wwPDB X-ray Structure Validation Summary Report (i)

May 15, 2020 – 11:05 pm BST

PDB ID	:	6DUP
Title	:	CRYSTAL STRUCTURE OF PXR IN COMPLEX WITH COMPOUND 7
Authors	:	Chen, X.; Zhang, Y.; Mclean, L.R.
Deposited on	:	2018-06-21
Resolution	:	2.30 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 following versions of software and data (see references (i)) were used in the production of this report:

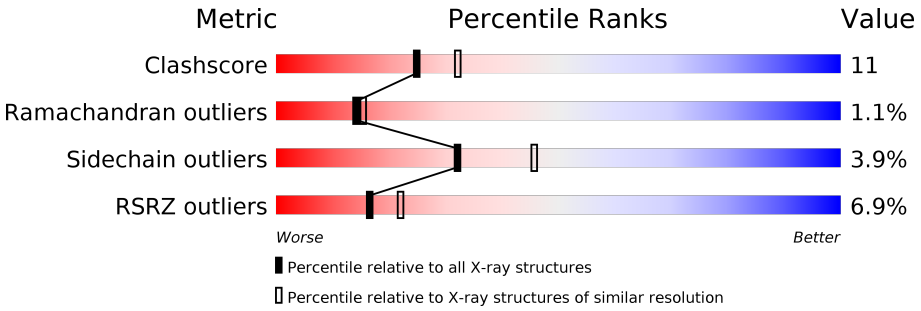
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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))
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 on 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	344	5% 63% 17% • 18%
1	B	344	6% 64% 17% • 19%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	HCJ	A	501	-	-	-	X

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4720 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 Nuclear receptor subfamily 1 group I member 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	282	Total	C	N	O	S	0	1	0
			2315	1486	400	410	19			
1	B	280	Total	C	N	O	S	0	1	0
			2299	1479	395	406	19			

There are 94 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	119	MET	-	expression tag	UNP O75469
A	120	LYS	-	expression tag	UNP O75469
A	121	LYS	-	expression tag	UNP O75469
A	122	GLY	-	expression tag	UNP O75469
A	123	HIS	-	expression tag	UNP O75469
A	124	HIS	-	expression tag	UNP O75469
A	125	HIS	-	expression tag	UNP O75469
A	126	HIS	-	expression tag	UNP O75469
A	127	HIS	-	expression tag	UNP O75469
A	128	HIS	-	expression tag	UNP O75469
A	129	GLY	-	expression tag	UNP O75469
A	427	GLU	-	expression tag	UNP O75469
A	428	LEU	-	expression tag	UNP O75469
A	429	PHE	-	expression tag	UNP O75469
A	430	GLY	-	expression tag	UNP O75469
A	431	ILE	-	expression tag	UNP O75469
A	432	THR	-	expression tag	UNP O75469
A	433	GLY	-	expression tag	UNP O75469
A	434	SER	-	expression tag	UNP O75469
A	435	GLY	-	expression tag	UNP O75469
A	436	GLY	-	expression tag	UNP O75469
A	437	SER	-	expression tag	UNP O75469
A	438	GLY	-	expression tag	UNP O75469
A	439	GLY	-	expression tag	UNP O75469
A	440	SER	-	expression tag	UNP O75469

Continued on next page...

Continued from previous page...

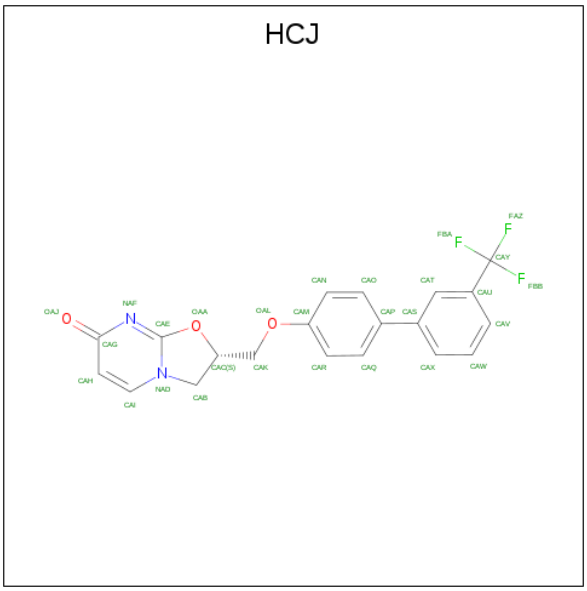
Chain	Residue	Modelled	Actual	Comment	Reference
A	441	SER	-	expression tag	UNP O75469
A	442	HIS	-	expression tag	UNP O75469
A	443	SER	-	expression tag	UNP O75469
A	444	SER	-	expression tag	UNP O75469
A	445	LEU	-	expression tag	UNP O75469
A	446	THR	-	expression tag	UNP O75469
A	447	GLU	-	expression tag	UNP O75469
A	448	ARG	-	expression tag	UNP O75469
A	449	HIS	-	expression tag	UNP O75469
A	450	LYS	-	expression tag	UNP O75469
A	451	ILE	-	expression tag	UNP O75469
A	452	LEU	-	expression tag	UNP O75469
A	453	HIS	-	expression tag	UNP O75469
A	454	ARG	-	expression tag	UNP O75469
A	455	LEU	-	expression tag	UNP O75469
A	456	LEU	-	expression tag	UNP O75469
A	457	GLN	-	expression tag	UNP O75469
A	458	GLU	-	expression tag	UNP O75469
A	459	GLY	-	expression tag	UNP O75469
A	460	SER	-	expression tag	UNP O75469
A	461	PRO	-	expression tag	UNP O75469
A	462	SER	-	expression tag	UNP O75469
B	119	MET	-	expression tag	UNP O75469
B	120	LYS	-	expression tag	UNP O75469
B	121	LYS	-	expression tag	UNP O75469
B	122	GLY	-	expression tag	UNP O75469
B	123	HIS	-	expression tag	UNP O75469
B	124	HIS	-	expression tag	UNP O75469
B	125	HIS	-	expression tag	UNP O75469
B	126	HIS	-	expression tag	UNP O75469
B	127	HIS	-	expression tag	UNP O75469
B	128	HIS	-	expression tag	UNP O75469
B	129	GLY	-	expression tag	UNP O75469
B	427	GLU	-	expression tag	UNP O75469
B	428	LEU	-	expression tag	UNP O75469
B	429	PHE	-	expression tag	UNP O75469
B	430	GLY	-	expression tag	UNP O75469
B	431	ILE	-	expression tag	UNP O75469
B	432	THR	-	expression tag	UNP O75469
B	433	GLY	-	expression tag	UNP O75469
B	434	SER	-	expression tag	UNP O75469
B	435	GLY	-	expression tag	UNP O75469

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	436	GLY	-	expression tag	UNP O75469
B	437	SER	-	expression tag	UNP O75469
B	438	GLY	-	expression tag	UNP O75469
B	439	GLY	-	expression tag	UNP O75469
B	440	SER	-	expression tag	UNP O75469
B	441	SER	-	expression tag	UNP O75469
B	442	HIS	-	expression tag	UNP O75469
B	443	SER	-	expression tag	UNP O75469
B	444	SER	-	expression tag	UNP O75469
B	445	LEU	-	expression tag	UNP O75469
B	446	THR	-	expression tag	UNP O75469
B	447	GLU	-	expression tag	UNP O75469
B	448	ARG	-	expression tag	UNP O75469
B	449	HIS	-	expression tag	UNP O75469
B	450	LYS	-	expression tag	UNP O75469
B	451	ILE	-	expression tag	UNP O75469
B	452	LEU	-	expression tag	UNP O75469
B	453	HIS	-	expression tag	UNP O75469
B	454	ARG	-	expression tag	UNP O75469
B	455	LEU	-	expression tag	UNP O75469
B	456	LEU	-	expression tag	UNP O75469
B	457	GLN	-	expression tag	UNP O75469
B	458	GLU	-	expression tag	UNP O75469
B	459	GLY	-	expression tag	UNP O75469
B	460	SER	-	expression tag	UNP O75469
B	461	PRO	-	expression tag	UNP O75469
B	462	SER	-	expression tag	UNP O75469

- Molecule 2 is (2S)-2-({[3'-(trifluoromethyl)[1,1'-biphenyl]-4-yl]oxy}methyl)-2,3-dihydro-7H-[1,3]oxazolo[3,2-a]pyrimidin-7-one (three-letter code: HCJ) (formula: $C_{20}H_{15}F_3N_2O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C F N O 28 20 3 2 3	0	0

- Molecule 3 is water.

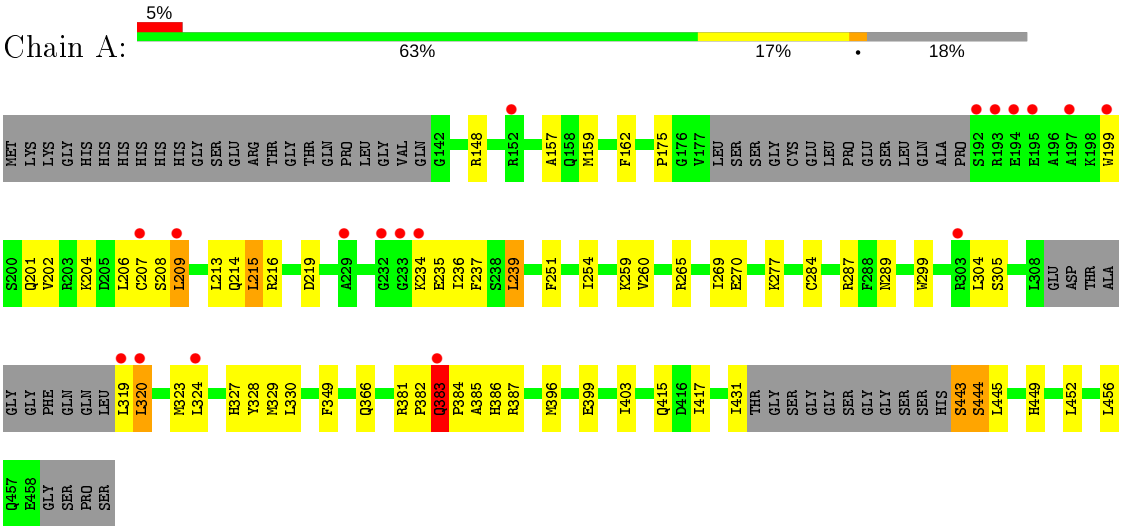
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	47	Total O 47 47	0	0
3	B	31	Total O 31 31	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

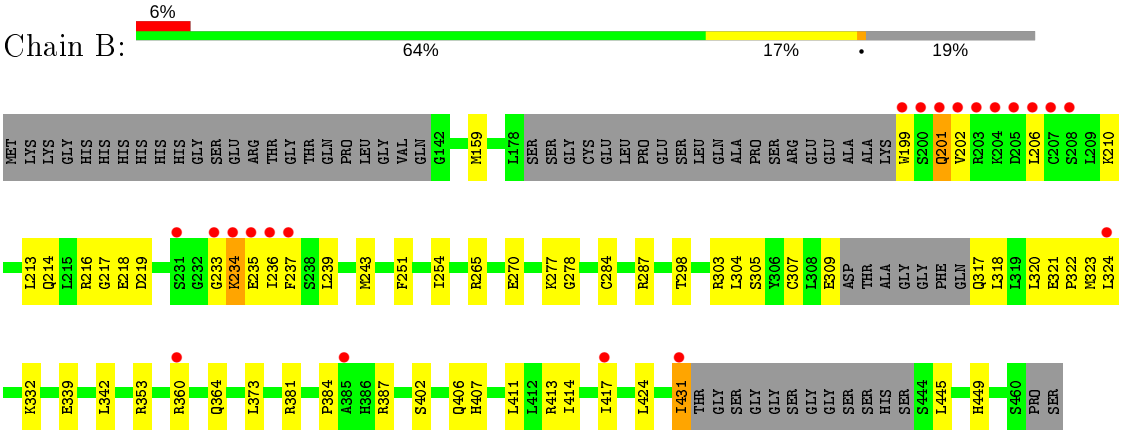
• Molecule 1: Nuclear receptor subfamily 1 group I member 2

Chain A:



• Molecule 1: Nuclear receptor subfamily 1 group I member 2

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	84.03Å 88.93Å 106.45Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.08 – 2.30 44.46 – 2.30	Depositor EDS
% Data completeness (in resolution range)	94.0 (39.08-2.30) 94.1 (44.46-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.34 (at 2.32Å)	Xtriage
Refinement program	CNX 2005	Depositor
R, R_{free}	0.226 , 0.266 0.225 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	46.7	Xtriage
Anisotropy	0.133	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 41.0	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	0.94	EDS
Total number of atoms	4720	wwPDB-VP
Average B, all atoms (Å^2)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.30% 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: HCJ

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	0.33	0/2364	0.48	0/3178
1	B	0.31	0/2348	0.45	0/3158
All	All	0.32	0/4712	0.47	0/6336

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

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	383	GLN	Peptide

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	2315	0	2335	58	0
1	B	2299	0	2321	41	0
2	A	28	0	0	4	0
3	A	47	0	0	0	0
3	B	31	0	0	0	0
All	All	4720	0	4656	100	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 11.

The worst 5 of 100 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:202:VAL:HG21	1:B:417:ILE:HD12	1.59	0.82
1:B:431:ILE:HD13	1:B:431:ILE:H	1.46	0.81
1:A:214:GLN:HE21	1:A:216:ARG:HH11	1.32	0.78
1:A:383:GLN:HB2	1:A:384:PRO:HD3	1.70	0.74
1:A:277:LYS:HD2	1:A:449:HIS:CE1	2.24	0.73

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	275/344 (80%)	260 (94%)	11 (4%)	4 (2%)	10	10
1	B	273/344 (79%)	262 (96%)	9 (3%)	2 (1%)	22	26
All	All	548/688 (80%)	522 (95%)	20 (4%)	6 (1%)	14	15

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	383	GLN
1	A	444	SER
1	B	218	GLU
1	B	234	LYS
1	A	208	SER

5.3.2 Protein sidechains

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	256/302 (85%)	244 (95%)	12 (5%)	26	37
1	B	255/302 (84%)	247 (97%)	8 (3%)	40	55
All	All	511/604 (85%)	491 (96%)	20 (4%)	32	46

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	383	GLN
1	A	443	SER
1	B	323	MET
1	A	320	LEU
1	A	329	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	407	HIS
1	A	449	HIS
1	B	272	GLN
1	A	386	HIS
1	B	317	GLN

5.3.3 RNA

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 carbohydrates 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	HCJ	A	501	-	28,31,31	1.59	4 (14%)	33,45,45	1.38	3 (9%)

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	HCJ	A	501	-	-	1/15/23/23	0/4/4/4

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	501	HCJ	CAE-NAF	4.42	1.40	1.32
2	A	501	HCJ	CAG-NAF	4.40	1.40	1.33
2	A	501	HCJ	CAS-CAP	-3.74	1.39	1.49
2	A	501	HCJ	CAI-NAD	2.84	1.40	1.35

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	501	HCJ	OAA-CAC-CAK	4.17	118.11	109.08
2	A	501	HCJ	OAA-CAC-CAB	-4.14	100.29	104.55
2	A	501	HCJ	OAL-CAK-CAC	2.67	114.32	107.94

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	501	HCJ	CAC-CAK-OAL-CAM

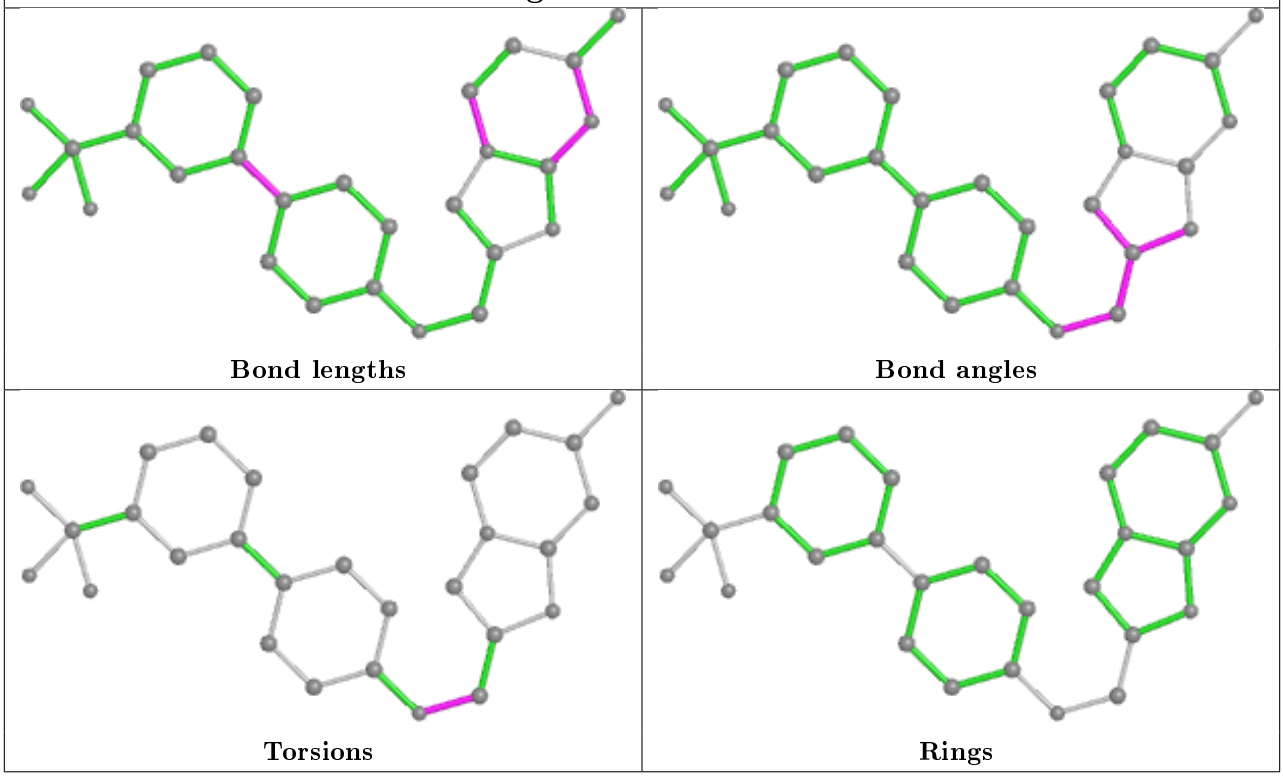
There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	501	HCJ	4	0

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 HCJ A 501



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	282/344 (81%)	0.33	18 (6%)	19	25	28, 54, 95, 123	0
1	B	280/344 (81%)	0.36	21 (7%)	14	19	26, 55, 95, 99	0
All	All	562/688 (81%)	0.35	39 (6%)	16	22	26, 54, 95, 123	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	209	LEU	6.1
1	A	193	ARG	6.0
1	B	431	ILE	5.5
1	A	233	GLY	5.3
1	B	207	CYS	4.6

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 carbohydrates in this entry.

6.4 Ligands

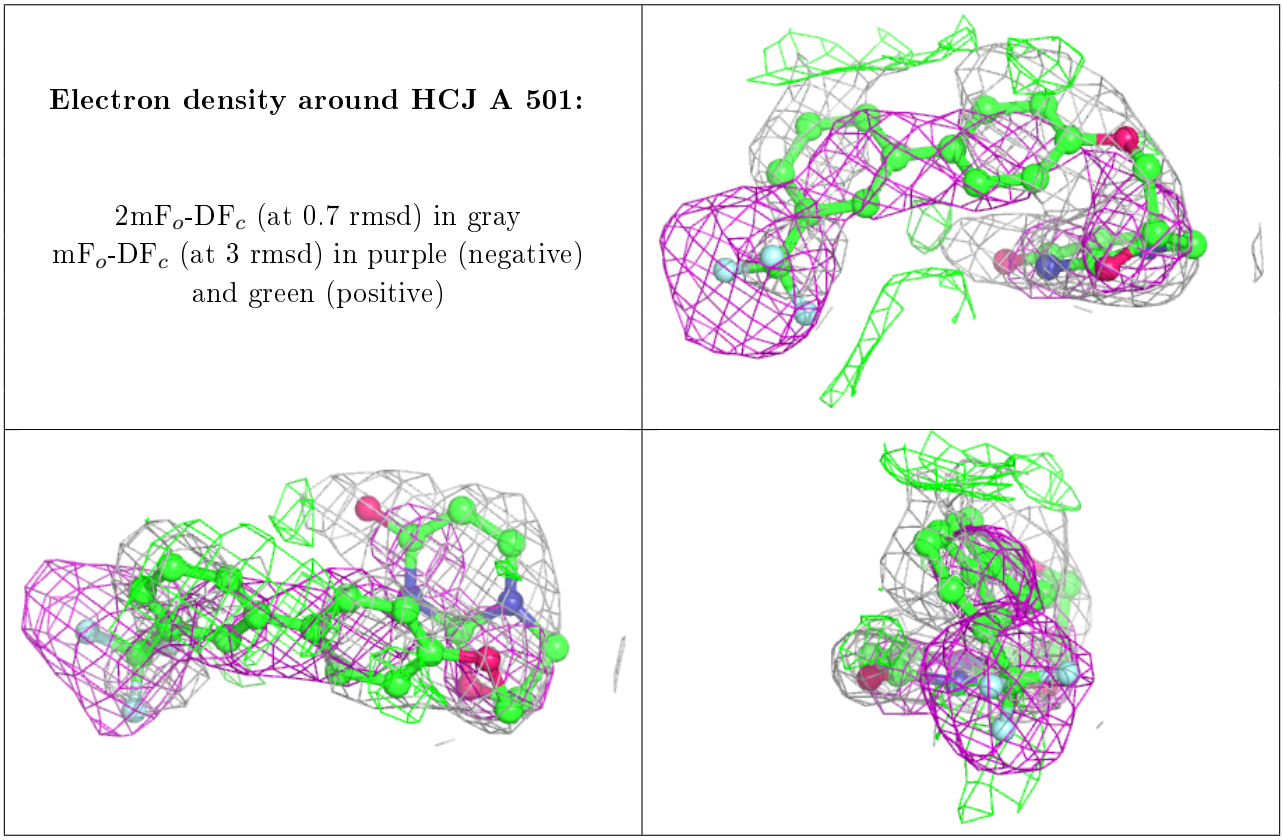
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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	HCJ	A	501	28/28	0.60	0.46	58,62,69,70	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 HCJ 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)



6.5 Other polymers

There are no such residues in this entry.