



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 23, 2020 – 08:52 pm BST

PDB ID : 4J08  
Title : Crystal structure of hcv ns5b polymerase in complex with 2-[[[4-METHYLPHENYL)SULFONYL]AMINO}-5-PHENOXYBENZOIC ACID  
Authors : Coulombe, R.  
Deposited on : 2013-01-30  
Resolution : 2.10 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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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

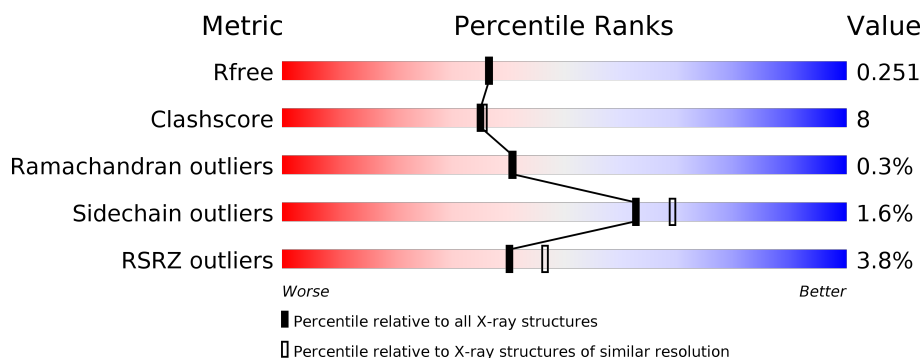
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	576	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="width: 100%; height: 10px; background-color: green; position: relative;"> <span style="position: absolute; left: 0; top: -5px;">%</span> <span style="position: absolute; left: 82%; top: -5px;">82%</span> <span style="position: absolute; left: 97%; top: -5px;">13%</span> <span style="position: absolute; left: 98%; top: -5px;">..</span> </div> </div>
1	B	576	<div> <div style="width: 6%; height: 10px; background-color: red;"></div> <div style="width: 100%; height: 10px; background-color: green; position: relative;"> <span style="position: absolute; left: 0; top: -5px;">6%</span> <span style="position: absolute; left: 77%; top: -5px;">77%</span> <span style="position: absolute; left: 97%; top: -5px;">19%</span> <span style="position: absolute; left: 98%; top: -5px;">..</span> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9027 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.

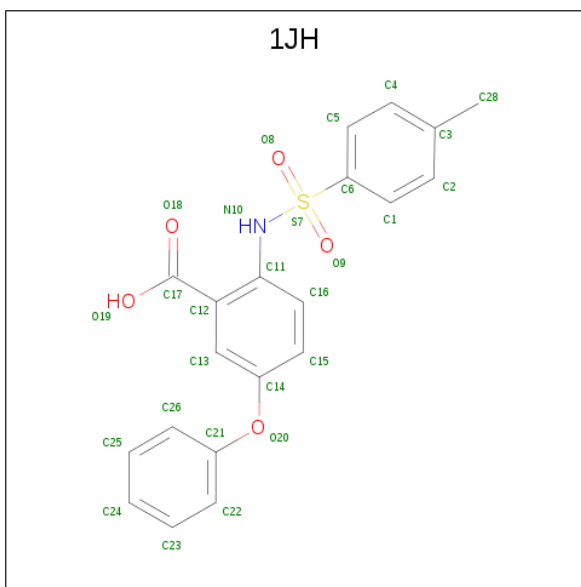
- Molecule 1 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	559	Total	C	N	O	S	0	0	0
			4358	2745	770	811	32			
1	B	558	Total	C	N	O	S	0	0	0
			4346	2737	768	809	32			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	571	HIS	-	EXPRESSION TAG	UNP O92972
A	572	HIS	-	EXPRESSION TAG	UNP O92972
A	573	HIS	-	EXPRESSION TAG	UNP O92972
A	574	HIS	-	EXPRESSION TAG	UNP O92972
A	575	HIS	-	EXPRESSION TAG	UNP O92972
A	576	HIS	-	EXPRESSION TAG	UNP O92972
B	571	HIS	-	EXPRESSION TAG	UNP O92972
B	572	HIS	-	EXPRESSION TAG	UNP O92972
B	573	HIS	-	EXPRESSION TAG	UNP O92972
B	574	HIS	-	EXPRESSION TAG	UNP O92972
B	575	HIS	-	EXPRESSION TAG	UNP O92972
B	576	HIS	-	EXPRESSION TAG	UNP O92972

- Molecule 2 is 2-[[[(4-methylphenyl)sulfonyl]amino]-5-phenoxybenzoic acid (three-letter code: 1JH) (formula: C<sub>20</sub>H<sub>17</sub>NO<sub>5</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			27	20	1	5	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	181	Total	O	0	0
			181	181		
3	B	115	Total	O	0	0
			115	115		

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.

Chain A:

82% 13%

Residues (Chain A):

Residue	State
S1	Green
P22	Green
I23	Green
N24	Green
S27	Green
L31	Green
R32	Green
P36	Green
R48	Green
L57	Green
A73	Green
S76	Green
A80	Green
K81	Green
L82	Green
L83	Green
S84	Green
I85	Green
E86	Green
E87	Green
P93	Green
P94	Green
G102	Green
R114	Green
I119	Green
V144	Green
P145	Green
C146	Green
P149	Green
GLU	Green
LYS	Green
GLY	Green
GLY	Green
R154	Green
I160	Green
L165	Green
V169	Green
K172	Green
M173	Green
P182	Green
Q184	Yellow
I233	Green
R234	Green
L245	Green
E248	Green
Q251	Green
R254	Green
T257	Green
Y261	Green
I262	Green
M268	Green
Q272	Green
M273	Green
C274	Green
C289	Green
T300	Green
K307	Green
K313	Green
L314	Green
V321	Green
L336	Yellow
Y346	Green
S347	Green
G351	Green
P356	Green
LEU	Green
SER	Green
ARG	Green
ALA	Green
ARG	Green
PRO	Green
ARG	Green
HIS	Green
HIS	Green
HIS	Green
HIS	Green
HIS	Green
HIS	Green
R401	Yellow
R402	Yellow
L434	Yellow
D444	Yellow

Chain B:

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	106.83Å 107.68Å 135.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.10 41.54 – 2.10	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-2.10) 99.6 (41.54-2.10)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.71 (at 2.10Å)	Xtriage
Refinement program	CNX 2002	Depositor
R, $R_{free}$	0.225 , 0.255 0.220 , 0.251	Depositor DCC
$R_{free}$ test set	9100 reflections (9.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	32.5	Xtriage
Anisotropy	0.376	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 41.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.013 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9027	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.78% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: 1JH

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	0.34	0/4453	0.62	2/6044 (0.0%)
1	B	0.35	0/4440	0.60	1/6025 (0.0%)
All	All	0.34	0/8893	0.61	3/12069 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	31	LEU	CA-CB-CG	5.52	127.99	115.30
1	B	351	GLY	N-CA-C	-5.22	100.04	113.10
1	A	351	GLY	N-CA-C	-5.13	100.27	113.10

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	4358	0	4371	60	0
1	B	4346	0	4359	86	0
2	A	27	0	16	0	0
3	A	181	0	0	4	0
3	B	115	0	0	0	0
All	All	9027	0	8746	143	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 8.

The worst 5 of 143 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:81:LYS:O	1:A:82:LEU:HB3	1.74	0.87
1:B:515:GLY:HA2	1:B:519:ALA:HB2	1.57	0.85
1:B:531:ARG:H	1:B:531:ARG:HH11	1.29	0.80
1:B:158:ARG:HG3	1:B:158:ARG:HH11	1.49	0.78
1:A:465:ARG:NH1	1:A:545:LEU:O	2.17	0.78

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	555/576 (96%)	538 (97%)	16 (3%)	1 (0%)	47	49
1	B	554/576 (96%)	528 (95%)	24 (4%)	2 (0%)	34	32
All	All	1109/1152 (96%)	1066 (96%)	40 (4%)	3 (0%)	41	41

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	82	LEU
1	B	541	ALA
1	B	515	GLY

### 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	477/491 (97%)	468 (98%)	9 (2%)	57	63
1	B	475/491 (97%)	469 (99%)	6 (1%)	69	75
All	All	952/982 (97%)	937 (98%)	15 (2%)	62	69

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	381	VAL
1	A	402	HIS
1	B	313	MET
1	A	336	LEU
1	B	310	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	544	GLN
1	A	562	HIS
1	B	309	GLN
1	A	514	GLN
1	B	273	ASN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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	1JH	A	601	-	27,29,29	2.71	19 (70%)	36,41,41	1.07	3 (8%)

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	1JH	A	601	-	-	0/15/19/19	0/3/3/3

The worst 5 of 19 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	1JH	S7-N10	6.72	1.74	1.63
2	A	601	1JH	C22-C21	3.54	1.45	1.38
2	A	601	1JH	C12-C11	3.52	1.45	1.40
2	A	601	1JH	O20-C21	3.36	1.47	1.39
2	A	601	1JH	C26-C21	3.22	1.45	1.38

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	601	1JH	O9-S7-C6	-2.69	104.66	107.97
2	A	601	1JH	O8-S7-N10	2.47	112.92	106.73
2	A	601	1JH	O20-C14-C15	2.43	127.93	119.38

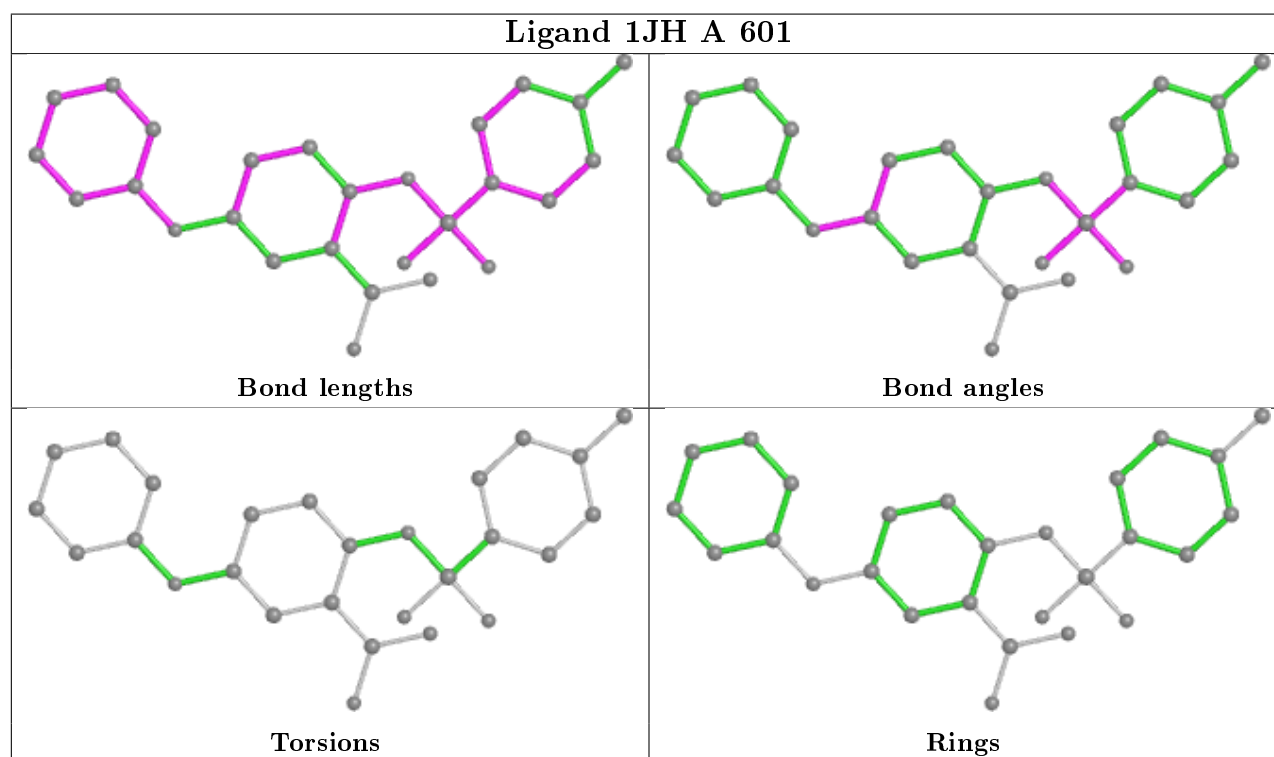
There are no chirality outliers.

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 than 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.



## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	559/576 (97%)	-0.12	7 (1%) 77 80	20, 31, 48, 57	0
1	B	558/576 (96%)	0.29	36 (6%) 18 23	22, 39, 54, 60	0
All	All	1117/1152 (96%)	0.08	43 (3%) 40 46	20, 35, 52, 60	0

The worst 5 of 43 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	23	ILE	5.9
1	B	24	ASN	5.5
1	B	542	ALA	4.6
1	B	543	SER	4.3
1	B	25	PRO	4.1

### 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](#)

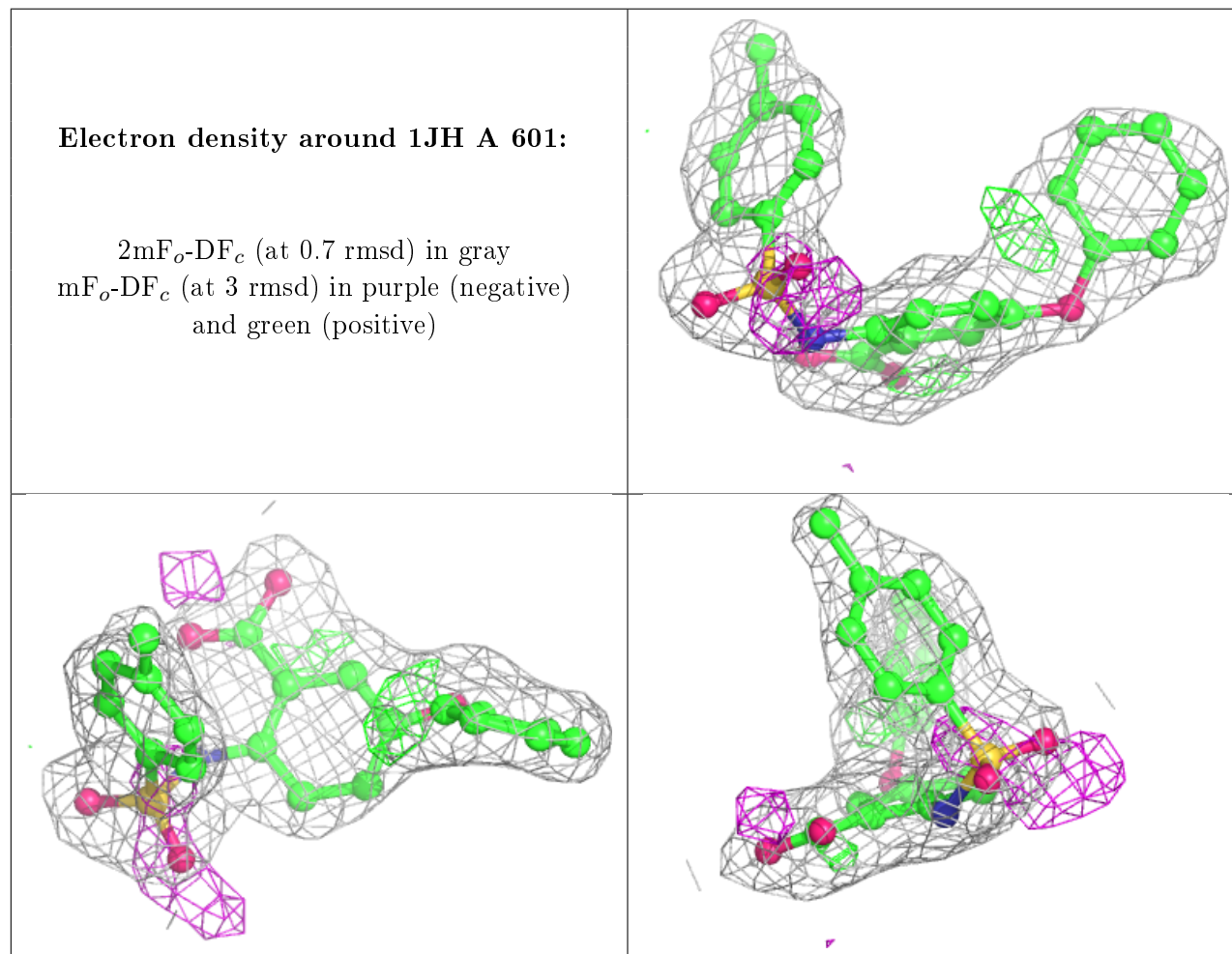
There are no carbohydrates in this entry.

### 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<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
2	1JH	A	601	27/27	0.89	0.14	37,44,46,49	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.



## 6.5 Other polymers [i](#)

There are no such residues in this entry.