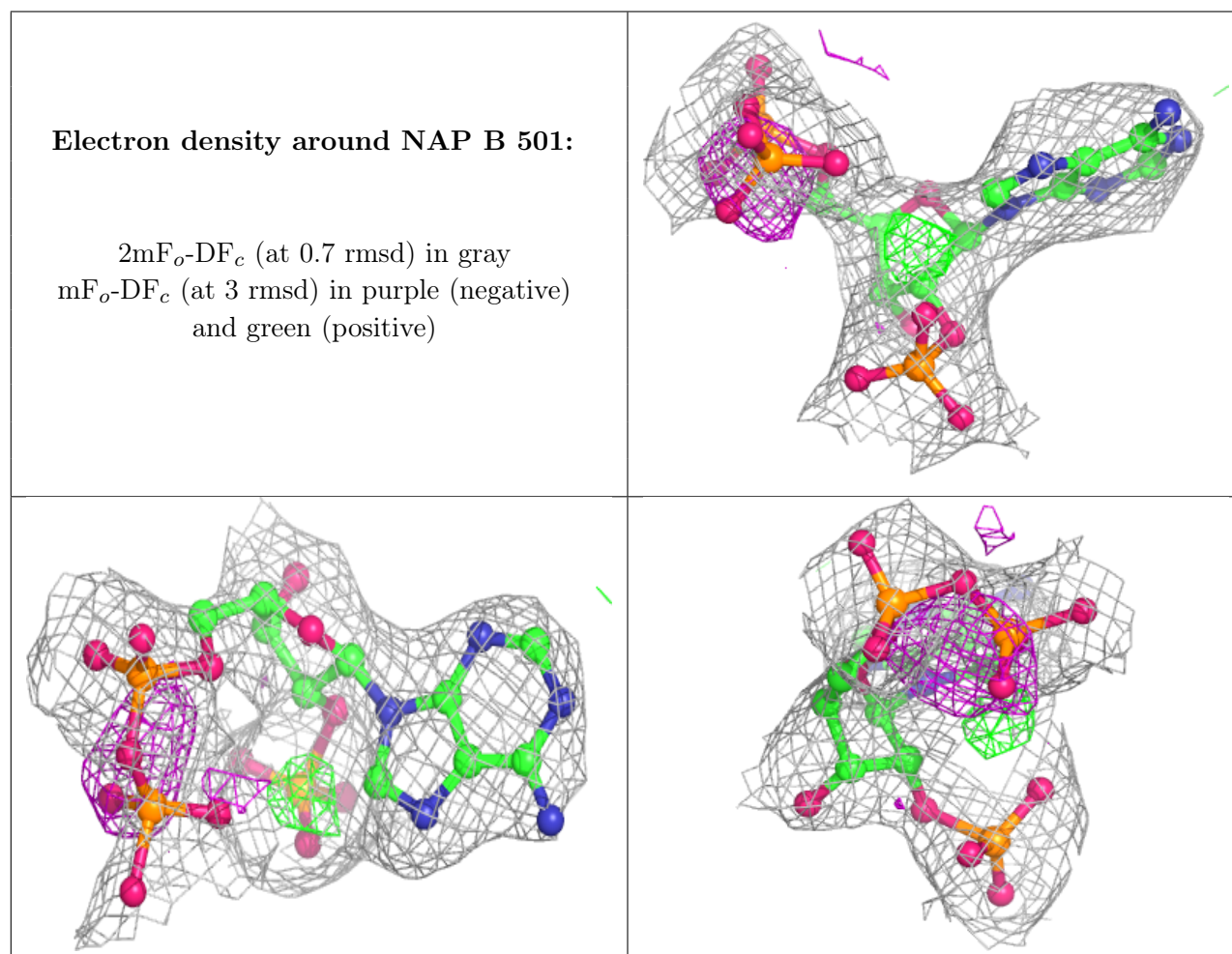
There are no oligosaccharides in this entry.

### 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.

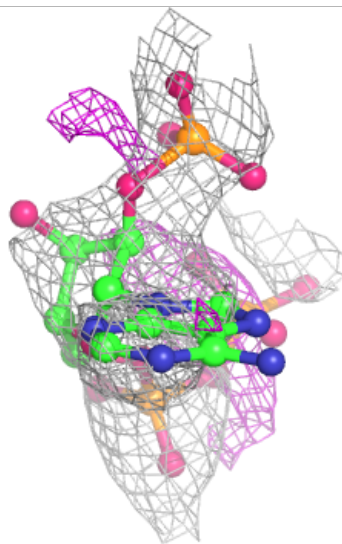
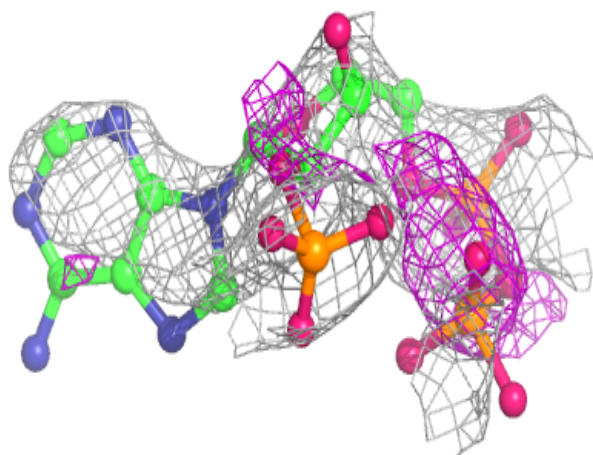
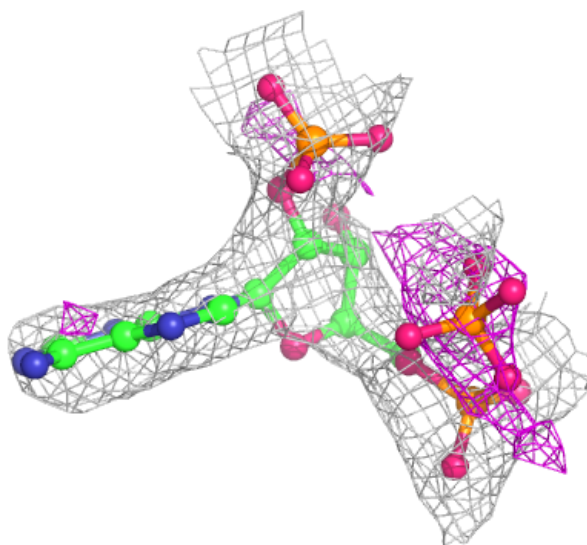
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	NAP	B	501	31/48	0.70	0.15	78,94,137,160	0
2	NAP	A	501	31/48	0.74	0.13	88,120,159,186	0
2	NAP	C	501	48/48	0.79	0.15	55,99,130,134	0
2	NAP	D	501	48/48	0.92	0.10	32,70,125,135	0

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.



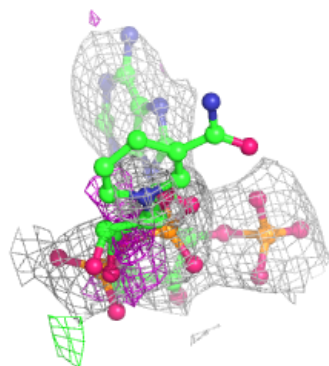
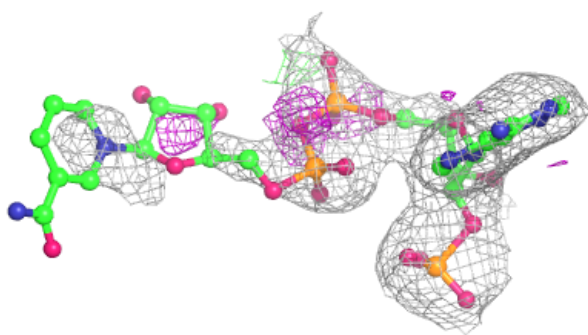
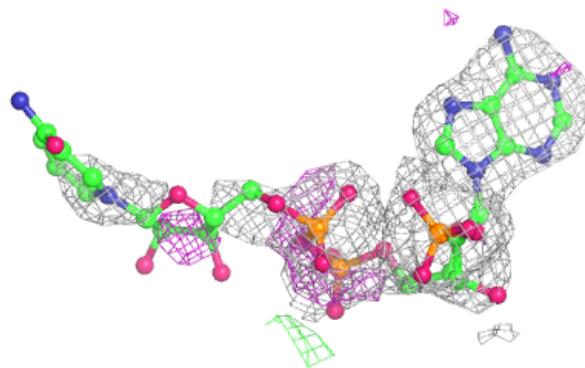
**Electron density around NAP A 501:**

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

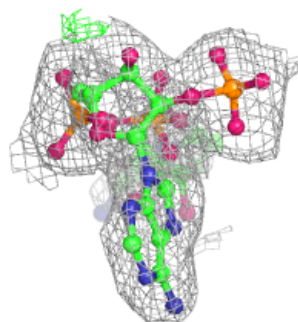
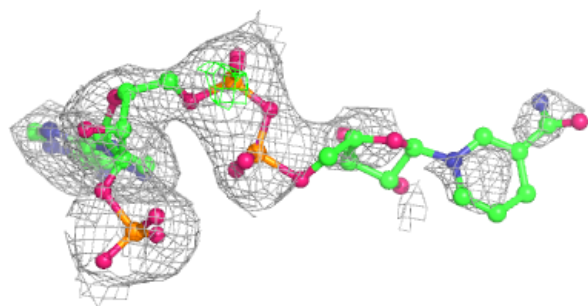
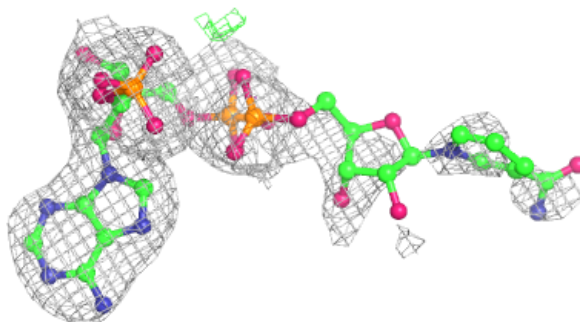


**Electron density around NAP C 501:**

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

**Electron density around NAP D 501:**

$2mF_o-DF_c$  (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.