WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jun 15, 2024 – 05:25 PM EDT

PDB ID	:	4R1F
Title	:	Re-refined Human DNA topoisomerase IIa (ATPase and transducer domains) in complex with ADP and SO4
Authors	:	Stanger, F.V.; Schirmer, T.
Deposited on	:	2014-08-05
Resolution	:	2.51 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

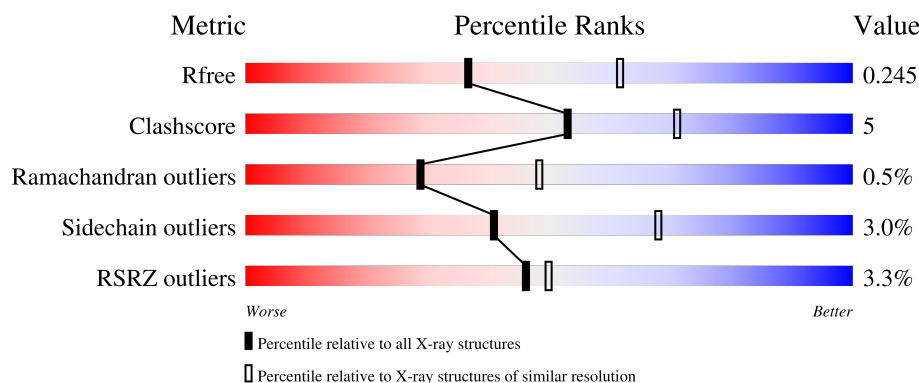|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.51 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 <=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	400	4% / 80% 8% · 8%
1	B	400	3% / 86% 8% · 5%
1	C	400	87% 8% ·
1	D	400	5% / 80% 12% · 7%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	SO4	B	502	-	-	X	-
5	SO4	C	503	-	-	X	-

2   Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 12360 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.
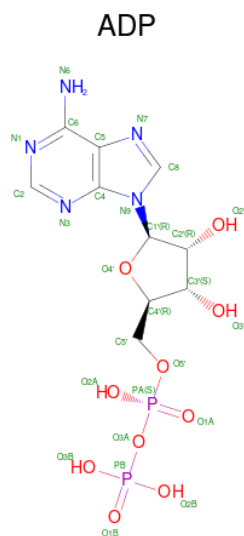
- Molecule 1 is a protein called DNA topoisomerase 2-alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	366	Total	C	N	O	S	0	0	0
			2955	1890	499	548	18			
1	B	379	Total	C	N	O	S	0	0	0
			3053	1952	518	564	19			
1	C	384	Total	C	N	O	S	0	0	0
			3080	1970	525	566	19			
1	D	373	Total	C	N	O	S	0	0	0
			3005	1922	510	554	19			

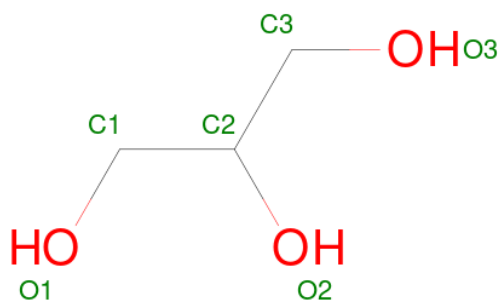- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 3 is ADENOSINE-5'-DIPHOSPHATE (three-letter code: ADP) (formula: $C_{10}H_{15}N_5O_{10}P_2$).

ADP

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
3	B	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
3	C	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
3	D	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total 6	C 3	O 3	0	0

• Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O$_4$S).
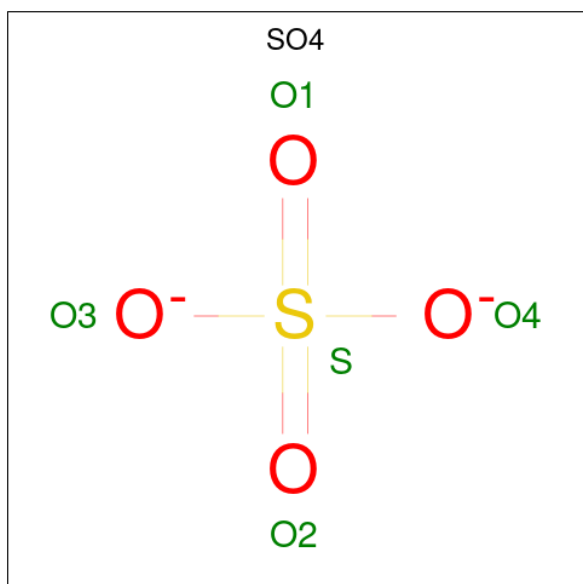
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total 5	O 4	S 1	0	0
5	C	1	Total 5	O 4	S 1	0	0
5	D	1	Total 5	O 4	S 1	0	0

• Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	37	Total 37	O 37	0	0
6	B	46	Total 46	O 46	0	0
6	C	40	Total 40	O 40	0	0
6	D	11	Total 11	O 11	0	0

3 Residue-property plots ⓘ

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 topoisomerase 2-alpha

Chain A:

- Molecule 1: DNA topoisomerase 2-alpha

Chain B:

- Molecule 1: DNA topoisomerase 2-alpha

Chain C:

- Molecule 1: DNA topoisomerase 2-alpha

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	71.02Å 90.49Å 148.29Å 90.00° 89.97° 90.00°	Depositor
Resolution (Å)	30.00 – 2.51 40.58 – 2.51	Depositor EDS
% Data completeness (in resolution range)	94.9 (30.00-2.51) 94.9 (40.58-2.51)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.80 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.208 , 0.243 0.212 , 0.245	Depositor DCC
R_{free} test set	4851 reflections (7.61%)	wwPDB-VP
Wilson B-factor (Å^2)	56.6	Xtriage
Anisotropy	0.163	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 39.1	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.023 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	12360	wwPDB-VP
Average B, all atoms (Å^2)	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.38% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: GOL, MG, ADP, SO4

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	0.65	0/3013	0.74	3/4061 (0.1%)
1	B	0.66	0/3112	0.72	3/4192 (0.1%)
1	C	0.61	0/3139	0.71	2/4228 (0.0%)
1	D	0.57	0/3063	0.71	2/4125 (0.0%)
All	All	0.62	0/12327	0.72	10/16606 (0.1%)

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
1	A	0	2

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	172	ILE	CG1-CB-CG2	-6.17	97.82	111.40
1	D	368	ILE	CG1-CB-CG2	-5.87	98.49	111.40
1	A	399	ILE	CG1-CB-CG2	-5.83	98.57	111.40
1	C	374	ASP	CB-CA-C	-5.50	99.40	110.40
1	A	135	MET	CG-SD-CE	5.25	108.60	100.20

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	409	GLU	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	416	LYS	Peptide

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	2955	0	2963	38	0
1	B	3053	0	3081	19	0
1	C	3080	0	3119	30	0
1	D	3005	0	3028	42	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	27	0	12	0	0
3	B	27	0	12	0	0
3	C	27	0	12	2	0
3	D	27	0	12	1	0
4	A	6	0	8	0	0
5	B	5	0	0	2	0
5	C	5	0	0	2	0
5	D	5	0	0	0	0
6	A	37	0	0	0	0
6	B	46	0	0	1	0
6	C	40	0	0	0	0
6	D	11	0	0	0	0
All	All	12360	0	12247	131	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 5.

The worst 5 of 131 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:336:LYS:HD2	1:A:399:ILE:HG22	1.20	1.11
1:A:336:LYS:CD	1:A:399:ILE:HG22	1.81	1.11

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:336:LYS:HD2	1:A:399:ILE:CG2	1.83	1.07
1:A:395:SER:O	1:A:399:ILE:HG13	1.59	1.02
1:D:171:ASN:HB3	1:D:198:MET:HE3	1.48	0.95

There are no symmetry-related clashes.

5.3　Torsion angles ⓘ

5.3.1　Protein backbone ⓘ

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	358/400 (90%)	335 (94%)	20 (6%)	3 (1%)	19	35
1	B	373/400 (93%)	352 (94%)	17 (5%)	4 (1%)	14	26
1	C	378/400 (94%)	365 (97%)	13 (3%)	0	100	100
1	D	365/400 (91%)	348 (95%)	16 (4%)	1 (0%)	41	61
All	All	1474/1600 (92%)	1400 (95%)	66 (4%)	8 (0%)	29	48

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	416	LYS
1	B	353	ALA
1	B	407	ILE
1	A	410	SER
1	A	417	PHE

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	330/358 (92%)	316 (96%)	14 (4%)	30	54
1	B	341/358 (95%)	332 (97%)	9 (3%)	46	72
1	C	343/358 (96%)	339 (99%)	4 (1%)	71	88
1	D	335/358 (94%)	322 (96%)	13 (4%)	32	57
All	All	1349/1432 (94%)	1309 (97%)	40 (3%)	41	68

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	189	MET
1	D	374	ASP
1	D	210	ASN
1	D	265	LYS
1	D	403	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	163	ASN
1	D	59	GLN
1	D	210	ASN
1	D	171	ASN
1	C	59	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 12 ligands modelled in this entry, 4 are monoatomic - leaving 8 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	ADP	B	503	2	24,29,29	1.08	1 (4%)	29,45,45	1.89	5 (17%)
5	SO4	C	503	-	4,4,4	0.25	0	6,6,6	0.40	0
3	ADP	C	502	2	24,29,29	1.04	2 (8%)	29,45,45	1.66	6 (20%)
4	GOL	A	503	-	5,5,5	0.65	0	5,5,5	0.90	0
3	ADP	A	502	2	24,29,29	1.24	2 (8%)	29,45,45	1.58	6 (20%)
5	SO4	D	502	-	4,4,4	0.33	0	6,6,6	0.25	0
3	ADP	D	503	2	24,29,29	1.12	2 (8%)	29,45,45	1.75	7 (24%)
5	SO4	B	502	-	4,4,4	0.55	0	6,6,6	0.35	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
3	ADP	B	503	2	-	8/12/32/32	0/3/3/3
3	ADP	C	502	2	-	2/12/32/32	0/3/3/3
4	GOL	A	503	-	-	2/4/4/4	-
3	ADP	A	502	2	-	6/12/32/32	0/3/3/3
3	ADP	D	503	2	-	2/12/32/32	0/3/3/3

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	502	ADP	C5-C4	2.80	1.48	1.40
3	D	503	ADP	O4'-C1'	2.72	1.44	1.41
3	D	503	ADP	C5-C4	2.63	1.47	1.40
3	A	502	ADP	O4'-C1'	2.46	1.44	1.41
3	C	502	ADP	O5'-C5'	-2.35	1.35	1.44

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	502	ADP	C3'-C2'-C1'	4.80	108.21	100.98
3	B	503	ADP	N3-C2-N1	-4.38	121.83	128.68
3	B	503	ADP	C3'-C2'-C1'	4.37	107.56	100.98
3	B	503	ADP	C2-N1-C6	3.73	125.14	118.75
3	D	503	ADP	C5'-C4'-C3'	-3.70	101.31	115.18

There are no chirality outliers.

5 of 20 torsion outliers are listed below:

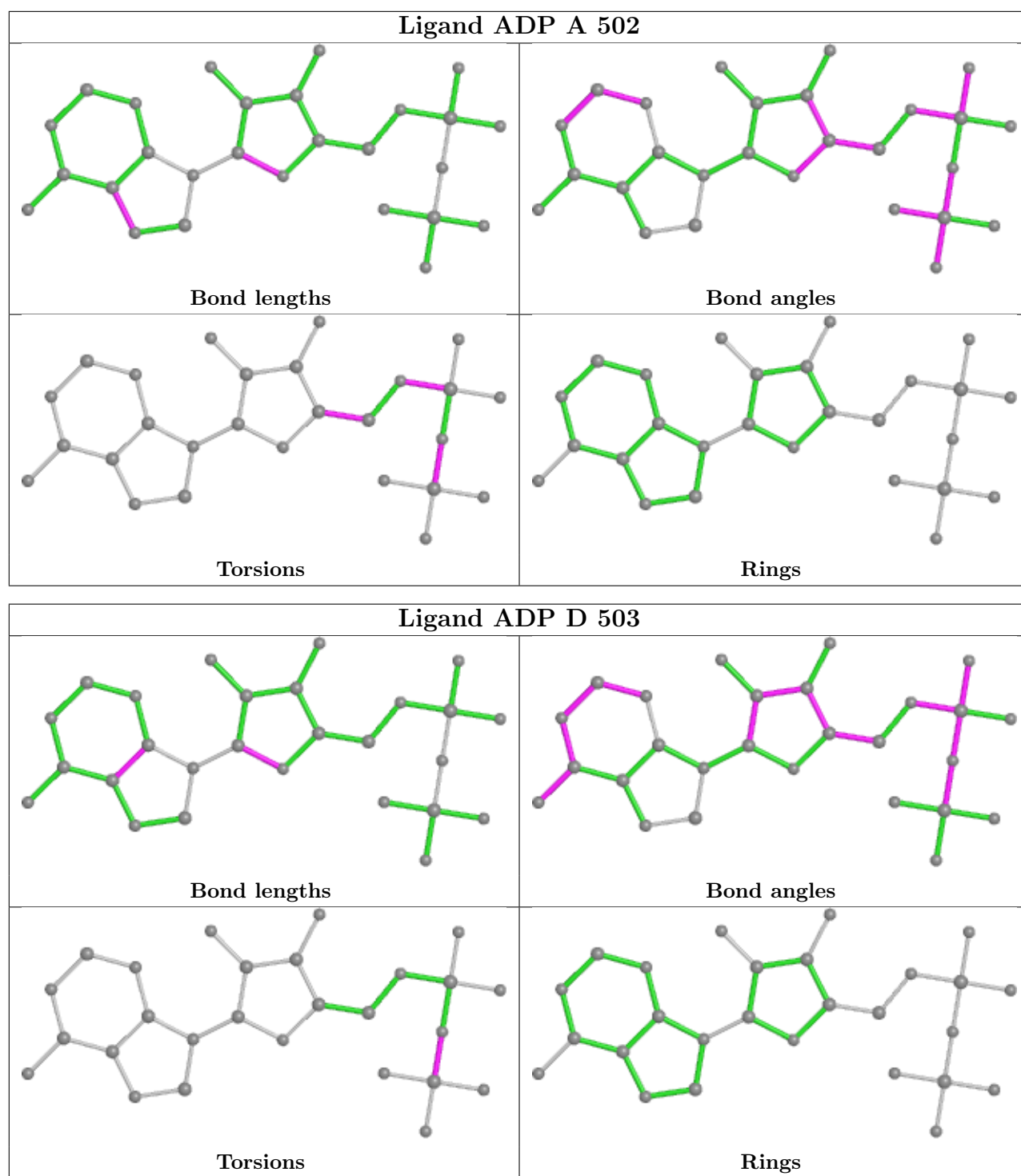| Mol	Chain	Res	Type	Atoms
3	A	502	ADP	C5'-O5'-PA-O1A
3	A	502	ADP	C5'-O5'-PA-O2A
3	B	503	ADP	C5'-O5'-PA-O1A
3	B	503	ADP	O4'-C4'-C5'-O5'
3	C	502	ADP	C3'-C4'-C5'-O5'

There are no ring outliers.

4 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	C	503	SO4	2	0
3	C	502	ADP	2	0
3	D	503	ADP	1	0
5	B	502	SO4	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 then 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.

Ligand ADP B 503

Bond lengths

Bond angles

Torsions

Rings

Ligand ADP C 502

Bond lengths

Bond angles

Torsions

Rings

Ligand ADP A 502

Bond lengths

Bond angles

Torsions

Rings

Ligand ADP D 503

Bond lengths

Bond angles

Torsions

Rings
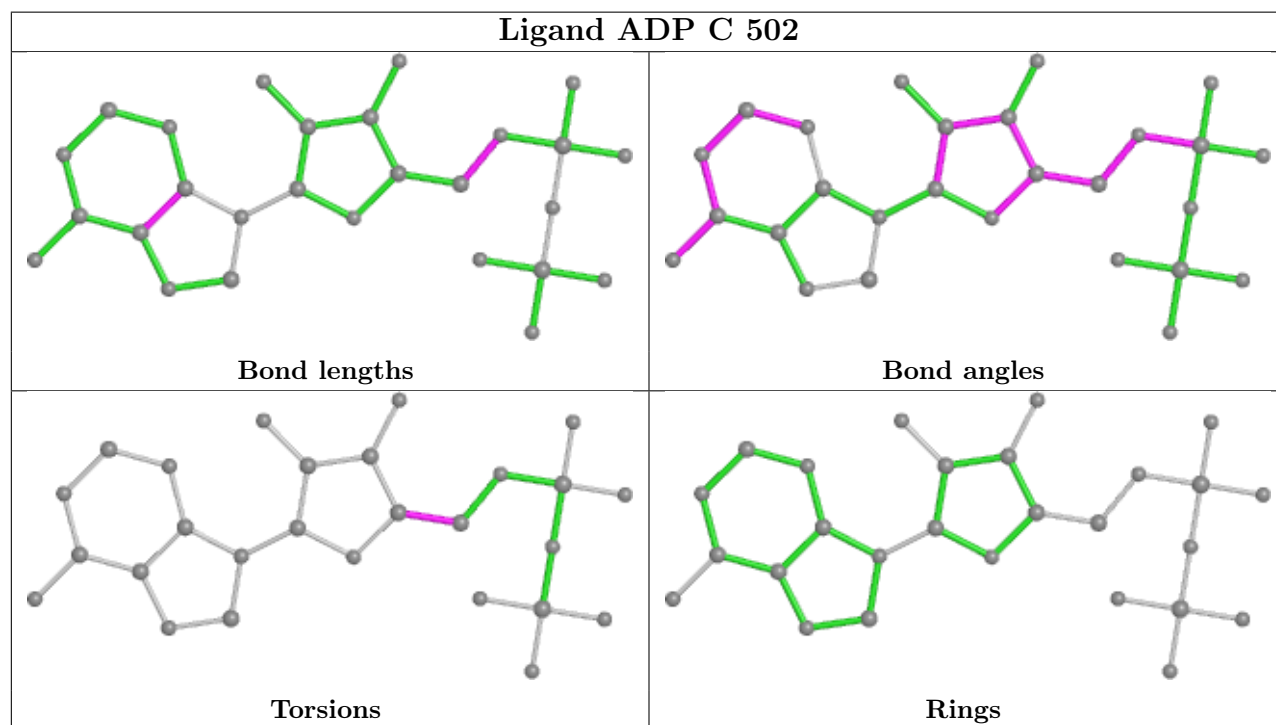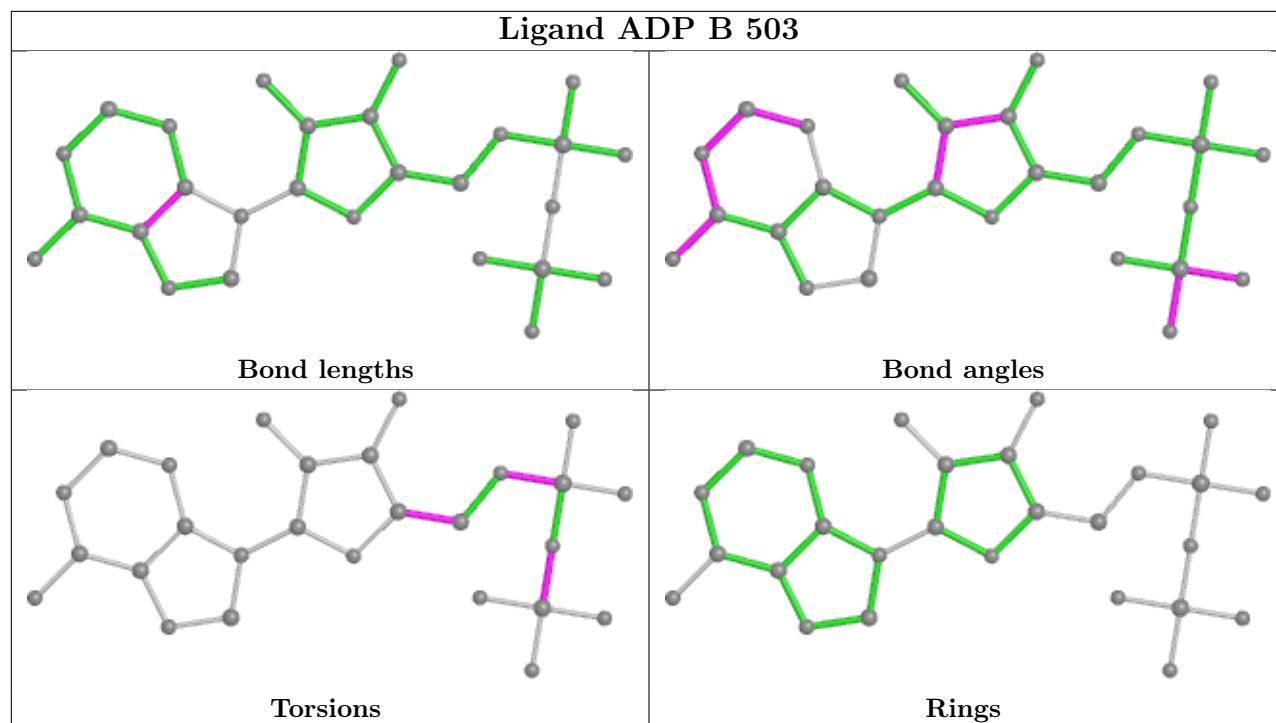
5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 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($\mathring{A}^2$)	Q<0.9
1	A	366/400 (91%)	0.19	18 (4%)	29	31	37, 64, 95, 117	0
1	B	379/400 (94%)	0.12	11 (2%)	51	55	35, 55, 88, 111	0
1	C	384/400 (96%)	-0.08	2 (0%)	91	91	39, 56, 80, 98	0
1	D	373/400 (93%)	0.34	19 (5%)	28	29	47, 71, 100, 117	0
All	All	1502/1600 (93%)	0.14	50 (3%)	46	50	35, 61, 93, 117	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	341	VAL	5.2
1	A	414	TRP	5.0
1	D	155	GLU	3.8
1	D	407	ILE	3.6
1	A	336	LYS	3.6

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

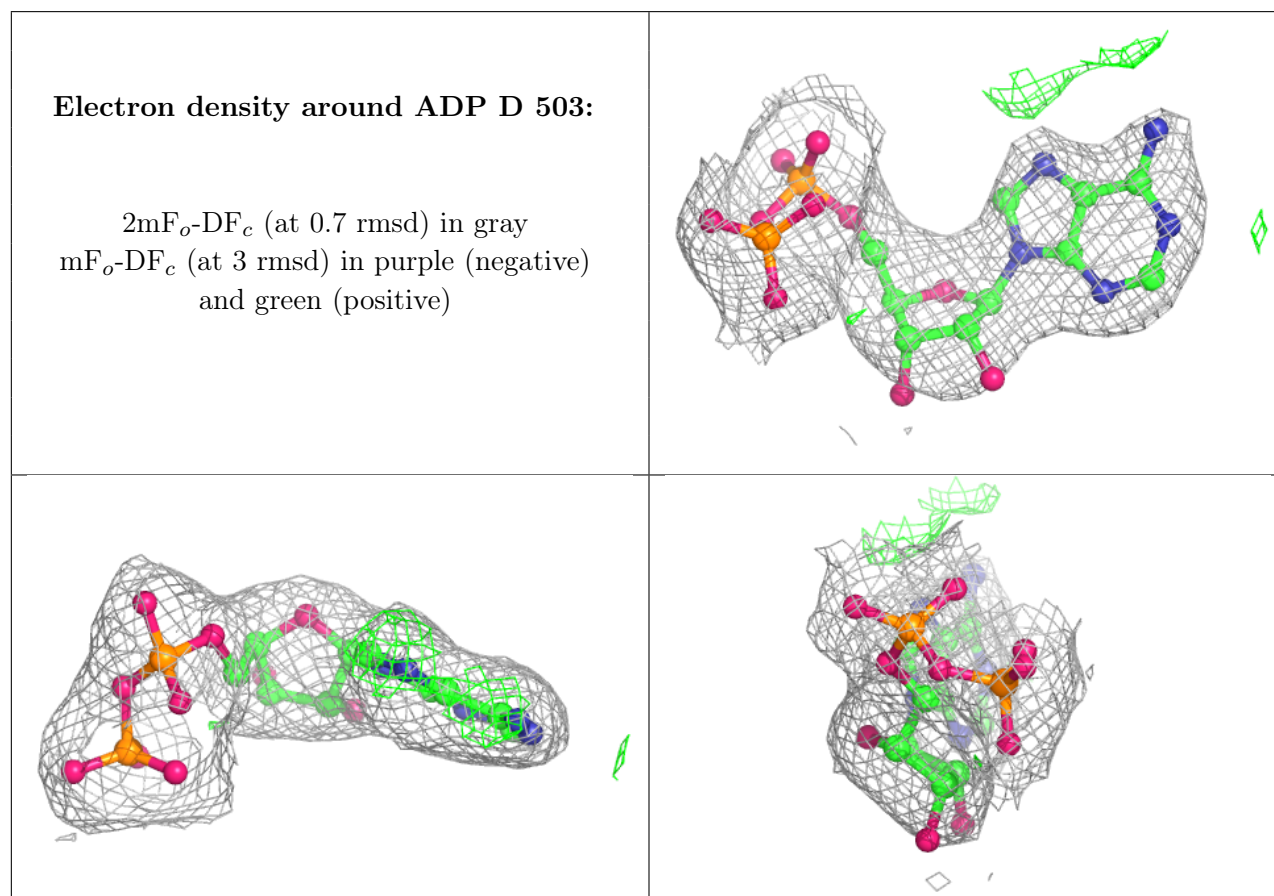There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides 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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	GOL	A	503	6/6	0.88	0.30	59,66,71,76	0
5	SO4	B	502	5/5	0.95	0.16	67,76,82,92	0
2	MG	D	501	1/1	0.96	0.23	50,50,50,50	0
5	SO4	D	502	5/5	0.96	0.17	93,97,112,112	0
5	SO4	C	503	5/5	0.97	0.12	74,81,82,83	0
2	MG	B	501	1/1	0.97	0.20	40,40,40,40	0
3	ADP	D	503	27/27	0.98	0.20	46,60,73,99	0
2	MG	C	501	1/1	0.98	0.18	45,45,45,45	0
2	MG	A	501	1/1	0.98	0.16	43,43,43,43	0
3	ADP	A	502	27/27	0.98	0.15	40,49,57,65	0
3	ADP	C	502	27/27	0.98	0.14	42,49,55,57	0
3	ADP	B	503	27/27	0.99	0.18	33,39,47,50	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.
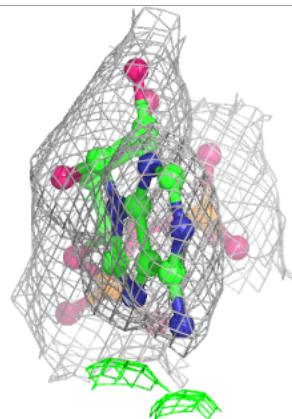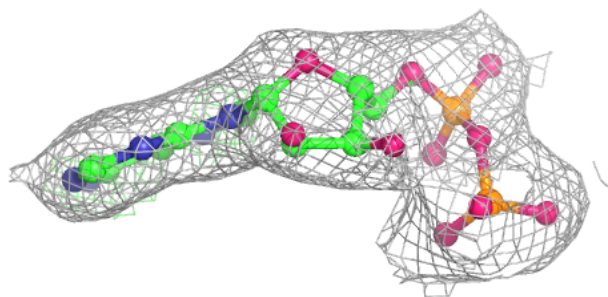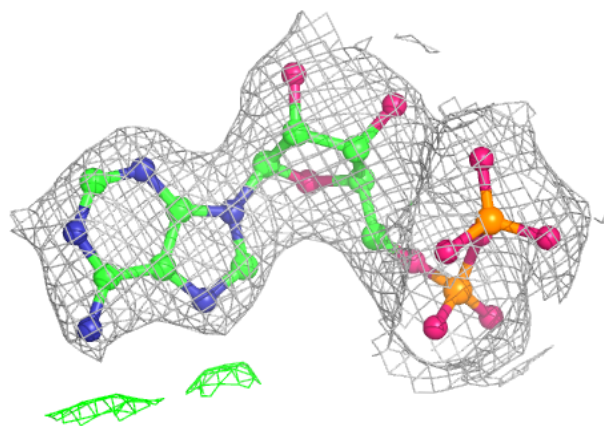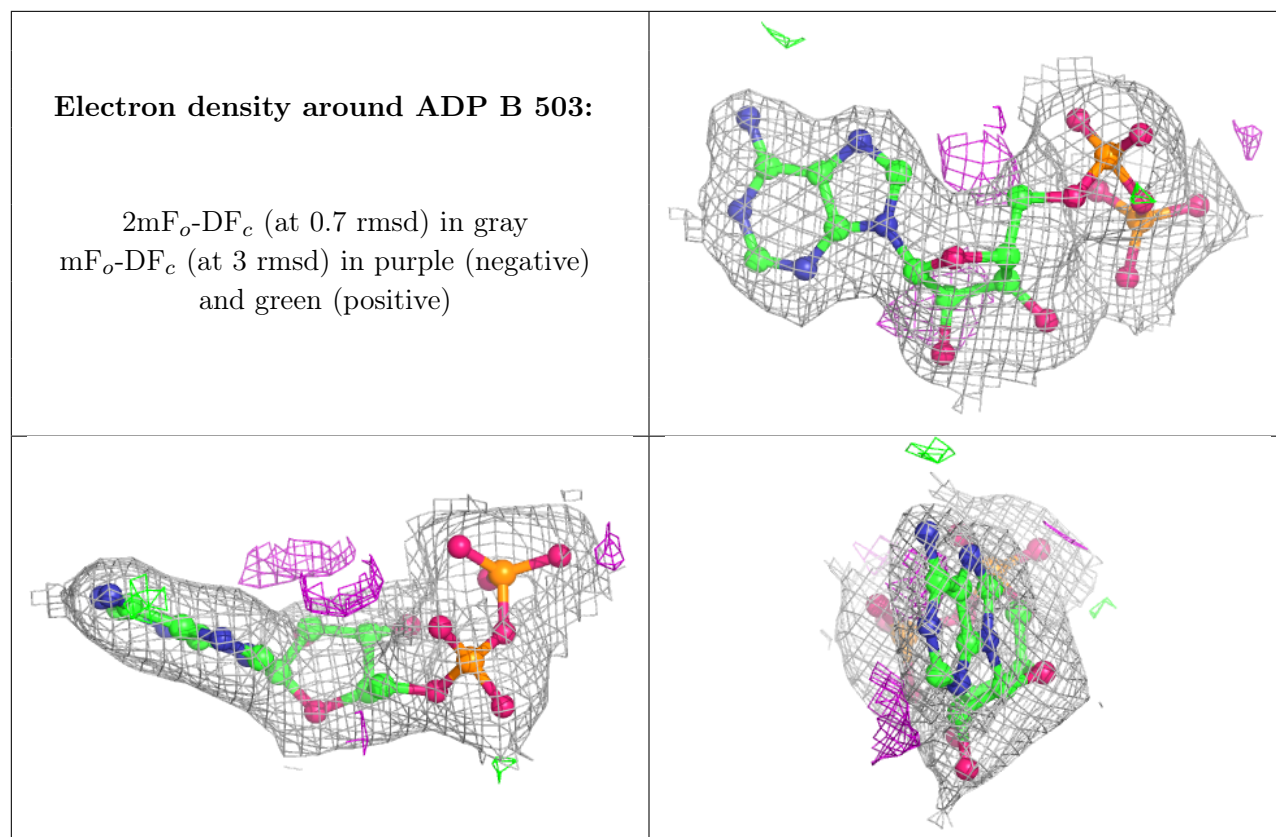

Electron density around ADP D 503:

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 ADP A 502:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)
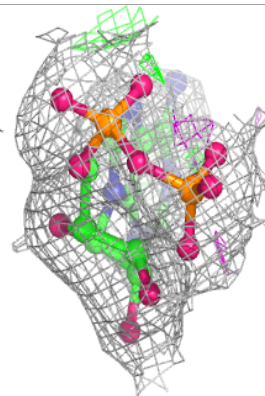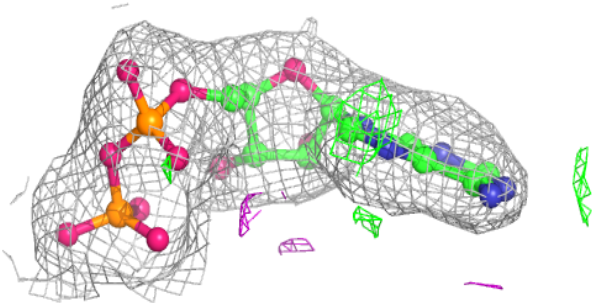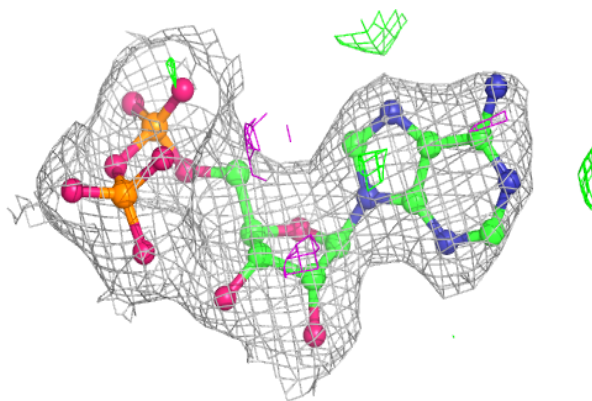
Electron density around ADP C 502:

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

Electron density around ADP B 503:

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

6.5 Other polymers ⓘ

There are no such residues in this entry.