



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 06:34 am BST

PDB ID : 4GMC  
Title : Crystal structure of HCV NS5B polymerase in complex with a thumb inhibitor  
Authors : Coulombe, R.  
Deposited on : 2012-08-15  
Resolution : 2.70 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

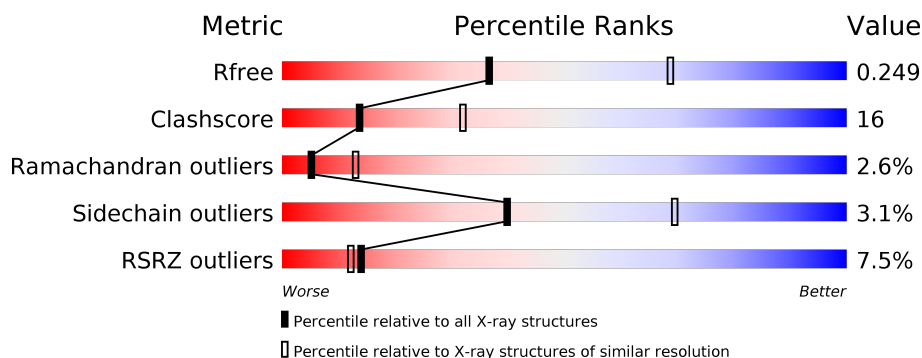
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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 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>%</div> <div> <div></div> <div>72%</div> <div>23%</div> <div>• •</div> </div> </div>
1	B	576	<div> <div>13%</div> <div> <div></div> <div>57%</div> <div>32%</div> <div>5% • 6%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8692 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 NS5B polymerase.

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	541	Total	C	N	O	S	0	0	0
			4210	2652	741	785	32			

There are 12 discrepancies between the modelled and reference sequences:

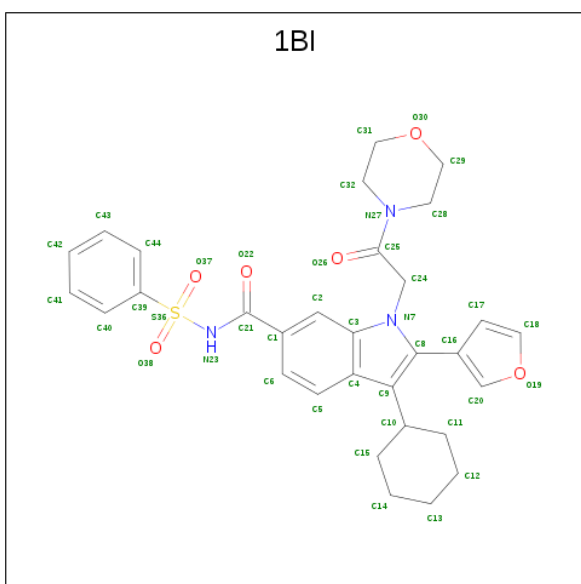
Chain	Residue	Modelled	Actual	Comment	Reference
A	571	HIS	-	EXPRESSION TAG	UNP O92972
A	572	HIS	-	EXPRESSION TAG	UNP O92972
A	573	HIS	-	EXPRESSION TAG	UNP O92972
A	574	HIS	-	EXPRESSION TAG	UNP O92972
A	575	HIS	-	EXPRESSION TAG	UNP O92972
A	576	HIS	-	EXPRESSION TAG	UNP O92972
B	571	HIS	-	EXPRESSION TAG	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 SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		

- Molecule 3 is 3-cyclohexyl-2-(furan-3-yl)-1-[2-(morpholin-4-yl)-2-oxoethyl]-N-(phenylsulfonyl)-1H-indole-6-carboxamide (three-letter code: 1BI) (formula: C<sub>31</sub>H<sub>33</sub>N<sub>3</sub>O<sub>6</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	B	1	Total	C	N	O	S	0	0
			41	31	3	6	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	42	Total	O	0	0
			42	42		
4	B	21	Total	O	0	0
			21	21		

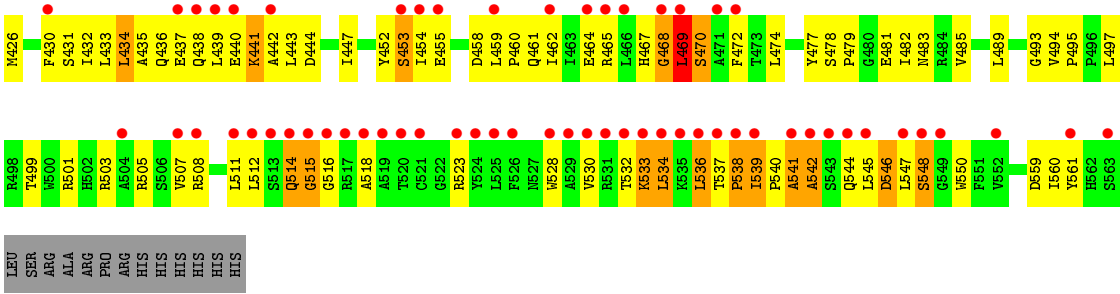
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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 ( $\text{RSRZ} > 2$ ). 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.

[illegible]

Chain B:

13% 57% 32% 5% 6%

S1 T5 I11 T12 P13 C14 A15 A16 E17 GLJ SER LYS LEU PRO ILE ASN PRO LEU ASN SER LEU LEU ARG HIS HIS ASN H36 V37 Y38 A39 GLY T40 T41 S42 S46 L47 R48 Q49 V52 A73 S76 T77 L82 L83 S84 S85 T92 P93 P94 H95 S96 S99 A105 K106 D107 V108 R109 S112 H118 I119 R120 S121 D129 I134 T138 M139 A140 K141 S142 E143 V144 F145 C146 V147 Q148 P149 LYS GLY G153 R154 A157 R158 L159 I160 L165 C170 M173 A174 L175 M187 Y191 G192 P193 Q194 S196 G199 R200 V201 W208 K212 C213 M215 D220 T221 R222 E248 A249 R250 Q251 A252 L253 R259 G264 P265 L266 M273 C274 Q275 Y276 L285 Y296 L297 K298 A306 K307 L308 Q309 D310 V321 G324 E325 S326 A327 G328 T329 O330 E331 A335 L336 S346 G351 P356 E357 Y358 D359 L362 I363 T364 S367 S368 V372 A373 A376 S377 Y381 Y382 Y383 L384 T385 R386 D387 P388 T389 T390 P391 R394 E398 R401 H402 T403 P404 I405 A406 S407 W408 L409 I412 I413 M414 Y415 A416 P417 T418 A421 R422 M423 I424 G425



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	104.24Å 106.26Å 133.33Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.70 49.65 – 2.60	Depositor EDS
% Data completeness (in resolution range)	100.0 (50.00-2.70) 99.7 (49.65-2.60)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.39 (at 2.58Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.219 , 0.257 0.211 , 0.249	Depositor DCC
$R_{free}$ test set	4643 reflections (10.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	47.7	Xtriage
Anisotropy	0.429	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 40.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.017 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	8692	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.97% 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: 1BI, SO4

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	0.38	0/4453	0.63	1/6044 (0.0%)
1	B	0.39	0/4299	0.71	3/5831 (0.1%)
All	All	0.38	0/8752	0.67	4/11875 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	516	GLY	N-CA-C	-11.62	84.06	113.10
1	A	351	GLY	N-CA-C	-5.67	98.92	113.10
1	B	402	HIS	N-CA-C	5.45	125.72	111.00
1	B	11	ILE	N-CA-C	-5.05	97.36	111.00

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	91	0
1	B	4210	0	4218	187	0
2	A	10	0	0	1	0
2	B	10	0	0	0	0
3	B	41	0	33	2	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	42	0	0	1	0
4	B	21	0	0	1	0
All	All	8692	0	8622	275	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 16.

The worst 5 of 275 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:536:LEU:HD12	1:B:536:LEU:H	1.14	1.10
1:B:405:ILE:HG21	1:B:443:LEU:HD13	1.43	1.00
1:B:13:PRO:HB3	1:B:17:GLU:HB2	1.45	0.97
1:B:434:LEU:HD23	1:B:434:LEU:H	1.38	0.89
1:B:514:GLN:HG2	1:B:515:GLY:N	1.85	0.88

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%)	521 (94%)	31 (6%)	3 (0%)	29	54
1	B	535/576 (93%)	457 (85%)	53 (10%)	25 (5%)	2	4
All	All	1090/1152 (95%)	978 (90%)	84 (8%)	28 (3%)	5	13

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	14	CYS
1	B	16	ALA

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	329	THR
1	B	403	THR
1	B	404	PRO

### 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	477/491 (97%)	461 (97%)	16 (3%)	37	66
1	B	458/491 (93%)	445 (97%)	13 (3%)	43	73
All	All	935/982 (95%)	906 (97%)	29 (3%)	40	69

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	474	LEU
1	B	41	THR
1	B	536	LEU
1	A	492	LEU
1	B	139	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	49	GLN
1	B	206	ASN
1	B	436	GLN
1	A	544	GLN
1	B	406	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 ⓘ

5 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	1BI	B	601	-	38,46,46	3.54	23 (60%)	45,66,66	2.27	8 (17%)
2	SO4	B	603	-	4,4,4	0.28	0	6,6,6	0.11	0
2	SO4	A	601	-	4,4,4	0.29	0	6,6,6	0.06	0
2	SO4	B	602	-	4,4,4	0.28	0	6,6,6	0.07	0
2	SO4	A	602	-	4,4,4	0.22	0	6,6,6	0.14	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
3	1BI	B	601	-	-	0/23/47/47	0/6/6/6

The worst 5 of 23 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	601	1BI	C21-N23	9.07	1.50	1.39
3	B	601	1BI	C25-N27	6.26	1.47	1.35
3	B	601	1BI	C9-C4	6.22	1.47	1.40
3	B	601	1BI	O37-S36	-6.01	1.36	1.43
3	B	601	1BI	C2-C1	5.83	1.47	1.37

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	601	1BI	C1-C21-N23	-7.91	106.51	116.09
3	B	601	1BI	O38-S36-O37	-7.54	110.28	119.55
3	B	601	1BI	O22-C21-N23	4.58	126.70	121.08
3	B	601	1BI	C25-C24-N7	4.17	115.72	110.76
3	B	601	1BI	C21-N23-S36	-3.89	118.34	123.36

There are no chirality outliers.

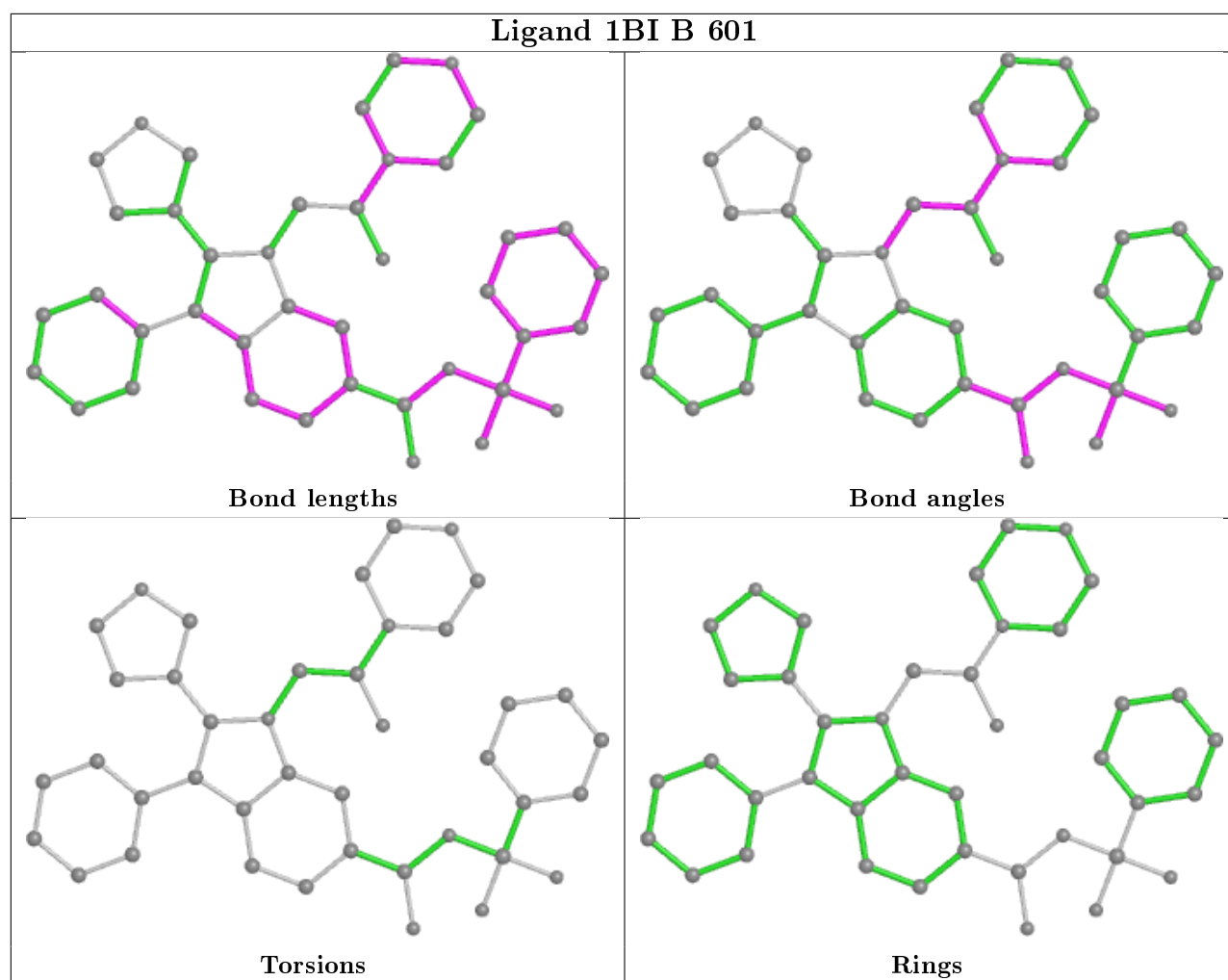
There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	601	1BI	2	0
2	A	601	SO4	1	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 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%)	-0.11	5 (0%) 84 85	23, 38, 63, 91	0
1	B	541/576 (93%)	0.59	77 (14%) 2 1	22, 58, 109, 119	0
All	All	1100/1152 (95%)	0.23	82 (7%) 14 12	22, 43, 101, 119	0

The worst 5 of 82 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	511	LEU	7.2
1	B	524	TYR	6.5
1	B	542	ALA	6.3
1	B	534	LEU	5.9
1	B	518	ALA	5.6

### 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](#)

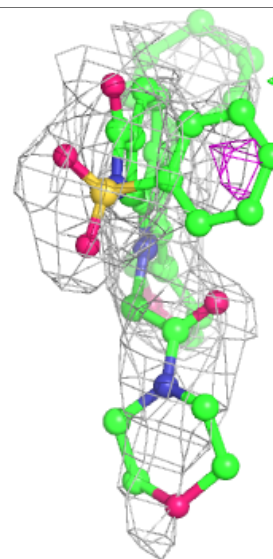
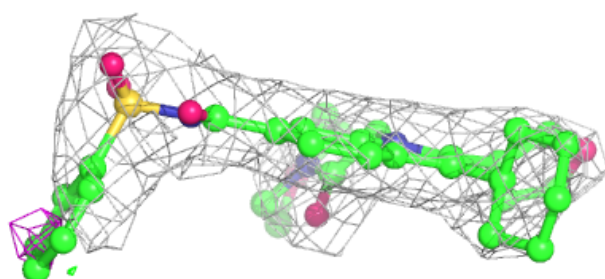
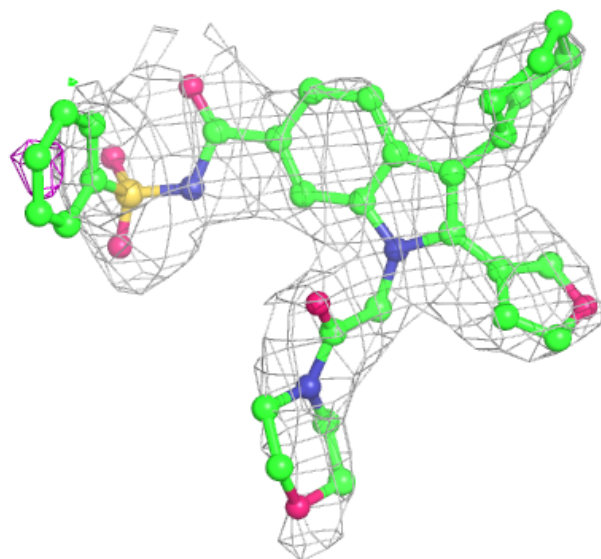
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 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
3	1BI	B	601	41/41	0.87	0.21	64,73,83,85	0
2	SO4	B	603	5/5	0.88	0.31	103,103,104,104	0
2	SO4	A	601	5/5	0.92	0.20	95,95,95,96	0
2	SO4	B	602	5/5	0.93	0.20	104,104,105,105	0
2	SO4	A	602	5/5	0.95	0.20	74,75,75,76	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 1BI B 601:**

2mF<sub>o</sub>-DF<sub>c</sub> (at 0.7 rmsd) in gray  
mF<sub>o</sub>-DF<sub>c</sub> (at 3 rmsd) in purple (negative)  
and green (positive)





## 6.5 Other polymers [i](#)

There are no such residues in this entry.