Full wwPDB X-ray Structure Validation Report (i)

Dec 2, 2021 – 01:11 pm GMT

PDB ID	:	7P3G
Title	:	EED in complex with compound 4
Authors	:	Read, J.A.
Deposited on	:	2021-07-07
Resolution	:	2.39 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 following versions of software and data (see references (i)) were used in the production of this report:

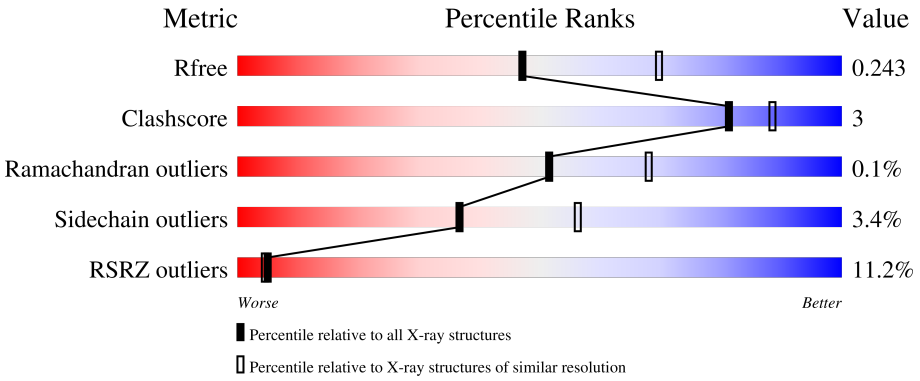
MolProbity	:	4.02b-467
Mogul	:	1.8.4 (270009), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.39 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	366	11% 86% 10%
1	B	366	10% 86% 7% 6%

2 Entry composition

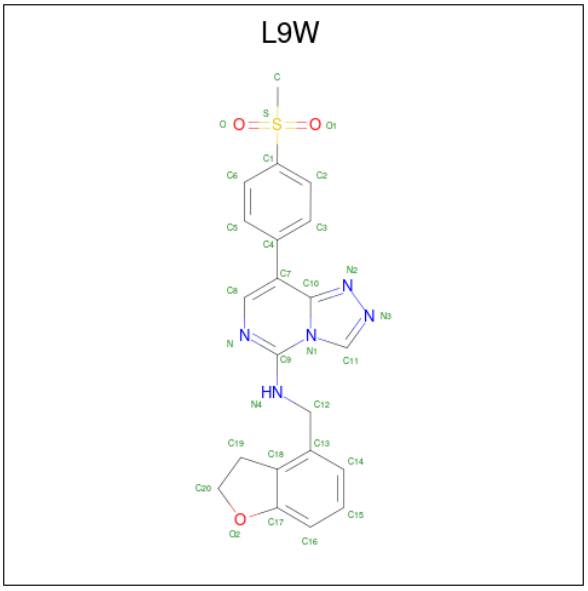
There are 4 unique types of molecules in this entry. The entry contains 5900 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Polycomb protein EED.

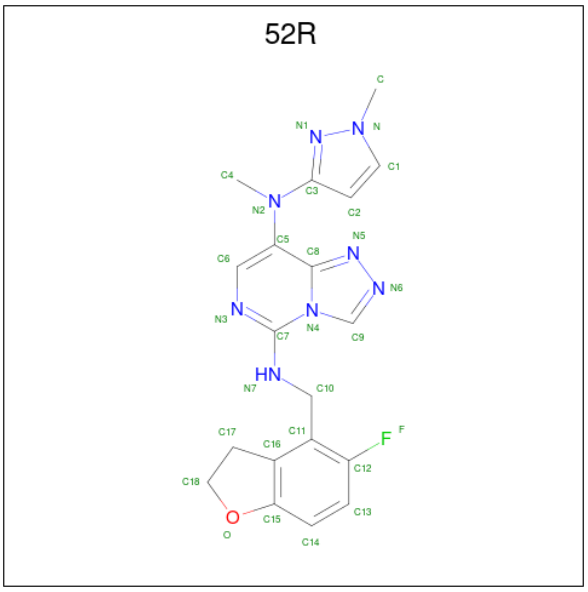
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	350	Total	C	N	O	S	0	0	0
			2790	1769	489	511	21			
1	B	344	Total	C	N	O	S	0	0	0
			2713	1721	467	504	21			

- Molecule 2 is N-(2,3-dihydro-1-benzofuran-4-ylmethyl)-8-(4-methylsulfonylphenyl)-[1,2,4]triazolo[4,3-c]pyrimidin-5-amine (three-letter code: L9W) (formula: $C_{21}H_{19}N_5O_3S$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			30	21	5	3	1		

- Molecule 3 is N5-[(5-fluoranyl-2,3-dihydro-1-benzofuran-4-yl)methyl]-N8-methyl-N8-(1-methylpyrazol-3-yl)-[1,2,4]triazolo[4,3-c]pyrimidine-5,8-diamine (three-letter code: 52R) (formula: $C_{19}H_{19}FN_8O$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total C F N O 29 19 1 8 1	0	0

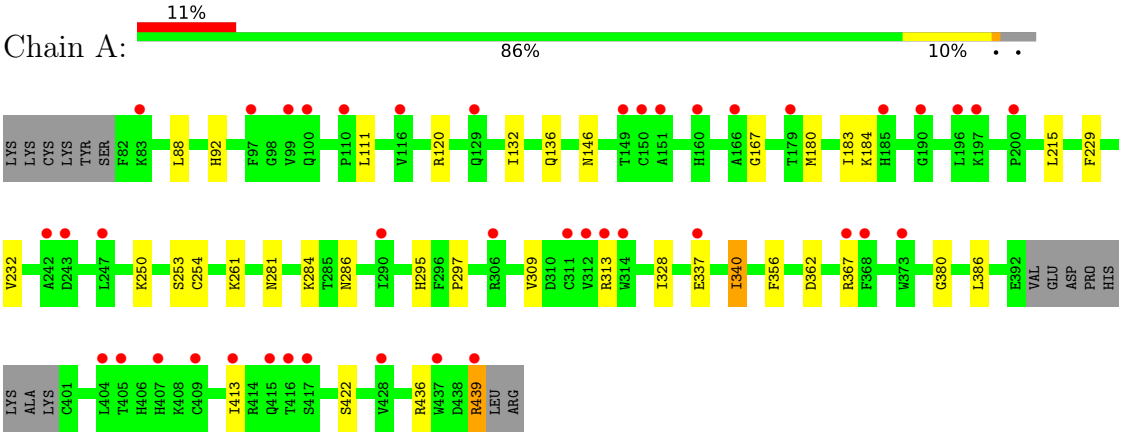
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	199	Total O 199 199	0	0
4	B	139	Total O 139 139	0	0

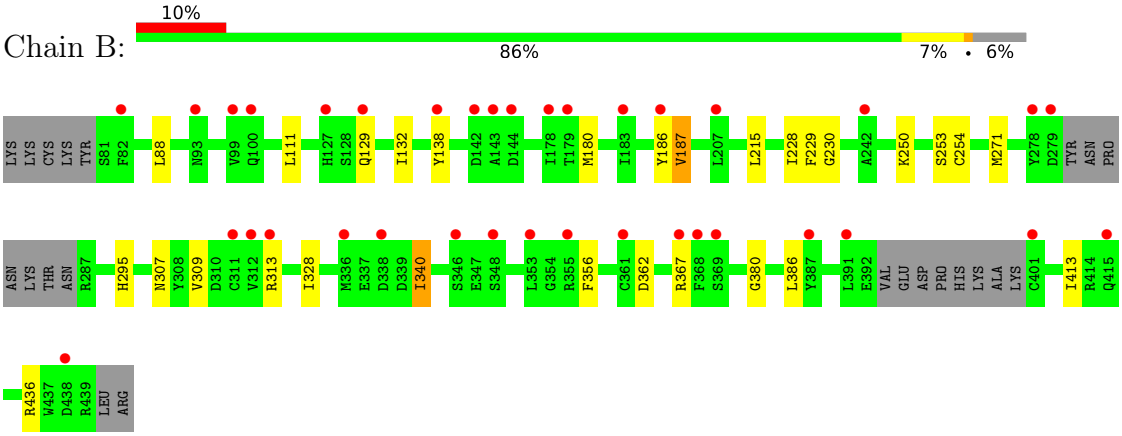
3 Residue-property plots (i)

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: Polycomb protein EED



• Molecule 1: Polycomb protein EED



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	87.27Å 89.31Å 98.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.90 – 2.39 30.87 – 2.39	Depositor EDS
% Data completeness (in resolution range)	99.8 (30.90-2.39) 99.8 (30.87-2.39)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.63 (at 2.39Å)	Xtriage
Refinement program	BUSTER	Depositor
R, R_{free}	0.182 , 0.239 0.193 , 0.243	Depositor DCC
R_{free} test set	1616 reflections (5.19%)	wwPDB-VP
Wilson B-factor (Å^2)	50.1	Xtriage
Anisotropy	0.625	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.000 for k,h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5900	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 5.49% 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: L9W, 52R

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	0.52	0/2861	0.70	0/3884
1	B	0.51	0/2780	0.69	0/3777
All	All	0.51	0/5641	0.70	0/7661

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	2790	0	2660	16	0
1	B	2713	0	2555	12	0
2	A	30	0	0	0	0
3	B	29	0	0	0	0
4	A	199	0	0	0	0
4	B	139	0	0	3	0
All	All	5900	0	5215	28	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 (28) 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:136:GLN:HB3	1:A:180:MET:HG3	1.67	0.76
1:A:250:LYS:HE3	1:A:340:ILE:HG12	1.77	0.67
1:B:250:LYS:HE3	1:B:340:ILE:HG12	1.78	0.64
1:A:215:LEU:HD21	1:A:253:SER:HB3	1.85	0.57
1:A:92:HIS:HD2	1:A:120:ARG:HH11	1.54	0.56
1:A:232:VAL:HG22	1:A:295:HIS:HB3	1.88	0.55
1:A:215:LEU:HB2	1:A:229:PHE:HB2	1.90	0.53
1:B:215:LEU:HB2	1:B:229:PHE:HB2	1.90	0.53
1:B:132:ILE:HD11	1:B:436:ARG:HB2	1.94	0.50
1:A:132:ILE:HD11	1:A:436:ARG:HB2	1.95	0.48
1:A:328:ILE:HB	1:A:356:PHE:HB2	1.96	0.48
1:A:146:ASN:O	1:A:167:GLY:HA3	2.13	0.48
1:B:230:GLY:O	1:B:295:HIS:HA	2.14	0.47
1:A:183:ILE:HG22	1:A:184:LYS:HG2	1.97	0.46
1:A:380:GLY:HA3	1:A:413:ILE:HB	1.98	0.45
1:B:138:TYR:HB2	1:B:180:MET:HG3	1.97	0.45
1:B:328:ILE:HB	1:B:356:PHE:HB2	1.98	0.45
1:A:261:LYS:HD3	1:A:297:PRO:HG3	1.99	0.45
1:B:380:GLY:HA3	1:B:413:ILE:HB	1.99	0.45
1:A:136:GLN:HB3	1:A:180:MET:CG	2.43	0.43
1:B:187:VAL:HG12	4:B:655:HOH:O	2.18	0.43
1:B:271:MET:HG3	4:B:689:HOH:O	2.19	0.43
1:B:254:CYS:HB2	1:B:309:VAL:HB	2.01	0.43
1:A:281:ASN:HB3	1:A:284:LYS:HG2	2.01	0.41
1:B:215:LEU:HD21	1:B:253:SER:HB3	2.01	0.41
1:B:186:TYR:HB2	4:B:695:HOH:O	2.21	0.41
1:A:254:CYS:HB2	1:A:309:VAL:HB	2.03	0.40
1:A:422:SER:HB3	1:A:439:ARG:HH21	1.86	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	346/366 (94%)	334 (96%)	12 (4%)	0	100 100
1	B	338/366 (92%)	325 (96%)	12 (4%)	1 (0%)	41 55
All	All	684/732 (93%)	659 (96%)	24 (4%)	1 (0%)	51 68

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	187	VAL

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	301/329 (92%)	291 (97%)	10 (3%)	38 57
1	B	290/329 (88%)	280 (97%)	10 (3%)	37 56
All	All	591/658 (90%)	571 (97%)	20 (3%)	37 56

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

Mol	Chain	Res	Type
1	A	88	LEU
1	A	111	LEU
1	A	286	ASN
1	A	313	ARG
1	A	337	GLU
1	A	340	ILE
1	A	362	ASP
1	A	367	ARG
1	A	386	LEU
1	A	439	ARG
1	B	88	LEU
1	B	111	LEU
1	B	129	GLN
1	B	228	ILE
1	B	307	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	313	ARG
1	B	340	ILE
1	B	362	ASP
1	B	367	ARG
1	B	386	LEU

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	100	GLN

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)

2 ligands are modelled in this entry.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

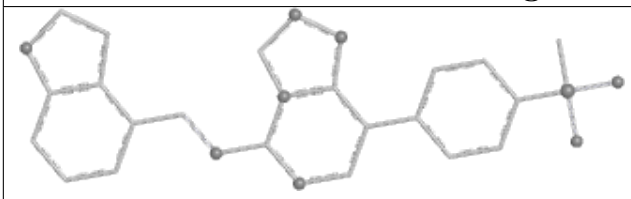
There are no ring outliers.

No monomer is involved in short contacts.

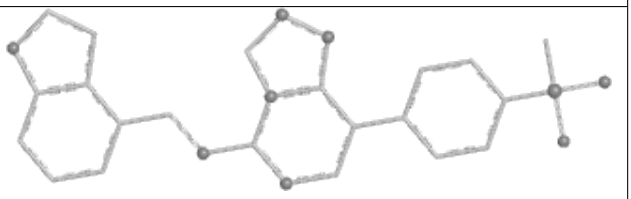
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included.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

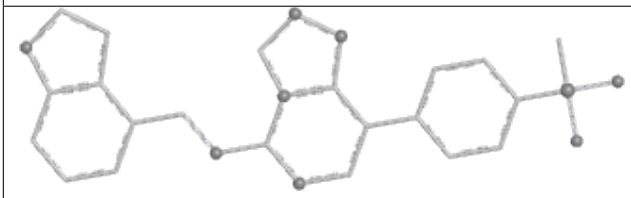
Ligand L9W A 501



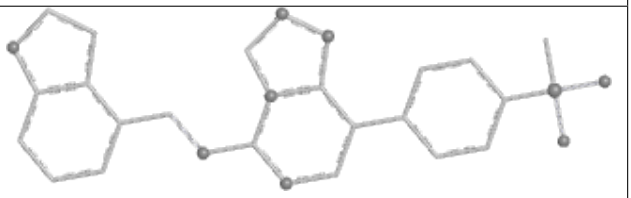
Bond lengths



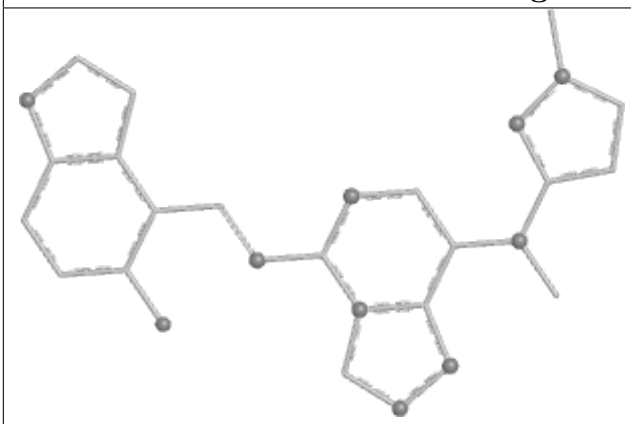
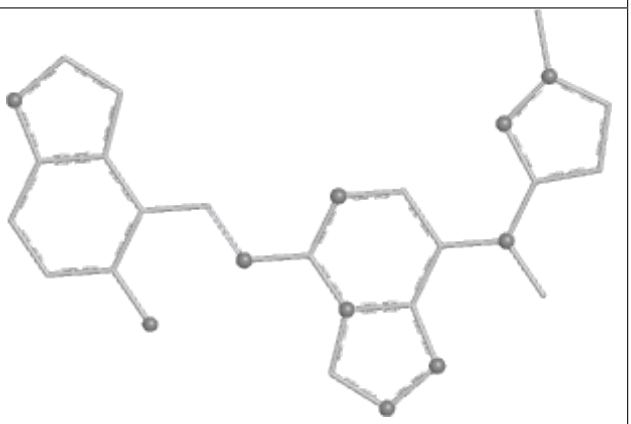
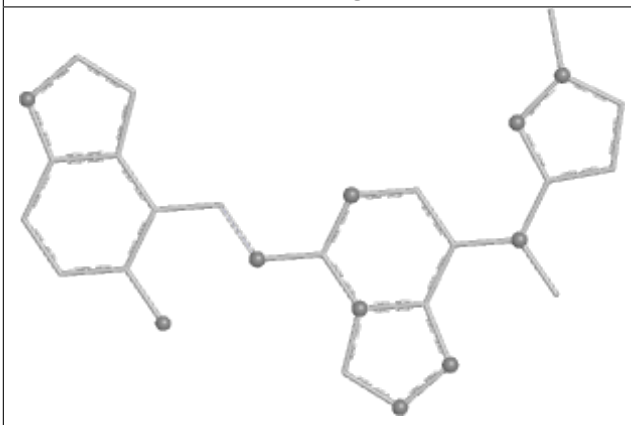
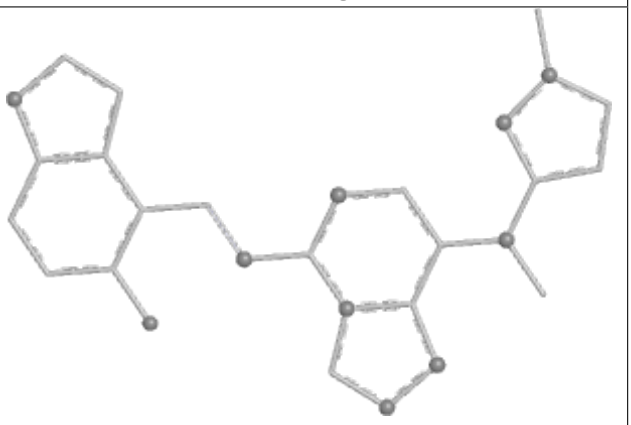
Bond angles



Torsions



Rings

Ligand 52R B 501	
	
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, 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	350/366 (95%)	0.70	42 (12%)	4	3	41, 59, 84, 101	0
1	B	344/366 (93%)	0.76	36 (10%)	6	5	44, 68, 95, 116	0
All	All	694/732 (94%)	0.73	78 (11%)	5	4	41, 63, 92, 116	0

All (78) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	338	ASP	6.8
1	B	82	PHE	5.8
1	A	415	GLN	4.0
1	A	190	GLY	4.0
1	B	143	ALA	3.9
1	B	312	VAL	3.7
1	A	312	VAL	3.7
1	A	110	PRO	3.7
1	A	179	THR	3.5
1	A	405	THR	3.5
1	B	361	CYS	3.5
1	A	99	VAL	3.5
1	B	99	VAL	3.5
1	A	150	CYS	3.5
1	A	100	GLN	3.5
1	B	142	ASP	3.4
1	A	416	THR	3.3
1	A	242	ALA	3.3
1	A	407	HIS	3.3
1	B	415	GLN	3.2
1	A	373	TRP	3.1
1	B	242	ALA	3.1
1	B	100	GLN	3.1
1	B	278	TYR	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	336	MET	3.1
1	B	127	HIS	3.0
1	B	93	ASN	3.0
1	B	183	ILE	3.0
1	A	149	THR	3.0
1	A	116	VAL	2.9
1	A	306	ARG	2.9
1	A	313	ARG	2.9
1	B	279	ASP	2.8
1	B	367	ARG	2.8
1	A	409	CYS	2.8
1	A	413	ILE	2.8
1	A	196	LEU	2.7
1	B	186	TYR	2.7
1	B	178	ILE	2.6
1	A	151	ALA	2.6
1	A	185	HIS	2.6
1	A	83	LYS	2.6
1	B	129	GLN	2.6
1	A	97	PHE	2.5
1	A	368	PHE	2.5
1	B	346	SER	2.5
1	B	401	CYS	2.5
1	A	337	GLU	2.4
1	B	207	LEU	2.4
1	A	404	LEU	2.4
1	A	290	ILE	2.4
1	A	367	ARG	2.4
1	A	311	CYS	2.4
1	A	243	ASP	2.3
1	B	179	THR	2.3
1	B	391	LEU	2.3
1	B	138	TYR	2.3
1	B	368	PHE	2.3
1	B	313	ARG	2.2
1	A	417	SER	2.2
1	B	144	ASP	2.2
1	A	247	LEU	2.2
1	A	200	PRO	2.2
1	B	348	SER	2.1
1	B	355	ARG	2.1
1	B	387	TYR	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	369	SER	2.1
1	A	197	LYS	2.1
1	B	353	LEU	2.1
1	B	438	ASP	2.1
1	A	129	GLN	2.0
1	A	428	VAL	2.0
1	A	160	HIS	2.0
1	B	311	CYS	2.0
1	A	439	ARG	2.0
1	A	314	TRP	2.0
1	A	437	TRP	2.0
1	A	166	ALA	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)

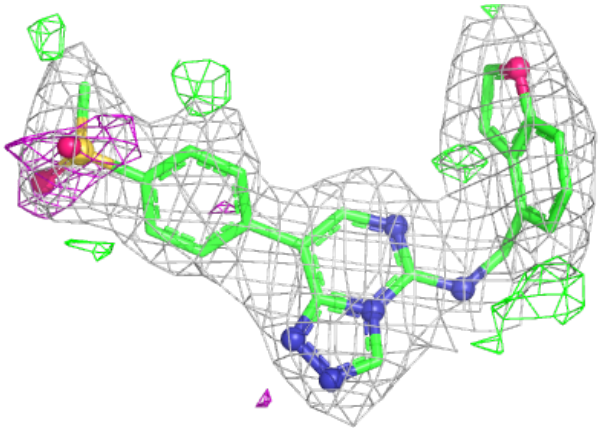
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

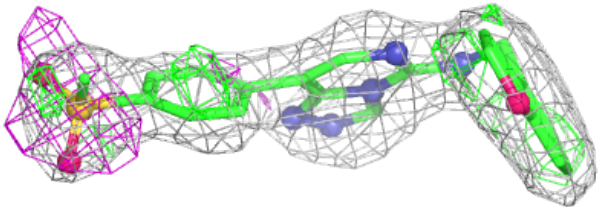
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	L9W	A	501	30/30	0.82	0.22	29,42,61,65	0
3	52R	B	501	29/29	0.87	0.20	41,54,68,70	0

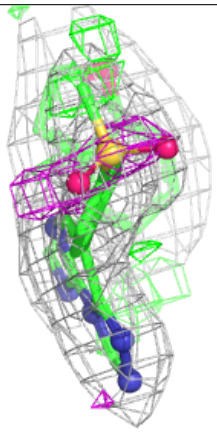
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around L9W A 501:

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

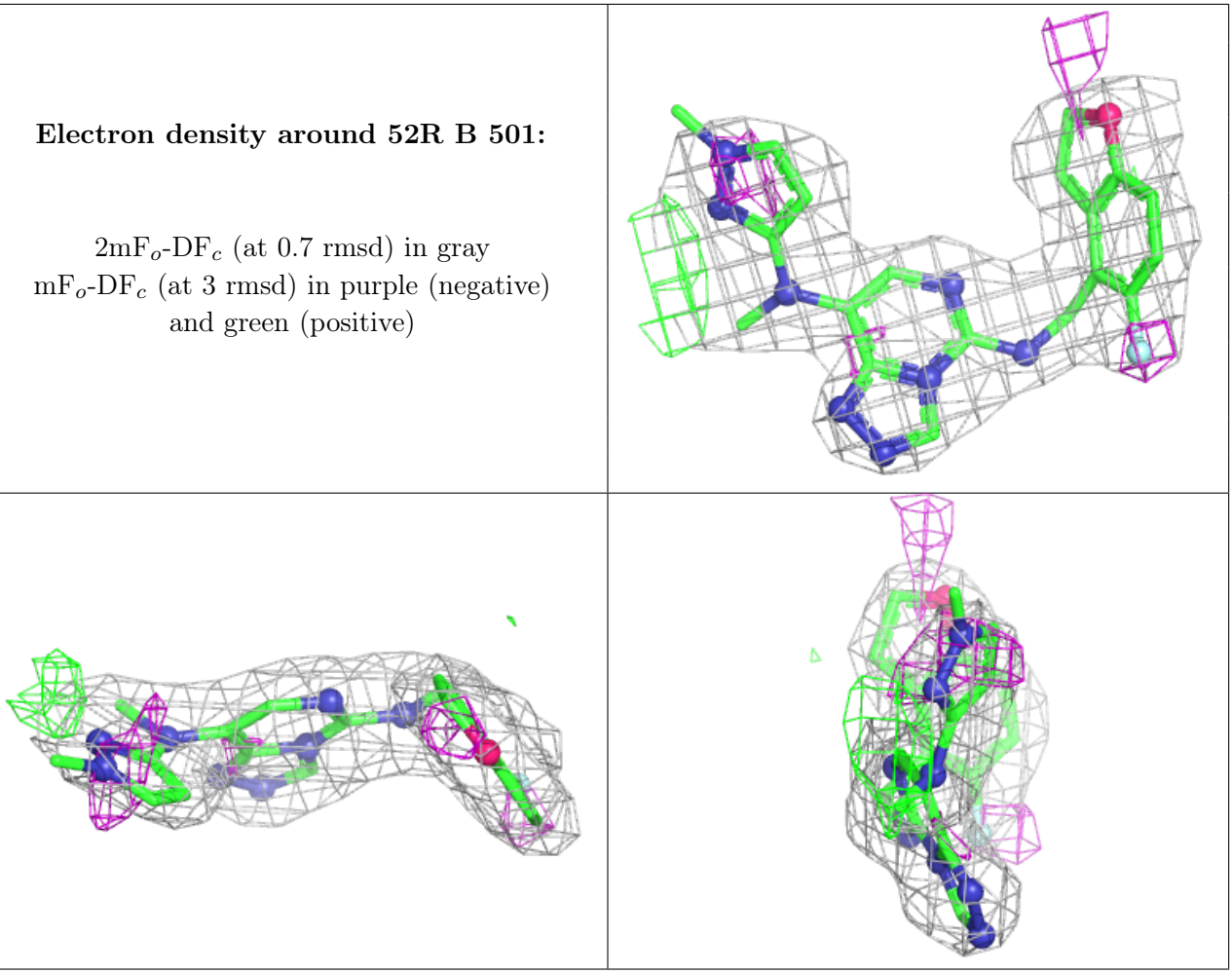






Electron density around 52R B 501:

$2mF_o$-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.