Full wwPDB X-ray Structure Validation Report (i)

Sep 5, 2022 – 10:09 pm BST

PDB ID	:	8AHI
Title	:	PAC-FragmentDEL: Photoactivated covalent capture of DNA encoded fragments for hit discovery
Authors	:	Baker, L.M.; Murray, J.B.; Hubbard, R.E.
Deposited on	:	2022-07-21
Resolution	:	2.69 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:

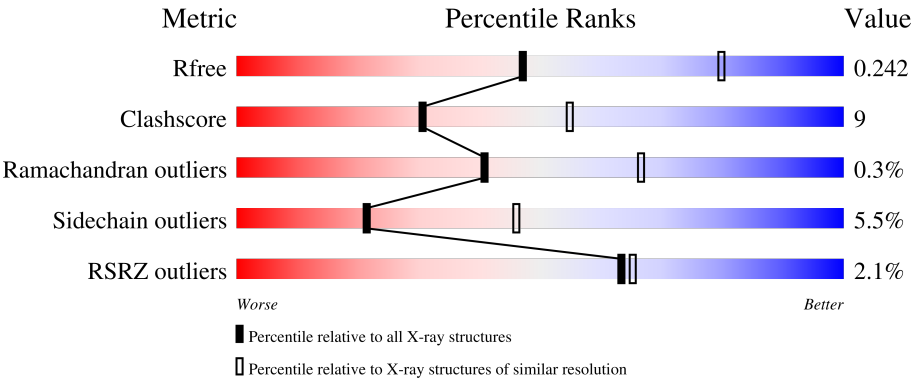
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.30
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.30

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 2.69 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	297	2% 72% 23%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4857 atoms, of which 2382 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.

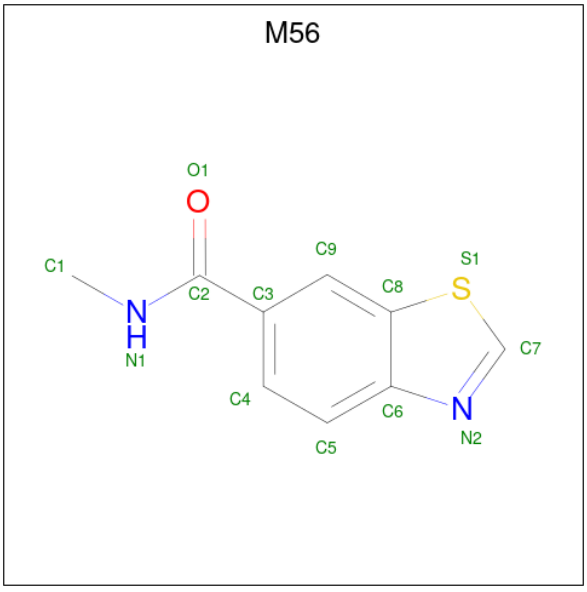
- Molecule 1 is a protein called Serine/threonine-protein kinase PAK 4.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P	S			
1	A	292	4684	1469	2374	410	416	1	14	133	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	295	GLY	-	expression tag	UNP O96013
A	296	PRO	-	expression tag	UNP O96013
A	297	LEU	-	expression tag	UNP O96013
A	298	GLY	-	expression tag	UNP O96013
A	299	SER	-	expression tag	UNP O96013
A	310	ALA	LEU	engineered mutation	UNP O96013

- Molecule 2 is {N}-methyl-1,3-benzothiazole-6-carboxamide (three-letter code: M56) (formula: $C_9H_8N_2OS$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	S	0	0
			21	9	8	2	1	1		

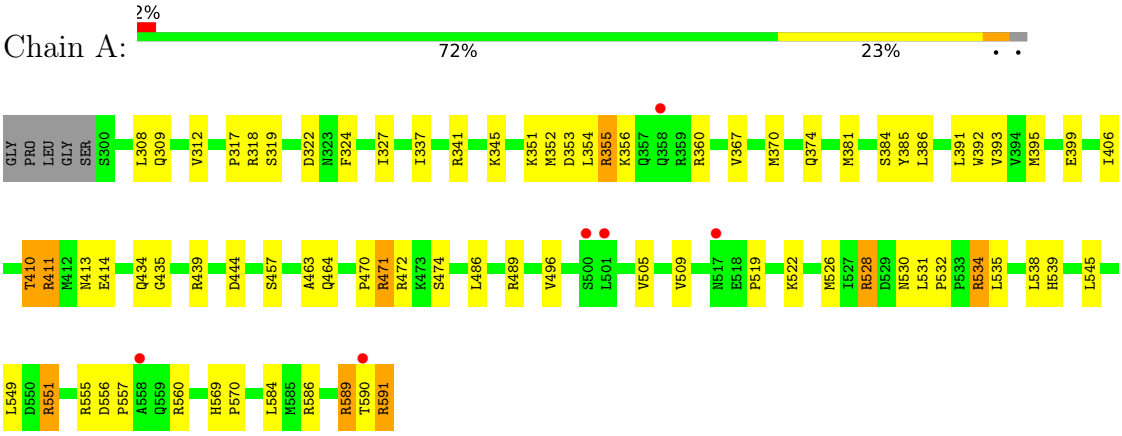
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	152	Total	O	0	0
			152	152		

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: Serine/threonine-protein kinase PAK 4



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	61.91Å 61.91Å 181.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.60 – 2.69 25.59 – 2.69	Depositor EDS
% Data completeness (in resolution range)	99.2 (25.60-2.69) 99.3 (25.59-2.69)	Depositor EDS
R_{merge}	0.21	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.23 (at 2.68Å)	Xtriage
Refinement program	REFMAC 5.8.0352	Depositor
R, R_{free}	0.155 , 0.249 0.164 , 0.242	Depositor DCC
R_{free} test set	514 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	32.1	Xtriage
Anisotropy	0.137	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4857	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.43% 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: M56, SEP

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	0.63	2/2347 (0.1%)	1.00	3/3177 (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	11

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	399	GLU	CD-OE2	5.21	1.31	1.25
1	A	414	GLU	CD-OE1	5.14	1.31	1.25

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	551	ARG	NE-CZ-NH2	-7.55	116.53	120.30
1	A	551	ARG	NE-CZ-NH1	5.86	123.23	120.30
1	A	472	ARG	NE-CZ-NH2	-5.59	117.50	120.30

There are no chirality outliers.

All (11) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	318	ARG	Sidechain
1	A	355	ARG	Sidechain

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	411	ARG	Sidechain
1	A	471	ARG	Sidechain
1	A	489	ARG	Sidechain
1	A	528	ARG	Sidechain
1	A	534	ARG	Sidechain
1	A	551	ARG	Sidechain
1	A	586	ARG	Sidechain
1	A	589	ARG	Sidechain
1	A	591	ARG	Sidechain

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	2310	2374	2365	43	0
2	A	13	8	0	0	0
3	A	152	0	0	7	0
All	All	2475	2382	2365	43	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.

All (43) 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:528:ARG:O	1:A:555:ARG:NH2	2.14	0.80
1:A:410:THR:HG21	1:A:584:LEU:HD22	1.70	0.73
1:A:406:ILE:O	1:A:410:THR:HB	1.93	0.68
1:A:545:LEU:HD23	1:A:545:LEU:C	2.16	0.65
1:A:309:GLN:NE2	1:A:386:LEU:HD23	2.12	0.64
1:A:539:HIS:CE1	3:A:748:HOH:O	2.50	0.64
1:A:560:ARG:O	3:A:701:HOH:O	2.15	0.61
1:A:370:MET:HE1	1:A:393:VAL:HG13	1.84	0.59
1:A:530:ASN:O	1:A:555:ARG:NH1	2.35	0.59
1:A:569:HIS:CG	1:A:570:PRO:HD2	2.38	0.59
1:A:360:ARG:NH1	3:A:706:HOH:O	2.33	0.58

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:444:ASP:OD1	3:A:702:HOH:O	2.17	0.57
1:A:353:ASP:HB3	1:A:356:LYS:HG3	1.87	0.57
1:A:374:GLN:HE22	1:A:381:MET:H	1.53	0.55
1:A:410:THR:CG2	1:A:584:LEU:HD22	2.36	0.54
1:A:308:LEU:HD12	1:A:354:LEU:HD21	1.90	0.54
1:A:322:ASP:OD1	1:A:341:ARG:NH1	2.42	0.53
1:A:531:LEU:HB3	1:A:532:PRO:HD2	1.92	0.52
1:A:522:LYS:HE2	1:A:526:MET:HE1	1.92	0.51
1:A:545:LEU:HD23	1:A:545:LEU:O	2.11	0.51
1:A:319:SER:O	1:A:341:ARG:NH2	2.44	0.50
1:A:539:HIS:HE1	3:A:748:HOH:O	1.92	0.50
1:A:352:MET:HE3	1:A:391:LEU:HD23	1.93	0.50
1:A:470:PRO:HD2	3:A:738:HOH:O	2.12	0.49
1:A:522:LYS:HE2	1:A:526:MET:CE	2.43	0.49
1:A:355:ARG:HG2	3:A:846:HOH:O	2.13	0.47
1:A:312:VAL:HG12	1:A:384:SER:OG	2.13	0.47
1:A:352:MET:CE	1:A:391:LEU:HD23	2.45	0.46
1:A:531:LEU:H	1:A:531:LEU:HD12	1.81	0.45
1:A:370:MET:HE2	1:A:381:MET:HB3	1.98	0.45
1:A:351:LYS:HG2	1:A:392:TRP:CD1	2.51	0.45
1:A:317:PRO:HD3	1:A:385:TYR:CG	2.53	0.44
1:A:413:ASN:OD1	1:A:413:ASN:C	2.55	0.44
1:A:526:MET:HG2	1:A:530:ASN:ND2	2.33	0.43
1:A:435:GLY:O	1:A:464:GLN:HA	2.18	0.42
1:A:327:ILE:HD13	1:A:337:ILE:CG2	2.50	0.42
1:A:486:LEU:HD23	1:A:486:LEU:O	2.19	0.42
1:A:324:PHE:HA	1:A:337:ILE:O	2.20	0.41
1:A:439:ARG:NH1	1:A:463:ALA:HB2	2.35	0.41
1:A:556:ASP:HA	1:A:557:PRO:HD2	1.93	0.41
1:A:486:LEU:HD23	1:A:486:LEU:C	2.40	0.41
1:A:505:VAL:HG11	1:A:549:LEU:HD13	2.04	0.40
1:A:509:VAL:HG11	1:A:535:LEU:HD22	2.03	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	289/297 (97%)	281 (97%)	7 (2%)	1 (0%)	41 66

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	589	ARG

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	253/256 (99%)	239 (94%)	14 (6%)	21 46

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

Mol	Chain	Res	Type
1	A	345	LYS
1	A	367	VAL
1	A	395	MET
1	A	410	THR
1	A	411	ARG
1	A	434	GLN
1	A	457	SER
1	A	471	ARG
1	A	496	VAL
1	A	519	PRO
1	A	534	ARG
1	A	538	LEU
1	A	590	THR
1	A	591	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	309	GLN
1	A	374	GLN
1	A	434	GLN
1	A	530	ASN
1	A	539	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)

1 non-standard protein/DNA/RNA residue is 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	$\#\|Z\| > 2$
1	SEP	A	474	1	8,9,10	0.55	0	8,12,14	1.53	1 (12%)

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
1	SEP	A	474	1	-	0/5/8/10	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	474	SEP	O3P-P-O2P	2.95	118.92	107.64

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

1 ligand is 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	M56	A	601	-	12,14,14	2.86	1 (8%)	13,19,19	1.03	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
2	M56	A	601	-	-	1/6/6/6	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	M56	C8-S1	-9.73	1.66	1.74

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

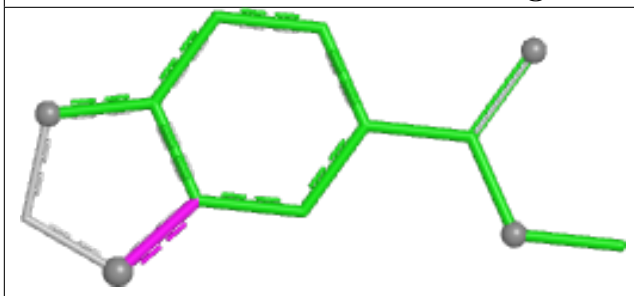
Mol	Chain	Res	Type	Atoms
2	A	601	M56	C3-C2-N1-C1

There are no ring outliers.

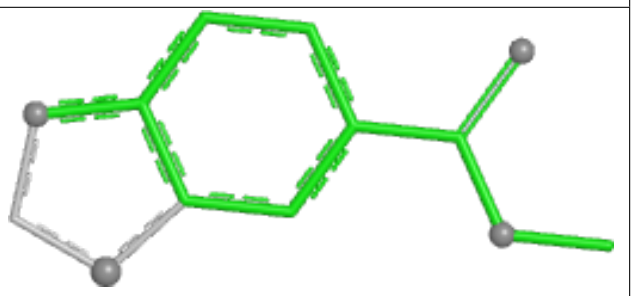
No monomer is involved in short contacts.

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.

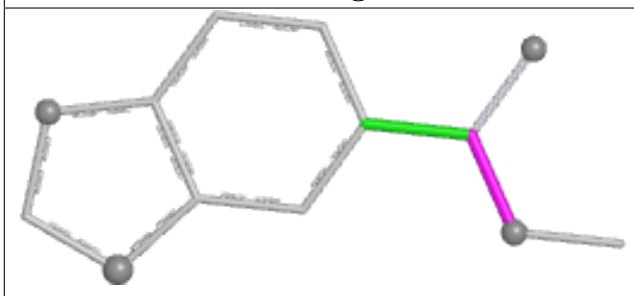
Ligand M56 A 601



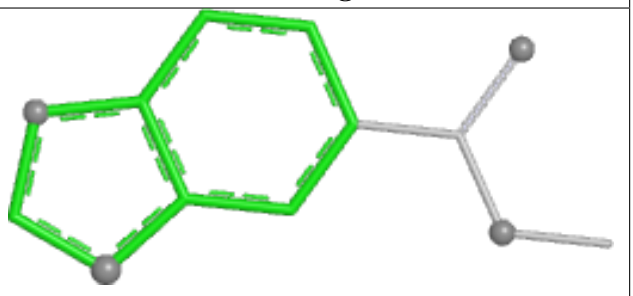
Bond lengths



Bond angles



Torsions



Rings

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	291/297 (97%)	-0.24	6 (2%)	63	65	15, 30, 51, 63	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	358	GLN	3.3
1	A	517	ASN	2.5
1	A	500	SER	2.5
1	A	558	ALA	2.4
1	A	590	THR	2.1
1	A	501	LEU	2.1

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

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
1	SEP	A	474	10/11	0.98	0.11	16,24,27,28	0

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,

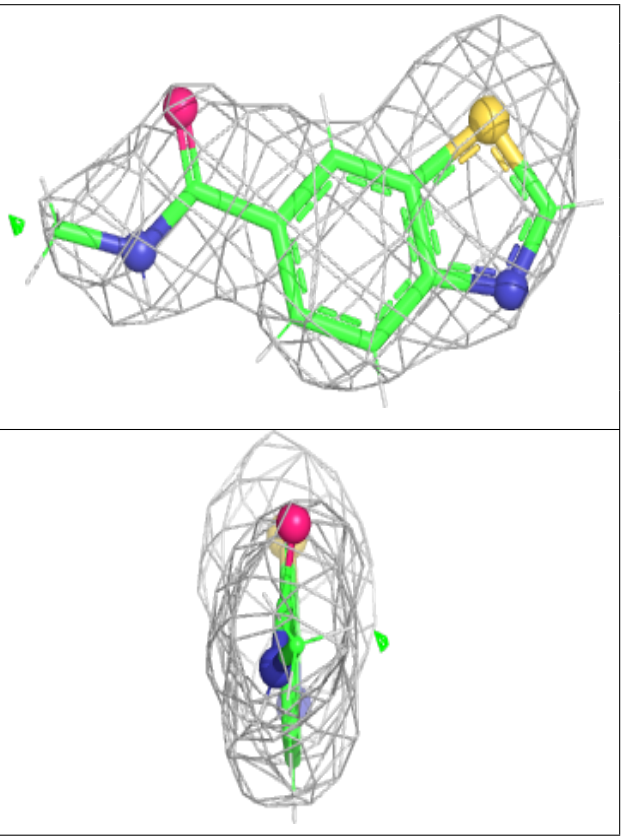
median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	M56	A	601	13/13	0.97	0.13	24,32,42,43	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 M56 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)



6.5 Other polymers (i)

There are no such residues in this entry.