Full wwPDB X-ray Structure Validation Report ⓘ

May 28, 2024 – 07:43 pm BST

PDB ID	:	8PO4
Title	:	Discovery and Optimisation of Potent, Efficacious and Selective Inhibitors Targeting EGFR Exon20 Insertion Mutations. Compound 33 bound to EGFR[V948R]
Authors	:	Hargreaves, D.
Deposited on	:	2023-07-03
Resolution	:	1.62 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 ⓘ 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:

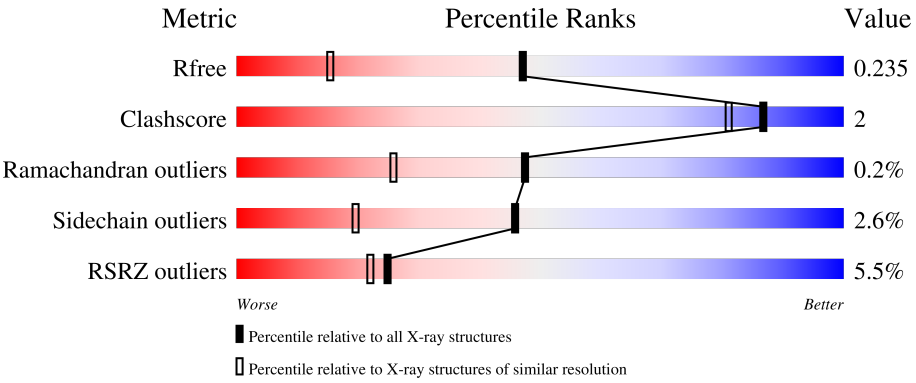
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 1.62 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	4693 (1.64-1.60)
Clashscore	141614	5002 (1.64-1.60)
Ramachandran outliers	138981	4888 (1.64-1.60)
Sidechain outliers	138945	4887 (1.64-1.60)
RSRZ outliers	127900	4609 (1.64-1.60)

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	328	4% 87% 5% · 6%
1	B	328	6% 84% 6% 10%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5251 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.

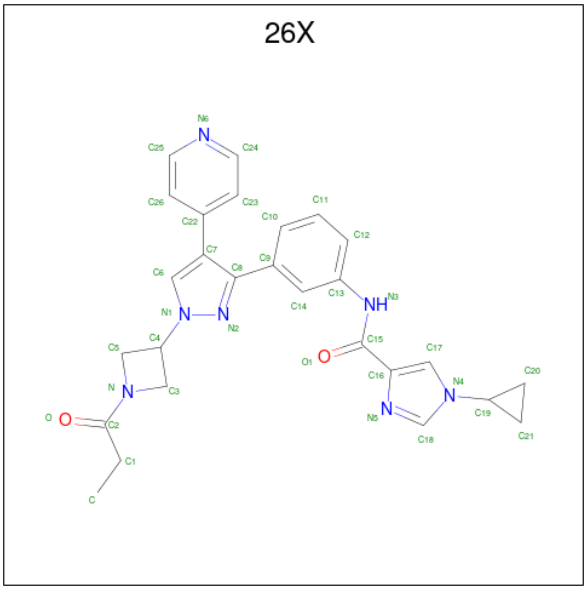
- Molecule 1 is a protein called Epidermal growth factor receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	307	Total	C	N	O	S	0	0	0
			2466	1579	418	451	18			
1	B	295	Total	C	N	O	S	0	0	0
			2376	1526	401	431	18			

There are 2 discrepancies between the modelled and reference sequences:

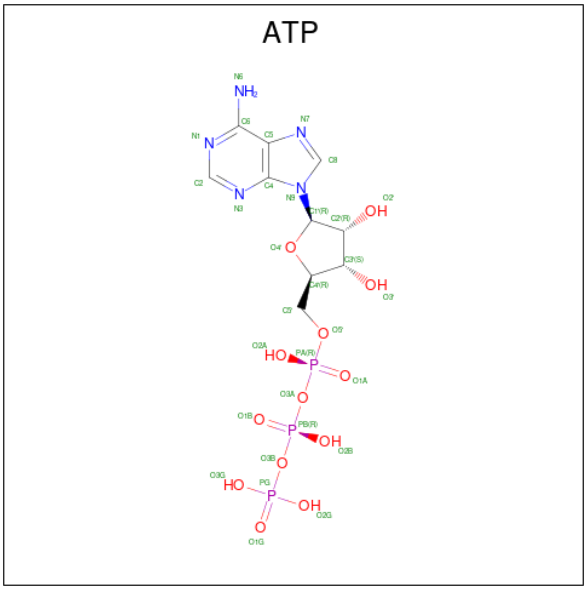
Chain	Residue	Modelled	Actual	Comment	Reference
A	948	ARG	VAL	engineered mutation	UNP P00533
B	948	ARG	VAL	engineered mutation	UNP P00533

- Molecule 2 is 1-cyclopropyl- {N}-[3-[1-(1-propanoylazetidin-3-yl)-4-pyridin-4-yl-pyrazol-3-yl]phenyl]imidazole-4-carboxamide (three-letter code: 26X) (formula: $C_{27}H_{27}N_7O_2$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			36	27	7	2		

- Molecule 3 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: $C_{10}H_{16}N_5O_{13}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	B	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		

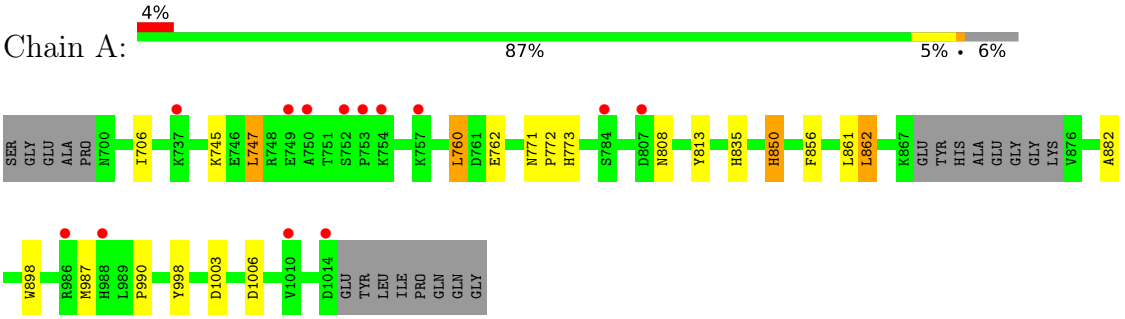
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	184	Total	O	0	0
			184	184		
5	B	157	Total	O	0	0
			157	157		

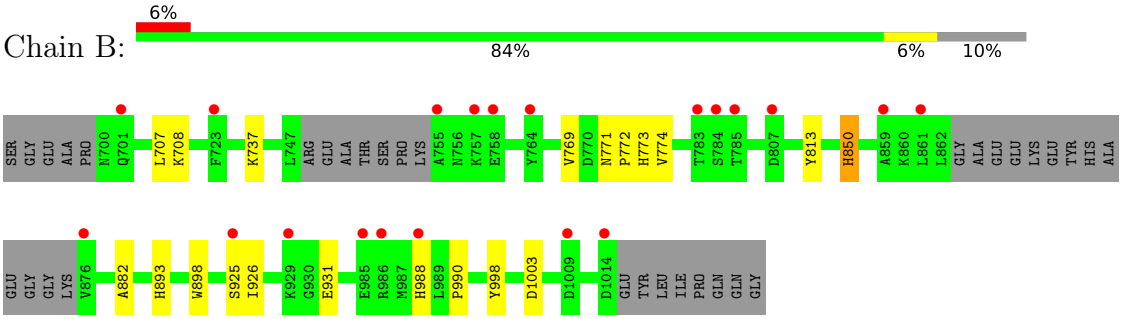
3 Residue-property plots (i)

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: Epidermal growth factor receptor



• Molecule 1: Epidermal growth factor receptor



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	156.48Å 71.20Å 76.54Å 90.00° 113.00° 90.00°	Depositor
Resolution (Å)	43.69 – 1.62 43.69 – 1.62	Depositor EDS
% Data completeness (in resolution range)	68.0 (43.69-1.62) 68.0 (43.69-1.62)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.89 (at 1.63Å)	Xtriage
Refinement program	BUSTER 2.11.8 (24-FEB-2021)	Depositor
R, R_{free}	0.209 , 0.243 0.203 , 0.235	Depositor DCC
R_{free} test set	3301 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	24.0	Xtriage
Anisotropy	0.061	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 44.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5251	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 45.96 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.2253e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: MG, ATP, 26X

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	0/2518	0.55	0/3406
1	B	0.41	0/2426	0.54	0/3282
All	All	0.41	0/4944	0.55	0/6688

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

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	2466	0	2501	12	0
1	B	2376	0	2412	9	0
2	A	36	0	0	0	0
3	B	31	0	12	0	0
4	B	1	0	0	0	0
5	A	184	0	0	0	0
5	B	157	0	0	1	0
All	All	5251	0	4925	21	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 2.

All (21) 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:850:HIS:HD2	1:A:1003:ASP:OD2	1.65	0.80
1:B:850:HIS:HD2	1:B:1003:ASP:OD2	1.80	0.65
1:B:772:PRO:O	1:B:850:HIS:HE1	1.86	0.58
1:A:772:PRO:O	1:A:850:HIS:HE1	1.84	0.58
1:A:762:GLU:HG2	1:A:861:LEU:HA	1.89	0.55
1:B:771:ASN:ND2	1:B:773:HIS:H	2.04	0.55
1:A:771:ASN:ND2	1:A:773:HIS:H	2.07	0.51
1:B:893:HIS:HD2	5:B:1247:HOH:O	1.93	0.51
1:A:850:HIS:CD2	1:A:1003:ASP:OD2	2.56	0.49
1:A:813:TYR:OH	1:A:990:PRO:HD3	2.13	0.47
1:B:882:ALA:HA	1:B:898:TRP:CD2	2.51	0.45
1:A:771:ASN:HD22	1:A:773:HIS:H	1.64	0.45
1:A:882:ALA:HA	1:A:898:TRP:CD2	2.52	0.45
1:A:706:ILE:HD11	1:A:760:LEU:HD21	1.98	0.45
1:B:926:ILE:HG23	1:B:931:GLU:HB2	2.00	0.43
1:B:769:VAL:HG11	1:B:774:VAL:HG11	2.02	0.42
1:A:747:LEU:HD23	1:A:862:LEU:HD11	2.00	0.41
1:B:769:VAL:HG11	1:B:774:VAL:CG1	2.49	0.41
1:B:813:TYR:OH	1:B:990:PRO:HD3	2.19	0.41
1:A:987:MET:HB2	1:A:987:MET:HE3	1.96	0.40
1:A:835:HIS:CD2	1:A:856:PHE:HB3	2.57	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	303/328 (92%)	295 (97%)	7 (2%)	1 (0%)	41 21
1	B	289/328 (88%)	284 (98%)	5 (2%)	0	100 100
All	All	592/656 (90%)	579 (98%)	12 (2%)	1 (0%)	47 26

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	808	ASN

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	273/288 (95%)	266 (97%)	7 (3%)	46	19
1	B	264/288 (92%)	257 (97%)	7 (3%)	44	18
All	All	537/576 (93%)	523 (97%)	14 (3%)	46	19

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	745	LYS
1	A	747	LEU
1	A	760	LEU
1	A	850	HIS
1	A	862	LEU
1	A	998	TYR
1	A	1006	ASP
1	B	707	LEU
1	B	708	LYS
1	B	737	LYS
1	B	850	HIS
1	B	925	SER
1	B	988	HIS
1	B	998	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (14) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	771	ASN
1	A	773	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	842	ASN
1	A	850	HIS
1	A	893	HIS
1	A	894	GLN
1	A	935	GLN
1	B	771	ASN
1	B	773	HIS
1	B	842	ASN
1	B	849	GLN
1	B	850	HIS
1	B	893	HIS
1	B	935	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

Of 3 ligands modelled in this entry, 1 is 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	$\#\|Z\| > 2$
3	ATP	B	1101	4	26,33,33	0.65	0	31,52,52	0.62	1 (3%)
2	26X	A	1101	1	32,41,41	0.70	1 (3%)	40,59,59	0.91	2 (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
3	ATP	B	1101	4	-	3/18/38/38	0/3/3/3
2	26X	A	1101	1	-	4/16/40/40	0/6/6/6

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1101	26X	C8-N2	-2.40	1.33	1.35

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1101	26X	C6-N1-C4	2.93	128.08	125.48
2	A	1101	26X	C17-N4-C19	-2.23	123.41	125.53
3	B	1101	ATP	C5-C6-N6	2.07	123.50	120.35

There are no chirality outliers.

All (7) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1101	26X	C-C1-C2-N
2	A	1101	26X	N2-C8-C9-C10
3	B	1101	ATP	PA-O3A-PB-O2B
2	A	1101	26X	N2-C8-C9-C14
2	A	1101	26X	C-C1-C2-O
3	B	1101	ATP	PB-O3B-PG-O3G
3	B	1101	ATP	PA-O3A-PB-O1B

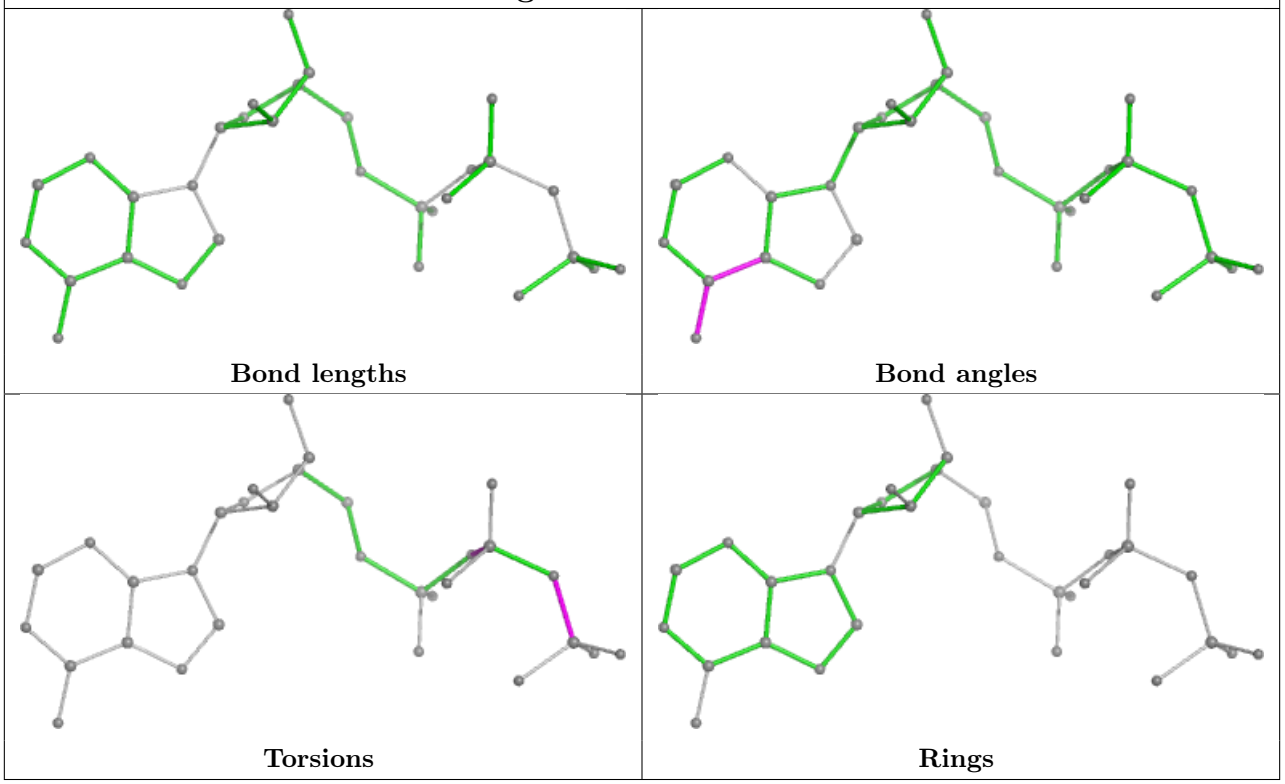
There are no ring outliers.

No monomer is involved in short contacts.

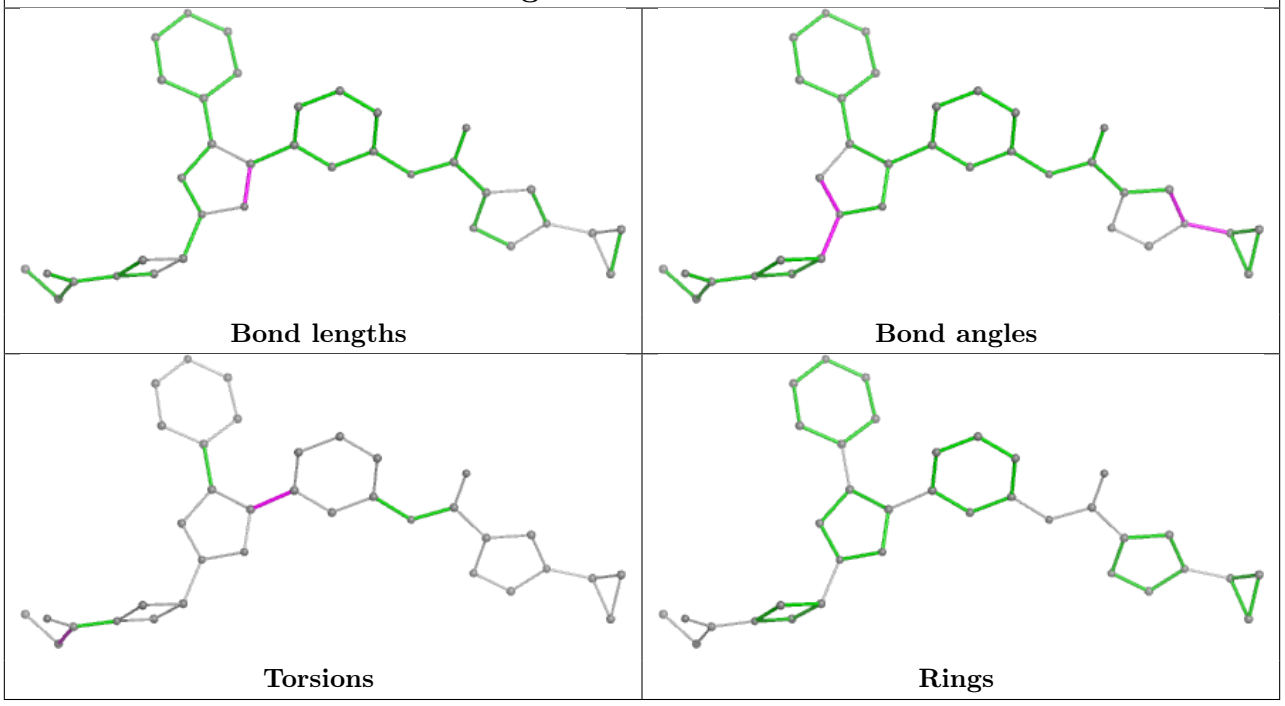
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less 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 ATP B 1101



Ligand 26X A 1101



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

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^{th} 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	307/328 (93%)	0.15	13 (4%)	36	32	14, 28, 52, 63	0
1	B	295/328 (89%)	0.28	20 (6%)	17	15	13, 30, 53, 62	0
All	All	602/656 (91%)	0.21	33 (5%)	25	22	13, 29, 52, 63	0

All (33) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	876	VAL	5.1
1	A	807	ASP	4.8
1	A	988	HIS	4.8
1	A	752	SER	3.9
1	B	861	LEU	3.8
1	B	723	PHE	3.7
1	B	988	HIS	3.7
1	B	859	ALA	3.5
1	A	986	ARG	3.5
1	B	755	ALA	3.4
1	A	749	GLU	3.3
1	A	757	LYS	3.2
1	A	1010	VAL	3.2
1	A	784	SER	3.2
1	A	750	ALA	3.1
1	B	758	GLU	3.1
1	B	701	GLN	3.1
1	A	1014	ASP	3.0
1	B	807	ASP	2.9
1	B	986	ARG	2.8
1	A	754	LYS	2.8
1	B	784	SER	2.7
1	B	783	THR	2.7
1	B	757	LYS	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	753	PRO	2.7
1	B	985	GLU	2.4
1	A	737	LYS	2.4
1	B	1014	ASP	2.3
1	B	929	LYS	2.3
1	B	1009	ASP	2.1
1	B	785	THR	2.1
1	B	764	TYR	2.0
1	B	925	SER	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)

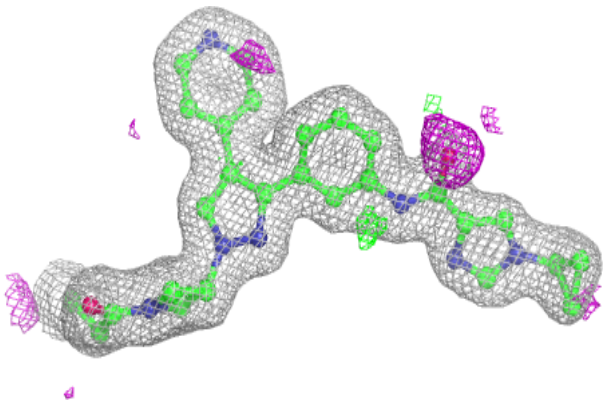
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

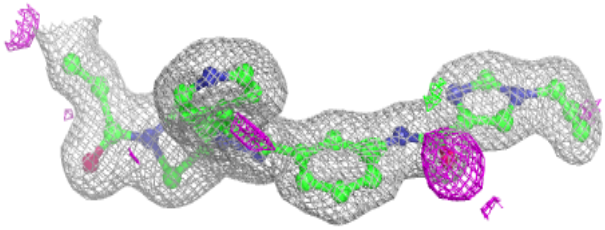
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	26X	A	1101	36/36	0.92	0.11	19,23,28,29	0
3	ATP	B	1101	31/31	0.95	0.07	25,29,30,31	0
4	MG	B	1102	1/1	0.98	0.05	26,26,26,26	0

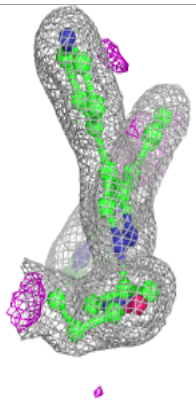
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around 26X A 1101:

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

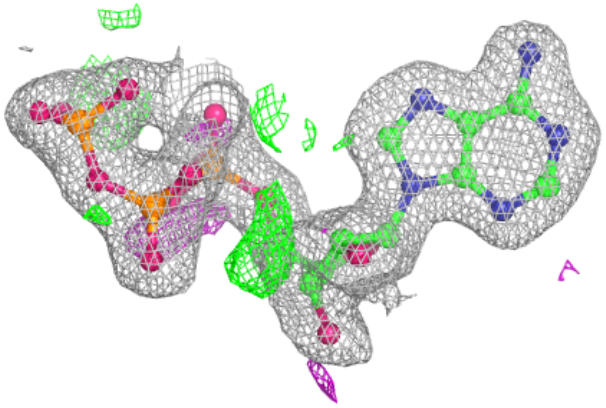


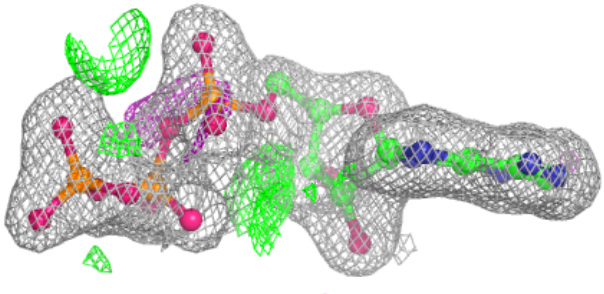


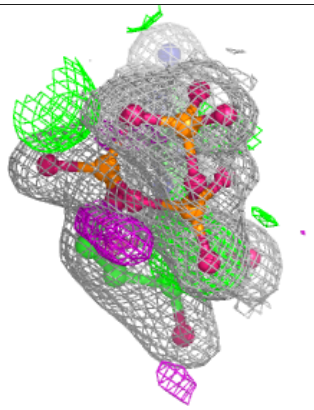


Electron density around ATP B 1101:

$2mF_o$-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.