Full wwPDB X-ray Structure Validation Report (i)

Nov 9, 2021 – 11:11 AM EST

PDB ID	:	5Q1C
Title	:	Ligand binding to FARNESOID-X-RECEPTOR
Authors	:	Rudolph, M.G.; Benz, J.; Burger, D.; Thoma, R.; Ruf, A.; Joseph, C.; Kuhn, B.; Shao, C.; Yang, H.; Burley, S.K.
Deposited on	:	2017-05-31
Resolution	:	2.30 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:

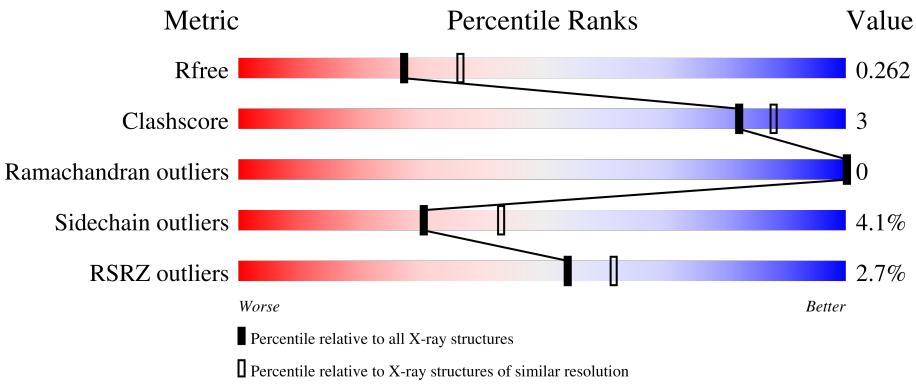
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	233	3%, 90%, 8%
1	C	233	3%, 89%, 8%
2	B	14	64%, 14%, 21%
2	D	14	50%, 21%, 29%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4211 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 Bile acid receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	229	1873	1198	316	348	11	0	0	0
1	C	228	1864	1193	314	346	11	0	0	0

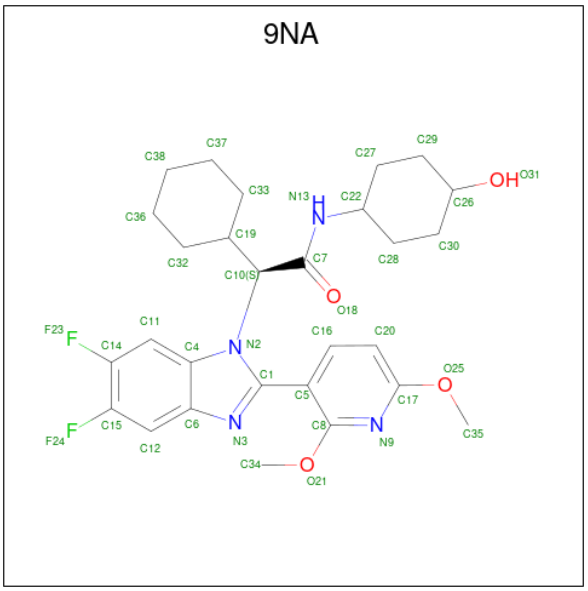
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	244	GLY	-	expression tag	UNP Q96RI1
A	245	SER	-	expression tag	UNP Q96RI1
A	246	HIS	-	expression tag	UNP Q96RI1
A	247	MET	-	expression tag	UNP Q96RI1
A	281	ALA	GLU	conflict	UNP Q96RI1
A	354	ALA	GLU	conflict	UNP Q96RI1
C	244	GLY	-	expression tag	UNP Q96RI1
C	245	SER	-	expression tag	UNP Q96RI1
C	246	HIS	-	expression tag	UNP Q96RI1
C	247	MET	-	expression tag	UNP Q96RI1
C	281	ALA	GLU	conflict	UNP Q96RI1
C	354	ALA	GLU	conflict	UNP Q96RI1

- Molecule 2 is a protein called COACTIVATOR PEPTIDE SRC-1 HD3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	B	11	96	63	18	15	0	0	0
2	D	10	91	60	17	14	0	0	0

- Molecule 3 is (2S)-2-cyclohexyl-2-[2-(2,6-dimethoxypyridin-3-yl)-5,6-difluoro-1H-benzimidazol-1-yl]-N-(trans-4-hydroxycyclohexyl)acetamide (three-letter code: 9NA) (formula: $C_{28}H_{34}F_2N_4O_4$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	F	N	O		
3	A	1	38	28	2	4	4	0	0
			Total	C	F	N	O		
3	C	1	38	28	2	4	4	0	0

- Molecule 4 is water.

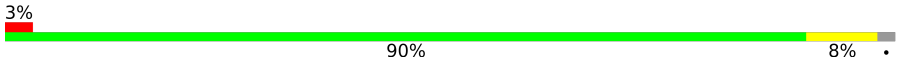
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	101	101	101	0	0
			Total	O		
4	B	7	7	7	0	0
			Total	O		
4	C	95	95	95	0	0
			Total	O		
4	D	8	8	8	0	0

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.

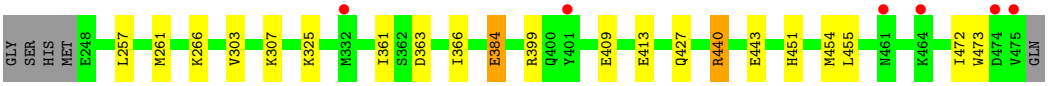
• Molecule 1: Bile acid receptor

Chain A:



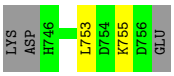
• Molecule 1: Bile acid receptor

Chain C:



• Molecule 2: COACTIVATOR PEPTIDE SRC-1 HD3

Chain B:



• Molecule 2: COACTIVATOR PEPTIDE SRC-1 HD3

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	72.48Å 84.46Å 190.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	27.25 – 2.30 27.22 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.5 (27.25-2.30) 99.5 (27.22-2.30)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.20 (at 2.31Å)	Xtriage
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.228 , 0.264 0.221 , 0.262	Depositor DCC
R_{free} test set	1303 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å^2)	43.1	Xtriage
Anisotropy	0.312	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 38.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4211	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.56% 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: 9NA

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	0.44	0/1913	0.59	0/2585
1	C	0.44	0/1904	0.59	0/2573
2	B	0.44	0/97	0.64	0/129
2	D	0.38	0/92	0.63	0/122
All	All	0.44	0/4006	0.59	0/5409

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	1873	0	1874	8	0
1	C	1864	0	1866	12	0
2	B	96	0	99	1	0
2	D	91	0	97	2	0
3	A	38	0	0	0	0
3	C	38	0	0	0	0
4	A	101	0	0	1	0
4	B	7	0	0	0	0
4	C	95	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	8	0	0	0	0
All	All	4211	0	3936	20	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 (20) 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:363:ASP:HA	1:A:366:ILE:HD12	1.76	0.66
1:C:363:ASP:HA	1:C:366:ILE:HD12	1.76	0.65
1:A:440:ARG:HH11	1:A:440:ARG:HA	1.64	0.62
1:C:440:ARG:HH11	1:C:440:ARG:HA	1.66	0.61
1:C:257:LEU:HD11	1:C:384:GLU:HB3	1.95	0.48
1:C:409:GLU:O	1:C:413:GLU:HG3	2.14	0.48
1:C:451:HIS:NE2	1:C:455:LEU:HD11	2.30	0.46
1:C:303:VAL:HG11	2:D:752:LEU:HB3	1.96	0.46
1:C:384:GLU:H	1:C:384:GLU:HG2	1.58	0.46
1:C:440:ARG:HA	1:C:440:ARG:NH1	2.32	0.45
1:A:324:LEU:HD12	2:B:753:LEU:HD11	1.99	0.45
1:C:472:ILE:HG22	1:C:473:TRP:CD1	2.51	0.44
1:A:289:LEU:HD12	4:A:673:HOH:O	2.17	0.44
1:C:307:LYS:HE3	2:D:753:LEU:O	2.18	0.43
1:A:440:ARG:HA	1:A:440:ARG:NH1	2.30	0.43
1:C:261:MET:SD	1:C:384:GLU:HA	2.59	0.42
1:A:472:ILE:HG22	1:A:473:TRP:CD1	2.55	0.42
1:A:379:LEU:HD21	1:A:436:CYS:HB3	2.03	0.41
1:A:334:LEU:HB2	1:A:444:LEU:HD11	2.03	0.41
1:C:361:ILE:HD11	1:C:454:MET:HE2	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	227/233 (97%)	223 (98%)	4 (2%)	0	100	100
1	C	226/233 (97%)	223 (99%)	3 (1%)	0	100	100
2	B	9/14 (64%)	9 (100%)	0	0	100	100
2	D	8/14 (57%)	8 (100%)	0	0	100	100
All	All	470/494 (95%)	463 (98%)	7 (2%)	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	210/213 (99%)	201 (96%)	9 (4%)	29	40
1	C	209/213 (98%)	202 (97%)	7 (3%)	38	53
2	B	10/14 (71%)	9 (90%)	1 (10%)	7	9
2	D	10/14 (71%)	9 (90%)	1 (10%)	7	9
All	All	439/454 (97%)	421 (96%)	18 (4%)	30	43

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	318	GLU
1	A	325	LYS
1	A	384	GLU
1	A	399	ARG
1	A	413	GLU
1	A	440	ARG
1	A	443	GLU
1	A	459	ARG
1	A	464	LYS
2	B	755	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	266	LYS
1	C	325	LYS
1	C	384	GLU
1	C	399	ARG
1	C	427	GLN
1	C	440	ARG
1	C	443	GLU
2	D	755	LYS

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	348	HIS
1	C	298	HIS
1	C	348	HIS
1	C	383	GLN
2	D	746	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)

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	$\#\|Z\| > 2$
3	9NA	C	501	-	39,42,42	0.88	0	48,60,60	0.78	2 (4%)
3	9NA	A	501	-	39,42,42	0.87	1 (2%)	48,60,60	0.76	2 (4%)

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	9NA	C	501	-	-	4/18/42/42	0/5/5/5
3	9NA	A	501	-	-	11/18/42/42	0/5/5/5

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	501	9NA	C30-C26	2.09	1.56	1.51

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	501	9NA	C14-C11-C4	-2.52	116.36	118.81
3	A	501	9NA	C14-C11-C4	-2.44	116.44	118.81
3	C	501	9NA	C11-C4-C6	2.21	122.73	120.54
3	A	501	9NA	C11-C4-C6	2.01	122.53	120.54

There are no chirality outliers.

All (15) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	501	9NA	C5-C8-O21-C34
3	A	501	9NA	C7-C10-C19-C32
3	C	501	9NA	N9-C8-O21-C34
3	C	501	9NA	C7-C10-C19-C32
3	C	501	9NA	C5-C8-O21-C34
3	A	501	9NA	N9-C8-O21-C34
3	A	501	9NA	N3-C1-C5-C16
3	C	501	9NA	N3-C1-C5-C16

Continued on next page...

Continued from previous page...

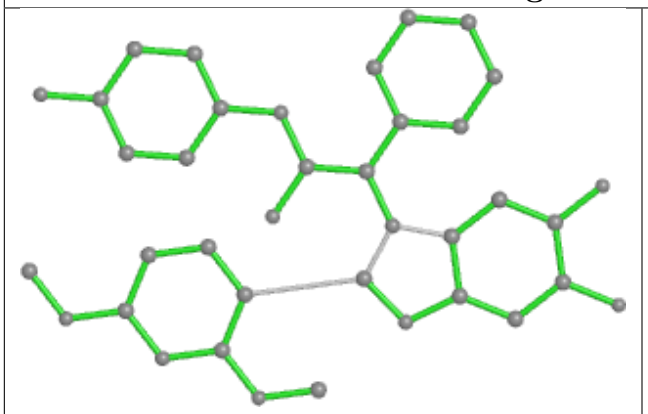
Mol	Chain	Res	Type	Atoms
3	A	501	9NA	C20-C17-O25-C35
3	A	501	9NA	N9-C17-O25-C35
3	A	501	9NA	C7-C10-C19-C33
3	A	501	9NA	N2-C10-C19-C32
3	A	501	9NA	N2-C10-C19-C33
3	A	501	9NA	N2-C10-C7-O18
3	A	501	9NA	N2-C10-C7-N13

There are no ring outliers.

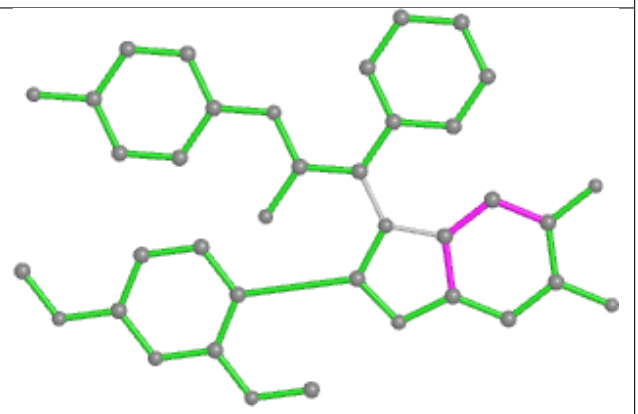
No monomer is involved in short contacts.

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.

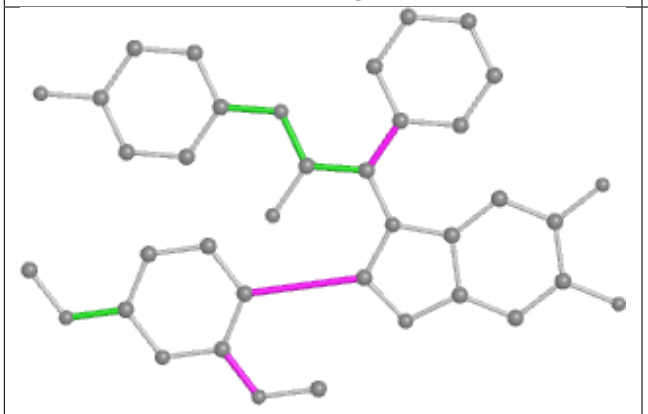
Ligand 9NA C 501



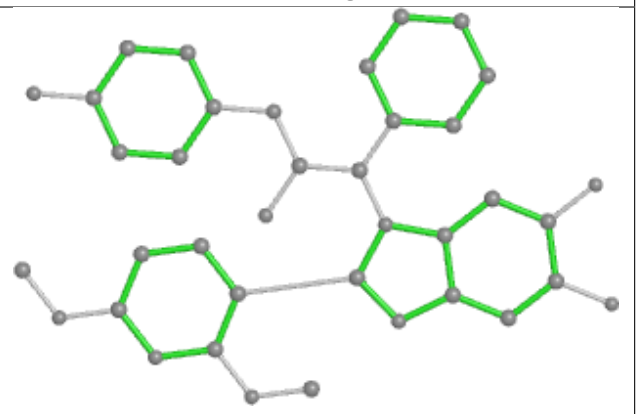
Bond lengths



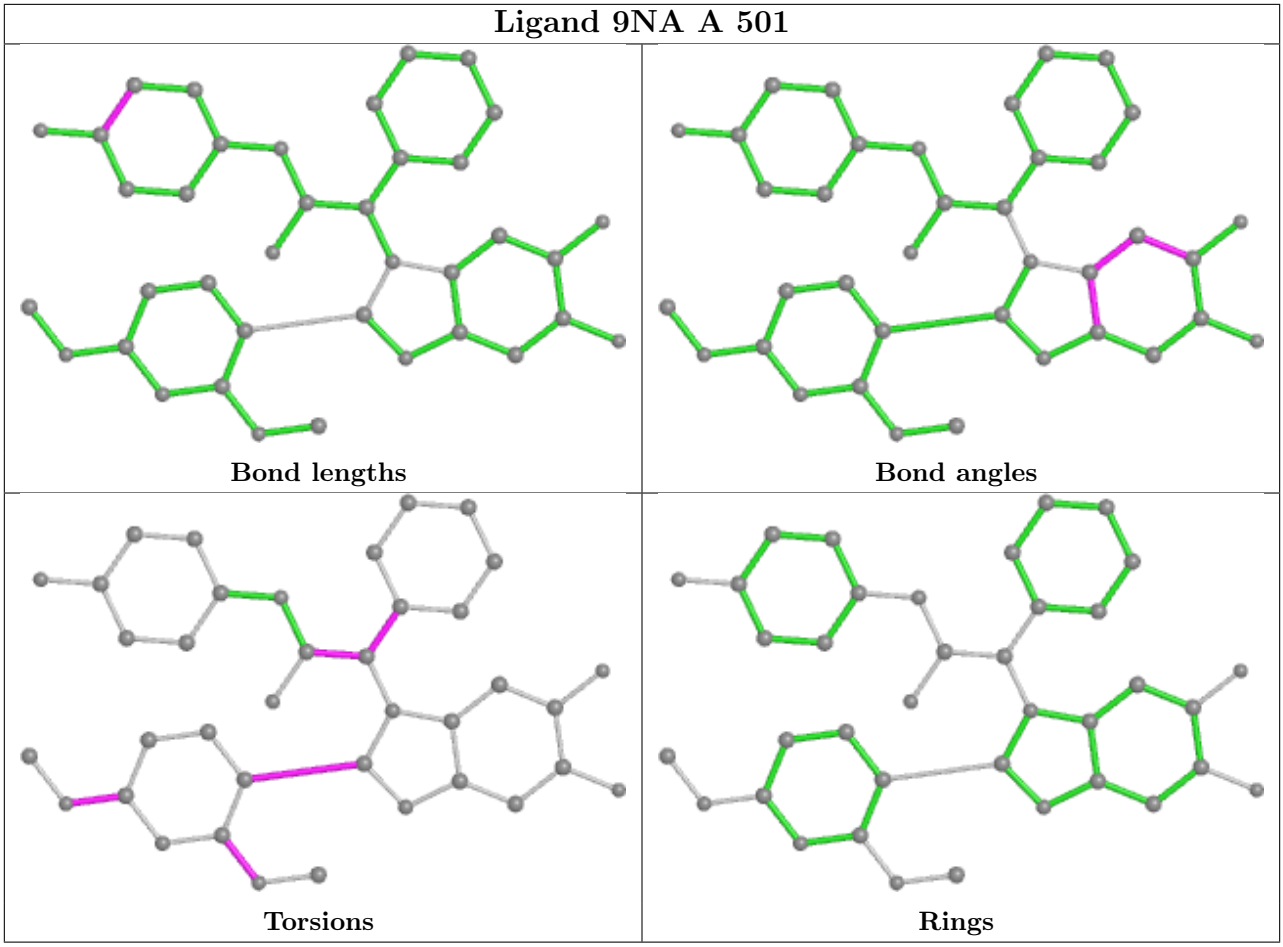
Bond angles



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 (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	229/233 (98%)	0.01	7 (3%) 49 56	29, 52, 87, 102	0
1	C	228/233 (97%)	0.11	6 (2%) 56 63	36, 57, 81, 106	0
2	B	11/14 (78%)	-0.48	0 100 100	42, 50, 78, 99	0
2	D	10/14 (71%)	0.13	0 100 100	55, 59, 83, 84	0
All	All	478/494 (96%)	0.05	13 (2%) 54 62	29, 55, 86, 106	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	474	ASP	4.0
1	A	429	GLU	3.9
1	A	346	SER	3.5
1	A	363	ASP	3.0
1	A	426	HIS	3.0
1	A	279	LYS	2.5
1	C	332	MET	2.4
1	C	475	VAL	2.4
1	C	461	ASN	2.2
1	C	401	TYR	2.2
1	A	343	LYS	2.2
1	A	449	HIS	2.2
1	C	464	LYS	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)

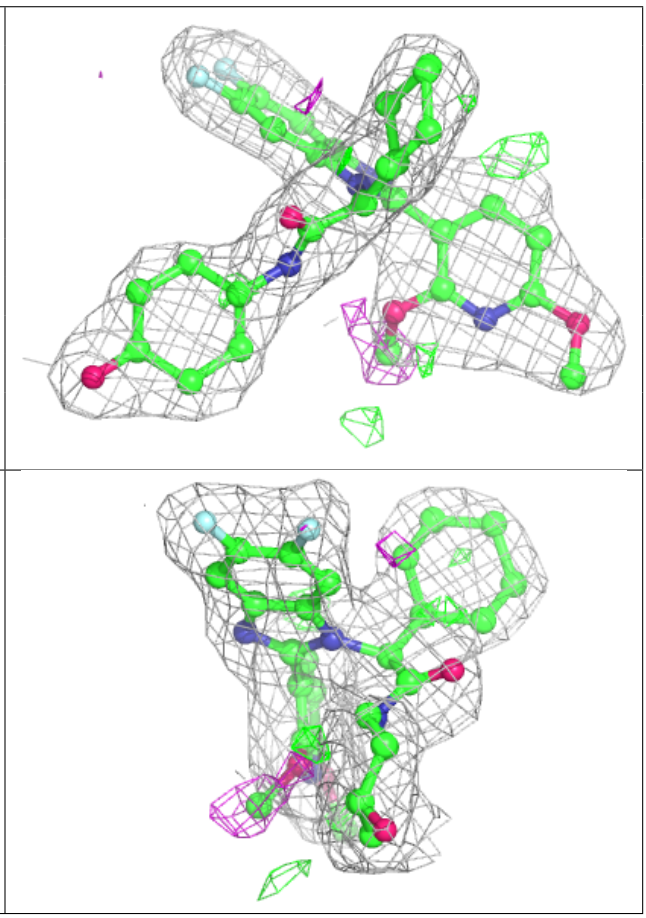
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists 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	9NA	A	501	38/38	0.94	0.19	33,40,47,52	0
3	9NA	C	501	38/38	0.94	0.21	34,49,56,60	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.

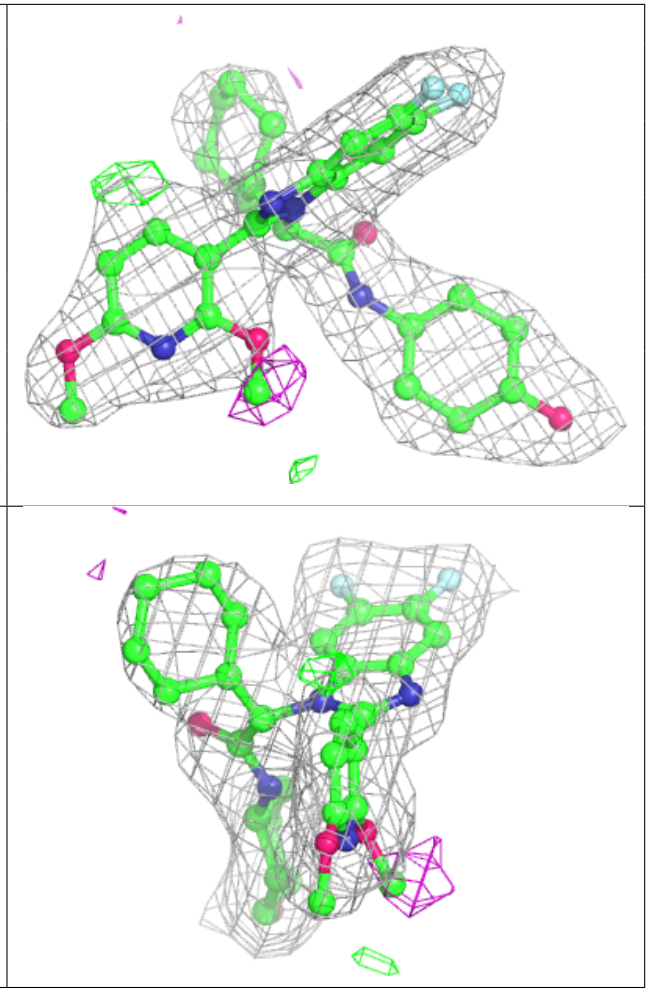
Electron density around 9NA A 501:

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 9NA C 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.