Full wwPDB X-ray Structure Validation Report (i)

Jun 23, 2024 – 03:51 PM EDT

PDB ID	:	5AGA
Title	:	Crystal structure of the Helicase domain of human DNA polymerase theta in complex with AMPPNP
Authors	:	Newman, J.A.; Cooper, C.D.O.; Aitkenhead, H.; Pinkas, D.M.; Kupinska, K.; Burgess-Brown, N.; von Delft, F.; Arrowsmith, C.H.; Edwards, A.; Bountra, C.; Gileadi, O.
Deposited on	:	2015-01-29
Resolution	:	2.90 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (i) 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 (i)) were used in the production of this report:

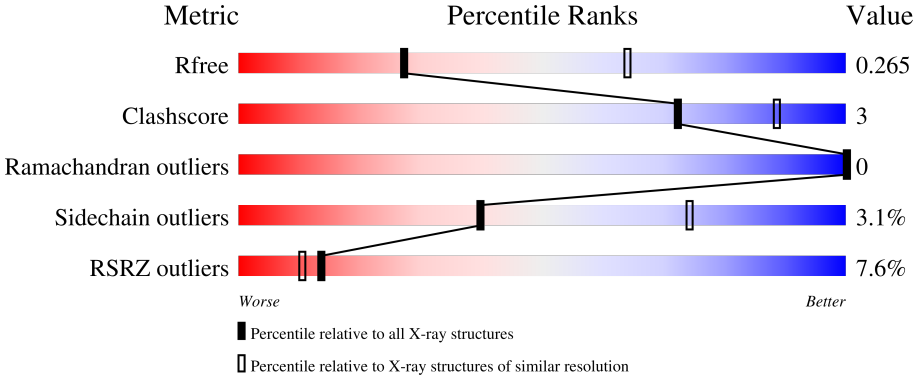
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh &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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	830	7% 84% 10% 5%

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
2	ANP	A	1892	-	-	-	X

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 6178 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.

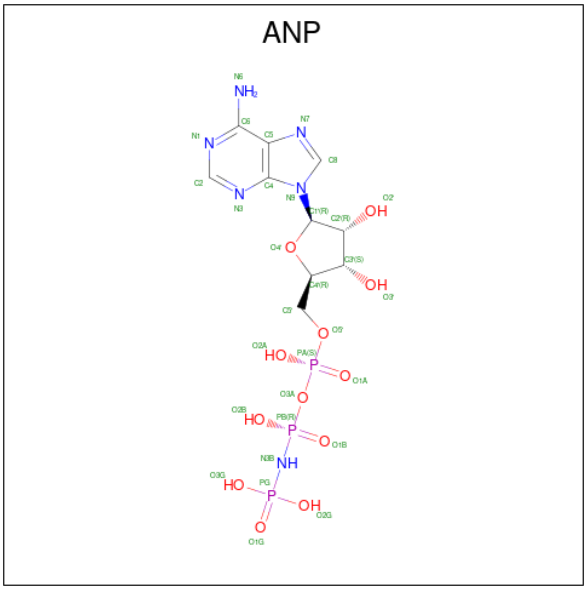
- Molecule 1 is a protein called DNA POLYMERASE THETA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	785	6093	3892	1051	1111	39	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

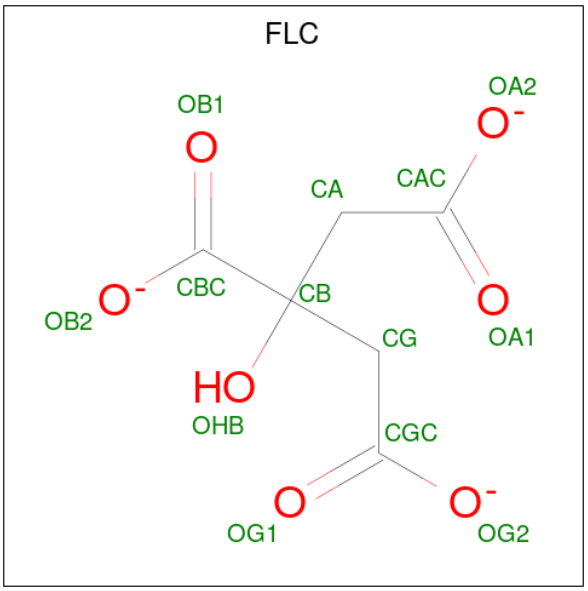
Chain	Residue	Modelled	Actual	Comment	Reference
A	65	SER	-	expression tag	UNP O75417
A	66	MET	-	expression tag	UNP O75417

- Molecule 2 is PHOSPHOAMINOPHOSPHONIC ACID-ADENYLATE ESTER (three-letter code: ANP) (formula: $C_{10}H_{17}N_6O_{12}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
2	A	1	31	10	6	12	3	0	0

- Molecule 3 is CITRATE ANION (three-letter code: FLC) (formula: $C_6H_5O_7$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 13 6 7	0	0

- Molecule 4 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total K 1 1	0	0

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Mg 1 1	0	0

- Molecule 6 is water.

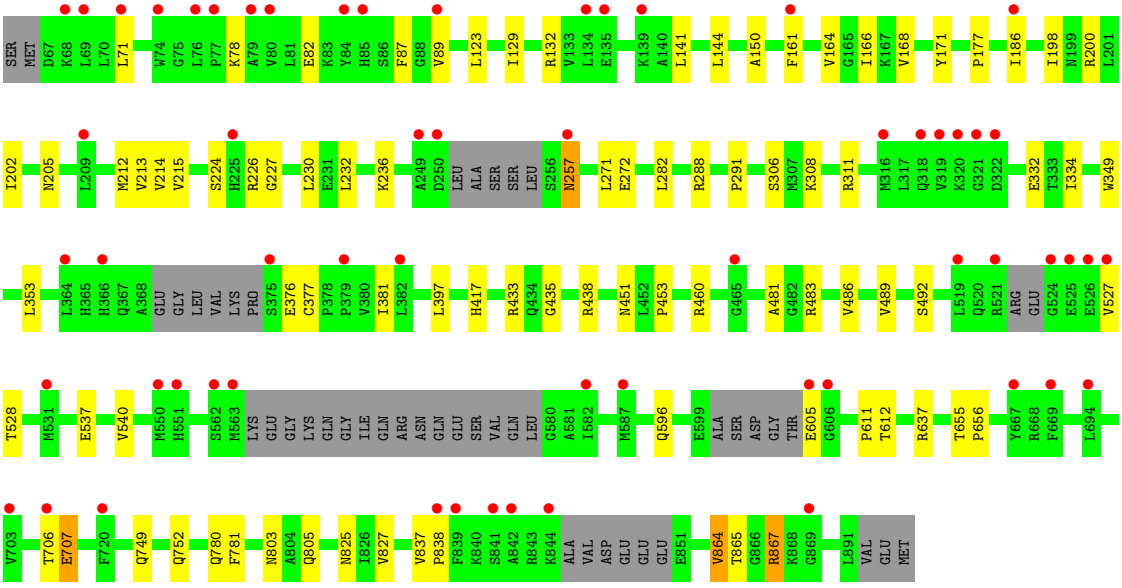
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	39	Total O 39 39	0	0

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: DNA POLYMERASE THETA

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	116.17Å 132.71Å 156.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.63 – 2.90 46.63 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.9 (46.63-2.90) 99.1 (46.63-2.90)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.81 (at 2.91Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.217 , 0.264 0.219 , 0.265	Depositor DCC
R_{free} test set	1361 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	84.3	Xtriage
Anisotropy	0.569	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 64.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6178	wwPDB-VP
Average B, all atoms (Å^2)	95.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.69% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: K, ANP, FLC, MG

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	0.21	0/6198	0.36	0/8388

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 (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	6093	0	6146	42	0
2	A	31	0	12	2	0
3	A	13	0	5	0	0
4	A	1	0	0	0	0
5	A	1	0	0	0	0
6	A	39	0	0	1	0
All	All	6178	0	6163	42	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 3.

All (42) 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:435:GLY:O	1:A:438:ARG:NH1	2.27	0.67
1:A:596:GLN:HG3	1:A:612:THR:HG22	1.76	0.67
1:A:433:ARG:NH1	1:A:451:ASN:O	2.29	0.64
1:A:168:VAL:HG22	1:A:186:ILE:HB	1.85	0.58
1:A:334:ILE:HD12	1:A:438:ARG:HB3	1.87	0.55
1:A:236:LYS:NZ	1:A:537:GLU:OE1	2.41	0.54
1:A:161:PHE:HB3	1:A:166:ILE:HB	1.91	0.52
1:A:291:PRO:HG2	1:A:489:VAL:HG12	1.93	0.51
1:A:376:GLU:HG3	1:A:377:CYS:H	1.75	0.50
1:A:129:ILE:HD11	1:A:214:VAL:HG21	1.93	0.50
1:A:144:LEU:HD13	1:A:150:ALA:HA	1.94	0.49
1:A:224:SER:HB3	1:A:227:GLY:HA3	1.95	0.49
1:A:171:TYR:CZ	1:A:177:PRO:HG2	2.48	0.49
1:A:453:PRO:HB2	1:A:483:ARG:HG2	1.95	0.48
1:A:141:LEU:HB3	1:A:213:VAL:HG12	1.96	0.48
1:A:864:VAL:HG12	1:A:865:THR:H	1.80	0.47
1:A:257:ASN:OD1	1:A:257:ASN:N	2.47	0.46
1:A:749:GLN:HA	1:A:752:GLN:HE21	1.81	0.46
1:A:198:ILE:O	1:A:202:ILE:HG12	2.16	0.46
1:A:825:ASN:OD1	1:A:827:VAL:HG12	2.15	0.46
1:A:123:LEU:HD22	2:A:1892:ANP:HN61	1.81	0.45
1:A:215:VAL:HG11	1:A:230:LEU:HD21	1.98	0.45
1:A:123:LEU:HD22	2:A:1892:ANP:N6	2.32	0.44
1:A:71:LEU:HD22	1:A:89:VAL:HG12	2.00	0.44
1:A:288:ARG:NH2	1:A:481:ALA:O	2.50	0.44
1:A:311:ARG:NH2	1:A:332:GLU:OE2	2.43	0.43
1:A:707:GLU:H	1:A:707:GLU:HG3	1.49	0.43
1:A:271:LEU:HG	1:A:282:LEU:HD11	1.99	0.43
1:A:306:SER:HB2	1:A:308:LYS:HD3	2.01	0.43
1:A:837:VAL:HA	1:A:838:PRO:HD3	1.79	0.43
1:A:353:LEU:HD12	1:A:353:LEU:HA	1.90	0.43
1:A:540:VAL:HA	1:A:611:PRO:HB3	2.00	0.42
1:A:232:LEU:O	1:A:236:LYS:HB2	2.20	0.42
1:A:803:ASN:ND2	1:A:805:GLN:OE1	2.53	0.42
1:A:132:ARG:HD2	1:A:212:MET:HB3	2.02	0.41
1:A:78:LYS:O	1:A:82:GLU:HG2	2.20	0.41
1:A:864:VAL:O	1:A:867:ARG:HB3	2.21	0.41
1:A:483:ARG:HB2	1:A:486:VAL:HB	2.02	0.41
1:A:205:ASN:O	6:A:2005:HOH:O	2.22	0.41
1:A:540:VAL:HG22	1:A:611:PRO:HB3	2.02	0.41
1:A:655:THR:HA	1:A:656:PRO:HD3	1.85	0.40
1:A:780:GLN:HG3	1:A:781:PHE:CD1	2.56	0.40

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	771/830 (93%)	730 (95%)	41 (5%)	0	100 100

There are no Ramachandran outliers to report.

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	654/716 (91%)	634 (97%)	20 (3%)	40 74

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	87	PHE
1	A	164	VAL
1	A	200	ARG
1	A	226	ARG
1	A	257	ASN
1	A	272	GLU
1	A	349	TRP
1	A	381	ILE
1	A	397	LEU
1	A	417	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	460	ARG
1	A	492	SER
1	A	527	VAL
1	A	528	THR
1	A	605	GLU
1	A	637	ARG
1	A	706	THR
1	A	707	GLU
1	A	864	VAL
1	A	867	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	318	GLN
1	A	739	GLN
1	A	752	GLN
1	A	790	GLN
1	A	816	HIS

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 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	$\#\|Z\| > 2$
2	ANP	A	1892	5	29,33,33	4.67	11 (37%)	31,52,52	2.14	4 (12%)
3	FLC	A	1893	-	12,12,12	1.15	0	17,17,17	1.54	1 (5%)

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	ANP	A	1892	5	-	8/14/38/38	0/3/3/3
3	FLC	A	1893	-	-	1/16/16/16	-

All (11) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1892	ANP	C2'-C1'	-16.22	1.29	1.53
2	A	1892	ANP	O4'-C1'	14.83	1.61	1.41
2	A	1892	ANP	O4'-C4'	-6.69	1.30	1.45
2	A	1892	ANP	PG-O1G	5.01	1.54	1.46
2	A	1892	ANP	O3'-C3'	-3.65	1.34	1.43
2	A	1892	ANP	O2'-C2'	3.45	1.51	1.43
2	A	1892	ANP	C2-N3	3.29	1.37	1.32
2	A	1892	ANP	C6-N6	2.76	1.44	1.34
2	A	1892	ANP	PB-O2B	-2.70	1.49	1.56
2	A	1892	ANP	C5-C4	-2.11	1.35	1.40
2	A	1892	ANP	PG-O3G	-2.05	1.51	1.56

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1892	ANP	C5-C6-N6	7.61	131.92	120.35
2	A	1892	ANP	N3-C2-N1	-5.61	119.92	128.68
2	A	1892	ANP	N6-C6-N1	-5.18	107.81	118.57
3	A	1893	FLC	OB2-CBC-CB	4.29	120.51	113.05
2	A	1892	ANP	C3'-C2'-C1'	2.31	104.45	100.98

There are no chirality outliers.

All (9) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1892	ANP	PA-O3A-PB-O1B
2	A	1892	ANP	PA-O3A-PB-O2B
2	A	1892	ANP	C5'-O5'-PA-O3A
2	A	1892	ANP	C3'-C4'-C5'-O5'
2	A	1892	ANP	O4'-C4'-C5'-O5'
2	A	1892	ANP	C4'-C5'-O5'-PA
2	A	1892	ANP	C5'-O5'-PA-O1A
3	A	1893	FLC	OHB-CB-CG-CGC
2	A	1892	ANP	PG-N3B-PB-O3A

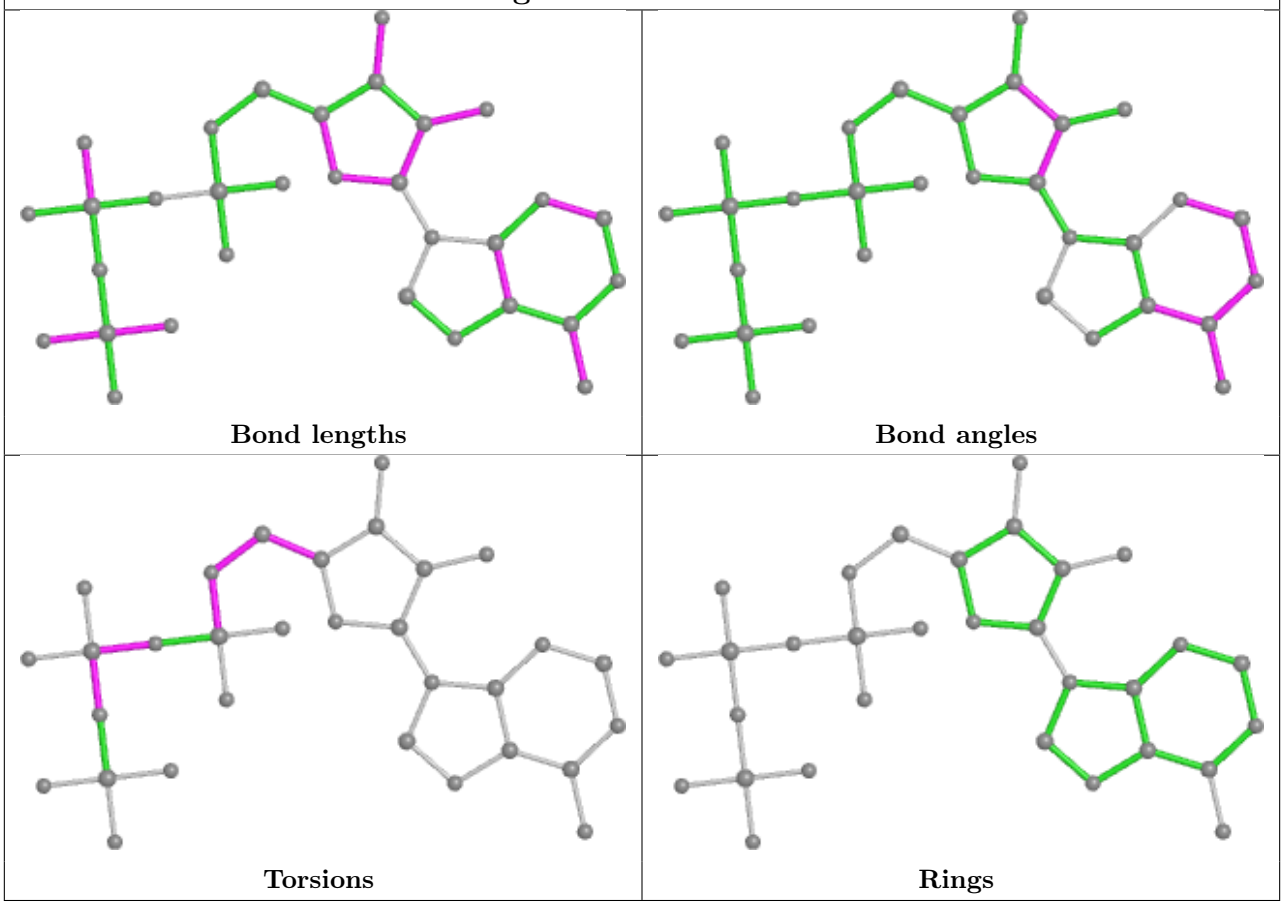
There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1892	ANP	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 ANP A 1892



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	785/830 (94%)	0.42	60 (7%) 13 10	62, 91, 134, 157	0

All (60) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	89	VAL	5.8
1	A	527	VAL	5.8
1	A	703	VAL	5.8
1	A	526	GLU	5.7
1	A	76	LEU	5.4
1	A	321	GLY	5.3
1	A	375	SER	5.3
1	A	842	ALA	4.6
1	A	524	GLY	4.6
1	A	250	ASP	4.3
1	A	706	THR	4.1
1	A	77	PRO	3.9
1	A	841	SER	3.6
1	A	525	GLU	3.6
1	A	531	MET	3.5
1	A	80	VAL	3.5
1	A	605	GLU	3.4
1	A	465	GLY	3.3
1	A	225	HIS	3.2
1	A	249	ALA	3.2
1	A	322	ASP	3.1
1	A	519	LEU	3.0
1	A	838	PRO	3.0
1	A	587	MET	2.9
1	A	161	PHE	2.8
1	A	694	LEU	2.8
1	A	71	LEU	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	84	TYR	2.7
1	A	667	TYR	2.7
1	A	521	ARG	2.6
1	A	134	LEU	2.6
1	A	319	VAL	2.5
1	A	318	GLN	2.5
1	A	839	PHE	2.4
1	A	79	ALA	2.4
1	A	606	GLY	2.4
1	A	364	LEU	2.4
1	A	68	LYS	2.4
1	A	74	TRP	2.4
1	A	135	GLU	2.4
1	A	562	SER	2.4
1	A	257	ASN	2.3
1	A	382	LEU	2.3
1	A	69	LEU	2.3
1	A	209	LEU	2.3
1	A	563	MET	2.3
1	A	550	MET	2.3
1	A	320	LYS	2.2
1	A	869	GLY	2.2
1	A	379	PRO	2.1
1	A	139	LYS	2.1
1	A	720	PHE	2.1
1	A	551	HIS	2.1
1	A	366	HIS	2.1
1	A	844	LYS	2.1
1	A	186	ILE	2.1
1	A	582	ILE	2.1
1	A	669	PHE	2.1
1	A	316	MET	2.0
1	A	85	HIS	2.0

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 monosaccharides in this entry.

6.4 Ligands (i)

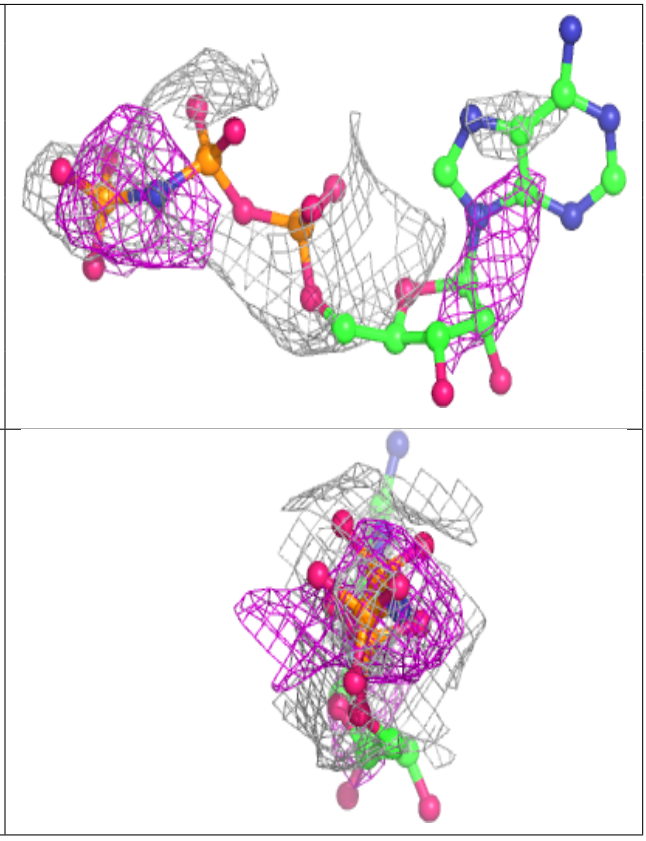
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	FLC	A	1893	13/13	0.49	0.29	100,120,133,135	0
5	MG	A	1895	1/1	0.61	0.25	116,116,116,116	0
2	ANP	A	1892	31/31	0.62	0.44	129,176,195,201	0
4	K	A	1894	1/1	0.92	0.10	78,78,78,78	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 ANP A 1892:

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.