wwPDB X-ray Structure Validation Summary Report (i)

Aug 29, 2024 – 02:05 PM EDT

PDB ID	:	9BKE
Title	:	STRUCTURE OF 4-HYDROXYPHENYLACETATE 3-MONOOXYGENASE (HPAB), OXYGENASE COMPONENT FROM ESCHERICHIA COLI MUTANT XS6 WITH AMP BOUND
Authors	:	Zhou, D.; Chen, L.; Rose, J.P.; Wang, B.C.
Deposited on	:	2024-04-27
Resolution	:	1.95 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*
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:

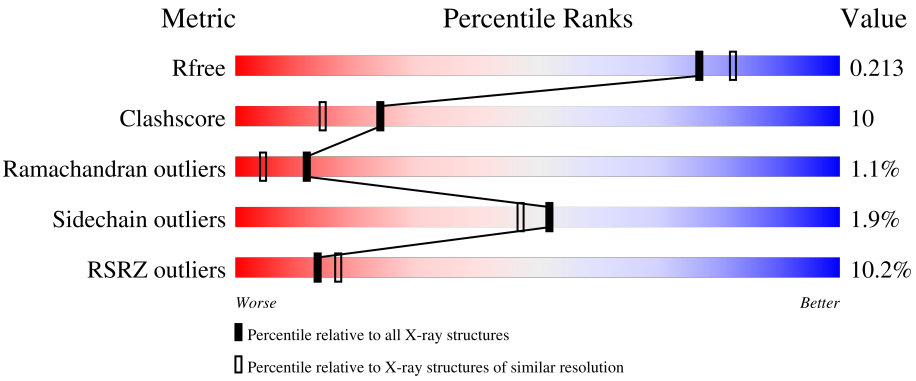
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.002 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.38.3

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 1.95 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	3187 (1.96-1.96)
Clashscore	180529	3412 (1.96-1.96)
Ramachandran outliers	177936	3390 (1.96-1.96)
Sidechain outliers	177891	3390 (1.96-1.96)
RSRZ outliers	164620	3186 (1.96-1.96)

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	527	11% 80% 16%
1	B	527	9% 81% 14%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 16943 atoms, of which 7930 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 4-hydroxyphenylacetate 3-monooxygenase oxygenase component.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	513	Total	C	H	N	O	S	0	0	0
			8054	2584	3963	713	768	26			
1	B	511	Total	C	H	N	O	S	0	0	0
			8015	2572	3943	708	766	26			

There are 30 discrepancies between the modelled and reference sequences:

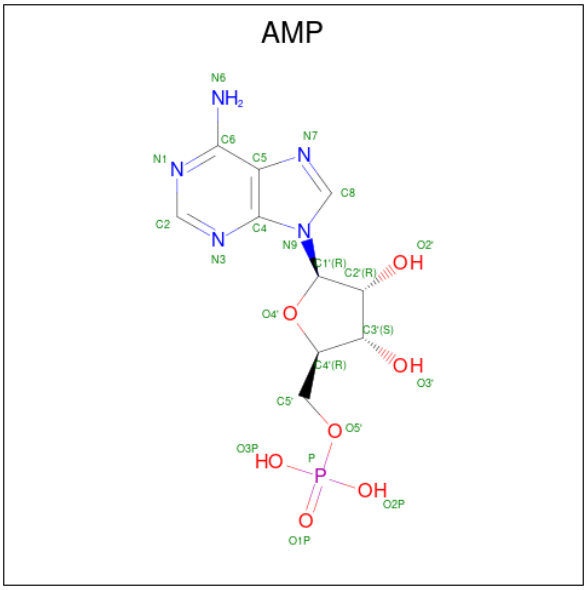
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP A0A2G8ZEZ1
A	2	HIS	-	expression tag	UNP A0A2G8ZEZ1
A	3	HIS	-	expression tag	UNP A0A2G8ZEZ1
A	4	HIS	-	expression tag	UNP A0A2G8ZEZ1
A	5	HIS	-	expression tag	UNP A0A2G8ZEZ1
A	6	HIS	-	expression tag	UNP A0A2G8ZEZ1
A	7	HIS	-	expression tag	UNP A0A2G8ZEZ1
A	8	HIS	-	expression tag	UNP A0A2G8ZEZ1
A	215	SER	PHE	conflict	UNP A0A2G8ZEZ1
A	218	ASP	ALA	conflict	UNP A0A2G8ZEZ1
A	219	LEU	GLN	conflict	UNP A0A2G8ZEZ1
A	220	GLY	VAL	conflict	UNP A0A2G8ZEZ1
A	221	SER	MET	conflict	UNP A0A2G8ZEZ1
A	223	SER	GLU	conflict	UNP A0A2G8ZEZ1
A	224	ASP	ASN	conflict	UNP A0A2G8ZEZ1
B	1	MET	-	initiating methionine	UNP A0A2G8ZEZ1
B	2	HIS	-	expression tag	UNP A0A2G8ZEZ1
B	3	HIS	-	expression tag	UNP A0A2G8ZEZ1
B	4	HIS	-	expression tag	UNP A0A2G8ZEZ1
B	5	HIS	-	expression tag	UNP A0A2G8ZEZ1
B	6	HIS	-	expression tag	UNP A0A2G8ZEZ1
B	7	HIS	-	expression tag	UNP A0A2G8ZEZ1
B	8	HIS	-	expression tag	UNP A0A2G8ZEZ1
B	215	SER	PHE	conflict	UNP A0A2G8ZEZ1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	218	ASP	ALA	conflict	UNP A0A2G8ZEZ1
B	219	LEU	GLN	conflict	UNP A0A2G8ZEZ1
B	220	GLY	VAL	conflict	UNP A0A2G8ZEZ1
B	221	SER	MET	conflict	UNP A0A2G8ZEZ1
B	223	SER	GLU	conflict	UNP A0A2G8ZEZ1
B	224	ASP	ASN	conflict	UNP A0A2G8ZEZ1

- Molecule 2 is ADENOSINE MONOPHOSPHATE (three-letter code: AMP) (formula: $C_{10}H_{14}N_5O_7P$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	P	0	0
			35	10	12	5	7	1		
2	A	1	Total	C	H	N	O	P	0	0
			35	10	12	5	7	1		

- Molecule 3 is water.

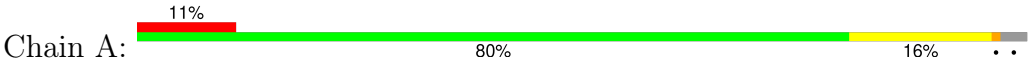
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	389	Total	O	0	0
			389	389		
3	B	415	Total	O	0	0
			415	415		

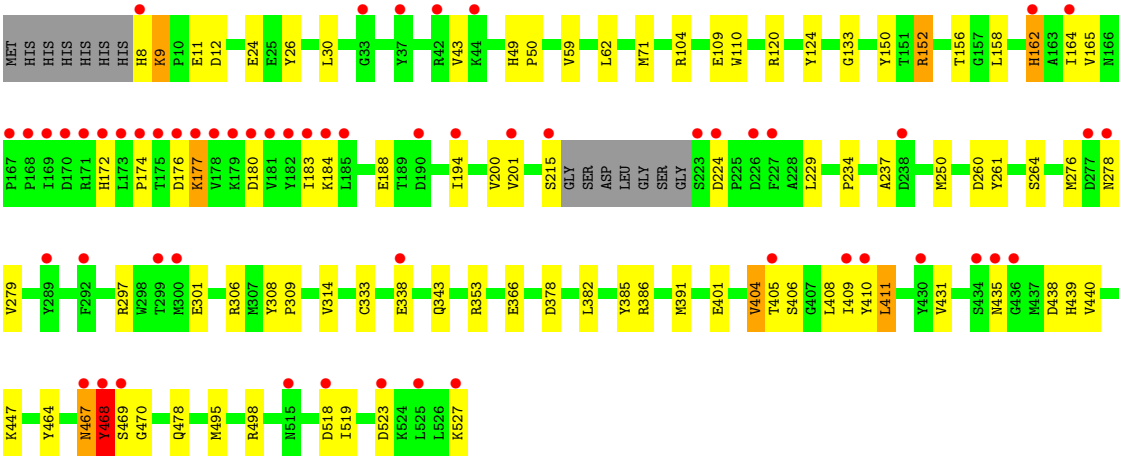
3 Residue-property plots (i)

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: 4-hydroxyphenylacetate 3-monooxygenase oxygenase component

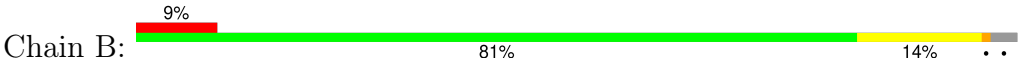
Chain A:

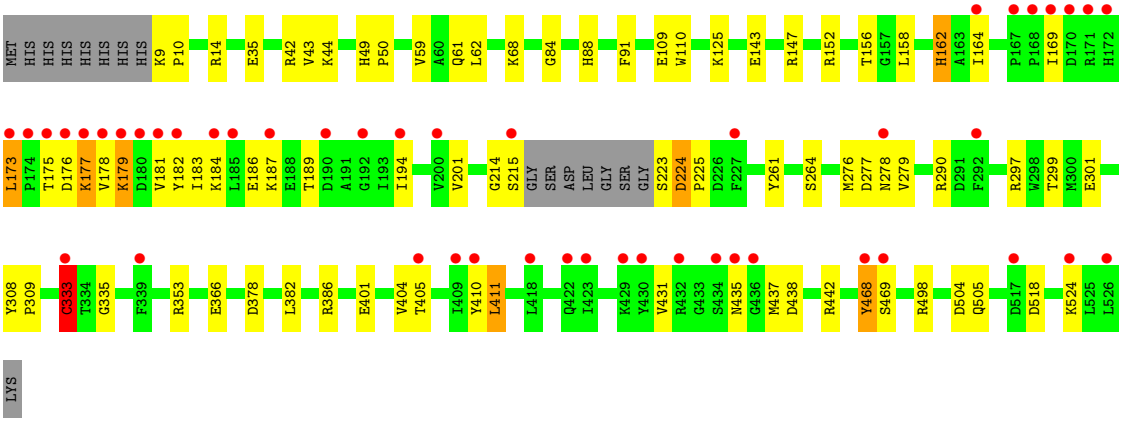




• Molecule 1: 4-hydroxyphenylacetate 3-monooxygenase oxygenase component

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 2 2	Depositor
Cell constants a, b, c, α, β, γ	100.23Å 100.23Å 336.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.74 – 1.95 39.74 – 1.95	Depositor EDS
% Data completeness (in resolution range)	76.0 (39.74-1.95) 88.4 (39.74-1.95)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.99 (at 1.95Å)	Xtriage
Refinement program	PHENIX (1.21_5207: ???)	Depositor
R, R_{free}	0.183 , 0.213 0.184 , 0.213	Depositor DCC
R_{free} test set	123240 reflections (1.80%)	wwPDB-VP
Wilson B-factor (Å^2)	19.0	Xtriage
Anisotropy	0.051	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 58.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	16943	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.35% 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: AMP

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	0.58	3/4188 (0.1%)	0.80	4/5675 (0.1%)
1	B	0.61	2/4168 (0.0%)	0.79	4/5649 (0.1%)
All	All	0.60	5/8356 (0.1%)	0.79	8/11324 (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	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	143	GLU	CB-CG	-8.05	1.36	1.52
1	A	333	CYS	CB-SG	-6.92	1.70	1.82
1	A	468	TYR	CD2-CE2	-6.82	1.29	1.39
1	B	333	CYS	CB-SG	-6.39	1.71	1.82
1	A	467	ASN	CB-CG	-5.01	1.39	1.51

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	468	TYR	CB-CG-CD2	-12.53	113.48	121.00
1	B	173	LEU	CB-CG-CD2	10.95	129.61	111.00
1	A	468	TYR	CB-CG-CD1	10.07	127.04	121.00
1	B	173	LEU	CA-CB-CG	9.04	136.10	115.30
1	B	173	LEU	CB-CG-CD1	-8.02	97.37	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	404	VAL	Peptide

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	4091	3963	3962	82	2
1	B	4072	3943	3942	79	2
2	A	46	24	24	0	0
3	A	389	0	0	31	2
3	B	415	0	0	35	2
All	All	9013	7930	7928	160	5

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 10.

The worst 5 of 160 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:B:411:LEU:N	3:B:603:HOH:O	1.79	1.12
1:A:411:LEU:N	3:A:701:HOH:O	1.82	1.09
1:B:505:GLN:OE1	3:B:601:HOH:O	1.70	1.09
1:B:109:GLU:OE2	3:B:602:HOH:O	1.75	1.02
1:A:467:ASN:O	1:A:468:TYR:HB2	1.61	1.01

All (5) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:278:ASN:O	1:B:187:LYS:NZ[1_655]	1.87	0.33
1:A:278:ASN:O	1:B:187:LYS:HZ2[1_655]	1.29	0.31
3:A:1038:HOH:O	3:B:854:HOH:O[7_555]	1.98	0.22
3:B:948:HOH:O	3:B:948:HOH:O[5_755]	2.09	0.11

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:823:HOH:O	3:A:921:HOH:O[7_555]	2.18	0.02

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	509/527 (97%)	487 (96%)	16 (3%)	6 (1%)	11	4
1	B	507/527 (96%)	487 (96%)	15 (3%)	5 (1%)	13	5
All	All	1016/1054 (96%)	974 (96%)	31 (3%)	11 (1%)	12	4

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	411	LEU
1	A	468	TYR
1	B	405	THR
1	B	411	LEU
1	A	177	LYS

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	432/443 (98%)	424 (98%)	8 (2%)	52	47
1	B	430/443 (97%)	422 (98%)	8 (2%)	52	47

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	862/886 (97%)	846 (98%)	16 (2%)	52 47

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	468	TYR
1	B	353	ARG
1	B	110	TRP
1	B	333	CYS
1	A	495	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	162	HIS

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

2 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	AMP	A	602	-	21,25,25	0.93	1 (4%)	23,38,38	1.39	3 (13%)
2	AMP	A	601	-	21,25,25	0.86	1 (4%)	23,38,38	1.45	3 (13%)

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	AMP	A	602	-	-	0/6/26/26	0/3/3/3
2	AMP	A	601	-	-	0/6/26/26	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	602	AMP	C2-N3	2.46	1.35	1.32
2	A	601	AMP	C2-N3	2.26	1.35	1.32

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	601	AMP	N3-C2-N1	-5.16	121.67	128.67
2	A	602	AMP	N3-C2-N1	-4.79	122.17	128.67
2	A	602	AMP	C4-C5-N7	-2.79	106.39	109.34
2	A	601	AMP	C4-C5-N7	-2.45	106.75	109.34
2	A	601	AMP	O3P-P-O2P	2.09	115.63	107.80

There are no chirality outliers.

There are no torsion outliers.

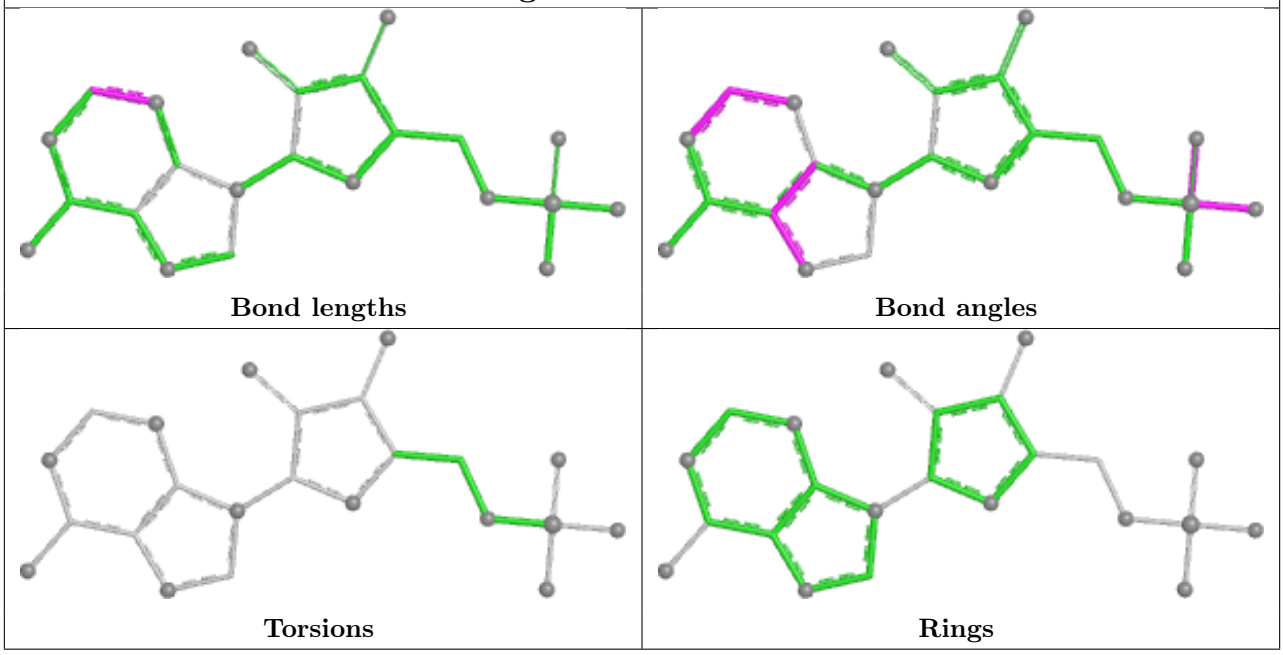
There are no ring outliers.

No monomer is involved in short contacts.

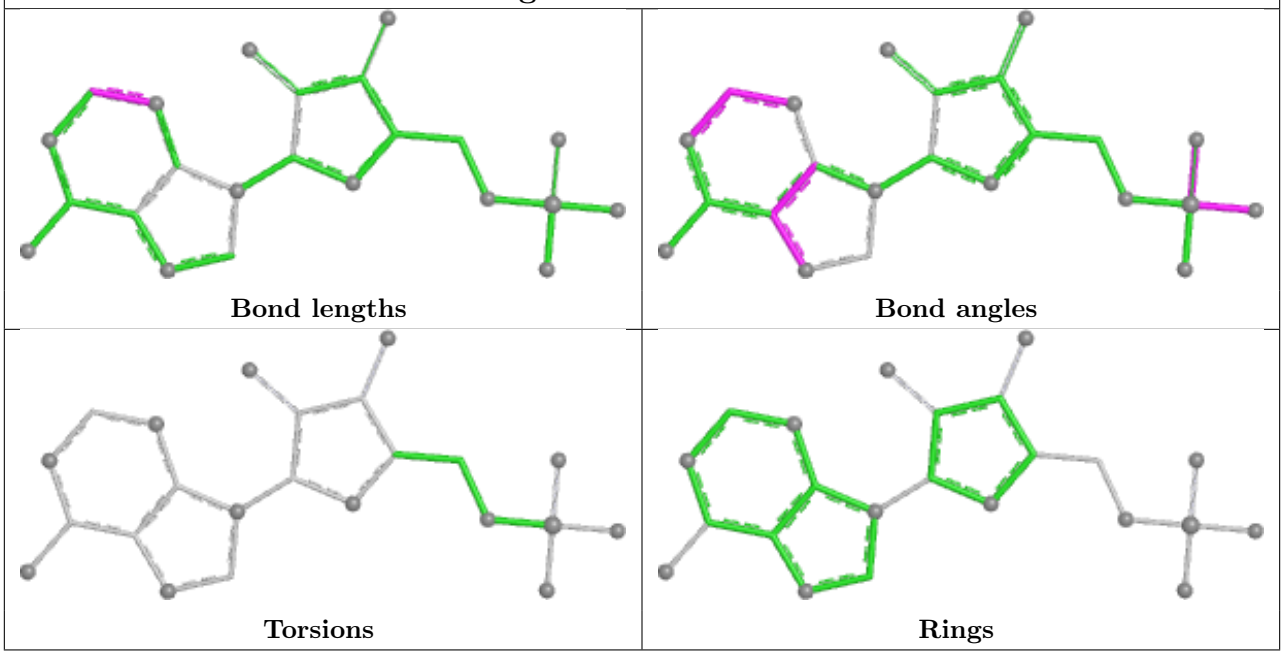
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. 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 AMP A 602



Ligand AMP A 601



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	513/527 (97%)	0.48	57 (11%) 12 15	7, 26, 66, 134	0
1	B	511/527 (96%)	0.28	47 (9%) 16 20	7, 23, 59, 145	0
All	All	1024/1054 (97%)	0.38	104 (10%) 13 17	7, 25, 62, 145	0

The worst 5 of 104 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	409	ILE	10.7
1	B	409	ILE	9.0
1	B	169	ILE	8.7
1	A	468	TYR	8.1
1	B	175	THR	7.6

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

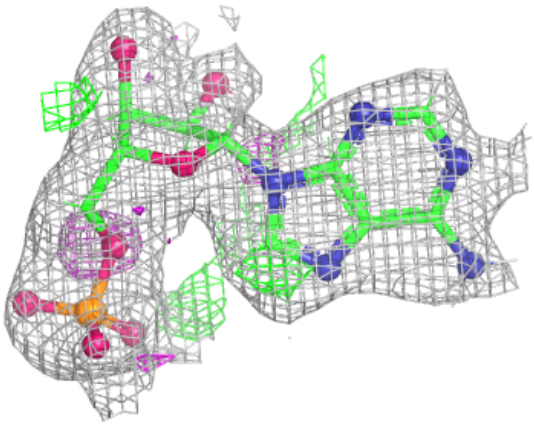
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

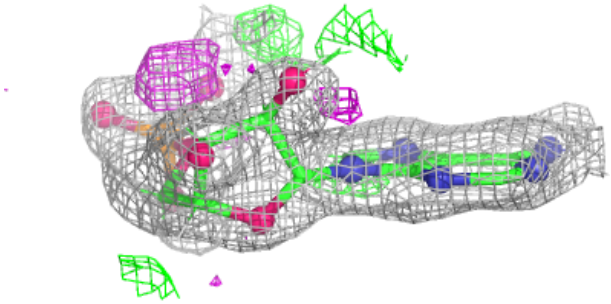
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	AMP	A	601	23/23	0.81	0.20	43,67,88,93	0
2	AMP	A	602	23/23	0.86	0.16	27,49,68,70	0

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 AMP A 601:

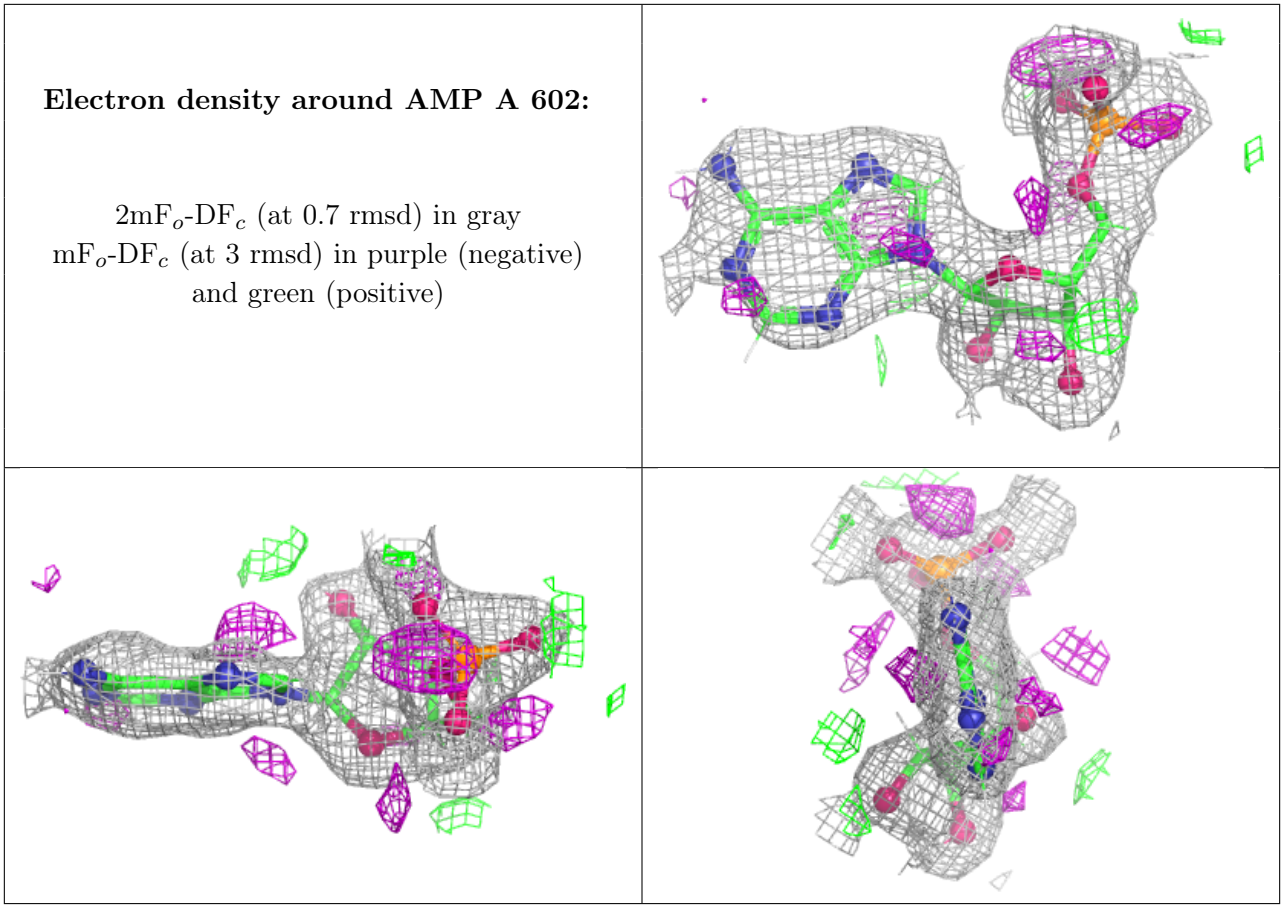
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)





Electron density around AMP A 602:

$2mF_o$-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

There are no such residues in this entry.