



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:55 PM GMT

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  
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

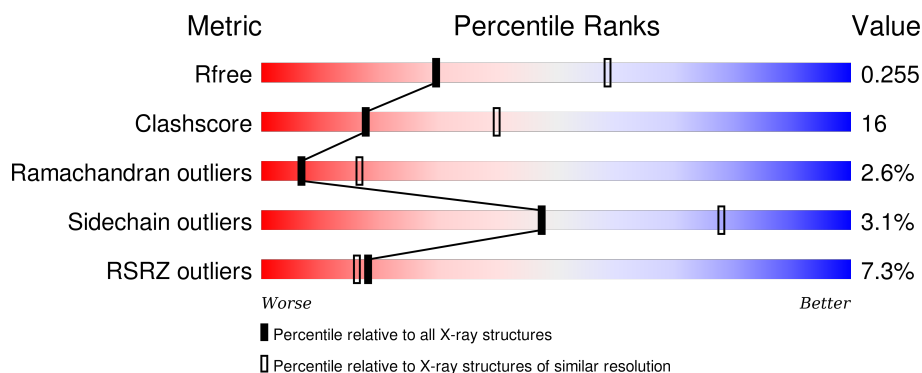
# 1 Overall quality at a glance

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))
$R_{free}$	91344	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (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. 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: linear-gradient(to right, red 1%, green 72%, yellow 23%, grey 5%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>%</span> <span>72%</span> <span>23%</span> <span>• •</span> </div> </div>
1	B	576	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 13%, green 57%, yellow 32%, orange 5%, grey 6%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>13%</span> <span>57%</span> <span>32%</span> <span>5% • 6%</span> </div> </div>

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	SO4	B	603	-	-	-	X

## 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:

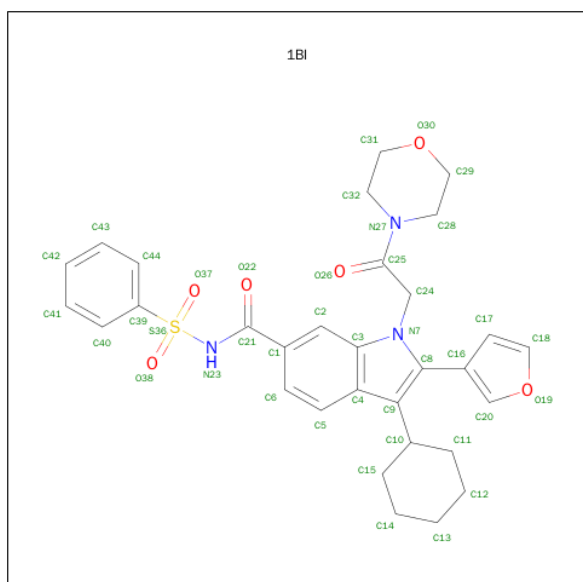
Chain	Residue	Modelled	Actual	Comment	Reference
A	571	HIS	-	EXPRESSION TAG	UNP O92972
A	572	HIS	-	EXPRESSION TAG	UNP O92972
A	573	HIS	-	EXPRESSION TAG	UNP O92972
A	574	HIS	-	EXPRESSION TAG	UNP O92972
A	575	HIS	-	EXPRESSION TAG	UNP O92972
A	576	HIS	-	EXPRESSION TAG	UNP O92972
B	571	HIS	-	EXPRESSION TAG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



R498	T499	W500	R501	H502	R503	A504	R505	S506	V507	R508		L511	L512	S513	Q514	G515	G516	R517	A518	A519	T520	C521	G522	R523	T524	L525	F526	N527	N528	A529	V530	R531	T532	K533	L534	K535	L536	T537	P538	I539	P540	A541	A542	S543	Q544	L545	D546	L547	S548	G549	F551	V552		D559	I560	Y561	H562	S563	
LEU	SER	ARG	ALA	ARG	PRO	ARG	HIS	HIS	HIS	HIS	HIS		L511	L512	S513	Q514	G515	G516	R517	A518	A519	T520	C521	G522	R523	T524	L525	F526	N527	N528	A529	V530	R531	T532	K533	L534	K535	L536	T537	P538	I539	P540	A541	A542	S543	Q544	L545	D546	L547	S548	G549	F551	V552		D559	I560	Y561	H562	S563



## 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 0.218 , 0.255	Depositor DCC
$R_{free}$ test set	4127 reflections (9.99%)	DCC
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
Estimated twinning fraction	0.017 for k,h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 46323 reflections	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, 0.375 respectively for untwinned datasets, and 0.333, 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%)	34	63
1	B	535/576 (93%)	457 (85%)	53 (10%)	25 (5%)	3	5
All	All	1090/1152 (95%)	978 (90%)	84 (8%)	28 (3%)	7	16

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%)	44	75
1	B	458/491 (93%)	445 (97%)	13 (3%)	51	81
All	All	935/982 (95%)	906 (97%)	29 (3%)	47	78

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 [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](#)

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	$\# Z  > 2$
2	SO4	A	601	-	4,4,4	0.24	0	6,6,6	0.06	0
2	SO4	A	602	-	4,4,4	0.18	0	6,6,6	0.18	0
3	1BI	B	601	-	40,46,46	3.63	24 (60%)	49,66,66	2.27	8 (16%)
2	SO4	B	602	-	4,4,4	0.24	0	6,6,6	0.06	0
2	SO4	B	603	-	4,4,4	0.25	0	6,6,6	0.11	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
2	SO4	A	601	-	-	0/0/0/0	0/0/0/0
2	SO4	A	602	-	-	0/0/0/0	0/0/0/0
3	1BI	B	601	-	-	0/23/47/47	0/5/6/6
2	SO4	B	602	-	-	0/0/0/0	0/0/0/0
2	SO4	B	603	-	-	0/0/0/0	0/0/0/0

The worst 5 of 24 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	601	1BI	O37-S36	-6.20	1.36	1.43
3	B	601	1BI	O38-S36	-5.42	1.37	1.43
3	B	601	1BI	S36-N23	-4.78	1.54	1.64
3	B	601	1BI	C41-C40	2.18	1.43	1.38
3	B	601	1BI	C15-C10	2.27	1.59	1.53

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	-8.34	106.51	116.05
3	B	601	1BI	O37-S36-O38	-6.98	110.28	119.54
3	B	601	1BI	C21-N23-S36	-4.26	118.34	123.33
3	B	601	1BI	C24-C25-N27	2.35	120.45	117.02
3	B	601	1BI	C28-N27-C32	3.34	118.75	112.56

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
2	A	601	SO4	1	0
3	B	601	1BI	2	0

## 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%)	-0.07	5 (0%) 85 86	23, 38, 63, 91	0
1	B	541/576 (93%)	0.64	75 (13%) 4 3	22, 58, 109, 119	0
All	All	1100/1152 (95%)	0.28	80 (7%) 18 16	22, 43, 101, 119	0

The worst 5 of 80 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	511	LEU	7.3
1	B	524	TYR	6.7
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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	SO4	B	603	5/5	0.88	0.31	6.06	103,103,104,104	0
2	SO4	A	601	5/5	0.92	0.21	1.53	95,95,95,96	0
2	SO4	A	602	5/5	0.95	0.20	1.33	74,75,75,76	0
3	1BI	B	601	41/41	0.87	0.21	1.14	64,73,83,85	0
2	SO4	B	602	5/5	0.92	0.21	0.96	104,104,105,105	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.