



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 10:21 pm BST

PDB ID : 4CL1  
Title : The crystal structure of NS5A domain 1 from genotype 1a reveals new clues to the mechanism of action for dimeric HCV inhibitors  
Authors : Lambert, S.M.; Langley, D.R.; Garnett, J.A.; Angell, R.; Hedgethorne, K.; Meanwell, N.A.; Matthews, S.J.  
Deposited on : 2014-01-10  
Resolution : 3.50 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
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

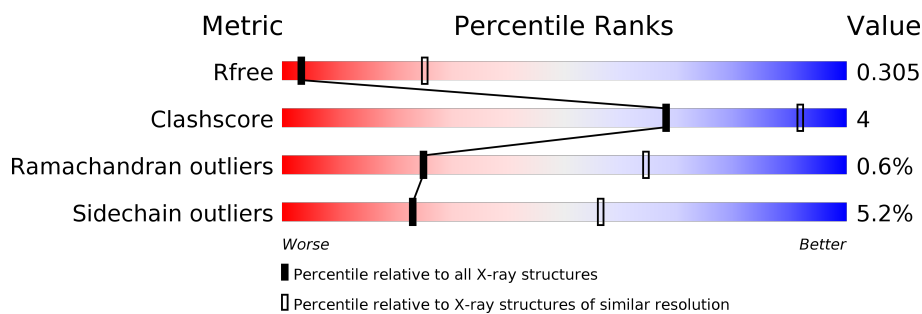
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.50 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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	177	
1	B	177	
1	C	177	
1	D	177	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4462 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 NON-STRUCTURAL PROTEIN 5A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	155	Total	C	N	O	S	0	0	0
			1101	703	199	187	12			
1	B	157	Total	C	N	O	S	0	0	0
			1128	716	206	193	13			
1	C	157	Total	C	N	O	S	0	0	0
			1091	691	195	192	13			
1	D	160	Total	C	N	O	S	0	0	0
			1133	718	202	200	13			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	GLY	-	expression tag	UNP K4KA16
A	3	SER	-	expression tag	UNP K4KA16
A	174	ASP	-	expression tag	UNP K4KA16
A	175	ASP	-	expression tag	UNP K4KA16
A	176	ASP	-	expression tag	UNP K4KA16
A	177	ASP	-	expression tag	UNP K4KA16
A	178	LYS	-	expression tag	UNP K4KA16
B	2	GLY	-	expression tag	UNP K4KA16
B	3	SER	-	expression tag	UNP K4KA16
B	174	ASP	-	expression tag	UNP K4KA16
B	175	ASP	-	expression tag	UNP K4KA16
B	176	ASP	-	expression tag	UNP K4KA16
B	177	ASP	-	expression tag	UNP K4KA16
B	178	LYS	-	expression tag	UNP K4KA16
C	2	GLY	-	expression tag	UNP K4KA16
C	3	SER	-	expression tag	UNP K4KA16
C	174	ASP	-	expression tag	UNP K4KA16
C	175	ASP	-	expression tag	UNP K4KA16
C	176	ASP	-	expression tag	UNP K4KA16
C	177	ASP	-	expression tag	UNP K4KA16
C	178	LYS	-	expression tag	UNP K4KA16

*Continued on next page...*

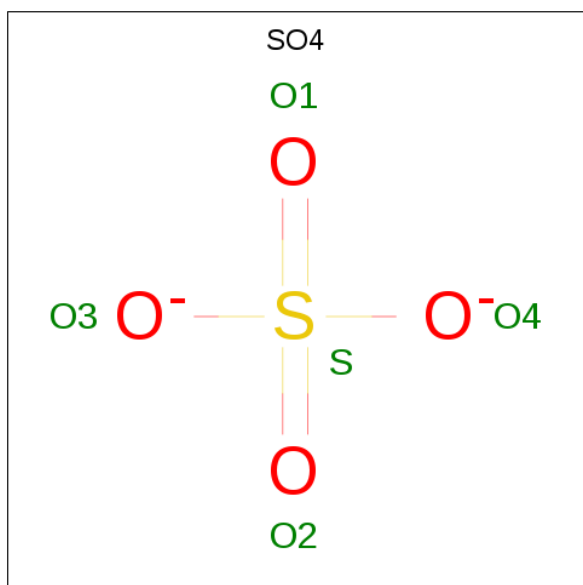
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	2	GLY	-	expression tag	UNP K4KA16
D	3	SER	-	expression tag	UNP K4KA16
D	174	ASP	-	expression tag	UNP K4KA16
D	175	ASP	-	expression tag	UNP K4KA16
D	176	ASP	-	expression tag	UNP K4KA16
D	177	ASP	-	expression tag	UNP K4KA16
D	178	LYS	-	expression tag	UNP K4KA16

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Zn	0	0
			1	1		
2	A	1	Total	Zn	0	0
			1	1		
2	D	1	Total	Zn	0	0
			1	1		
2	C	1	Total	Zn	0	0
			1	1		

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).

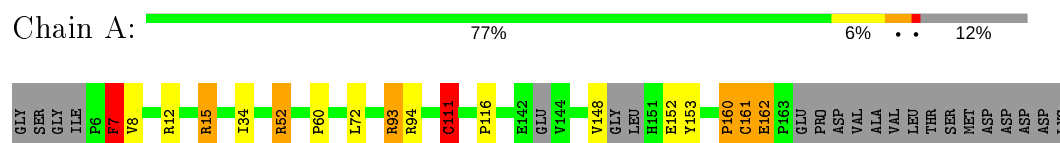


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		

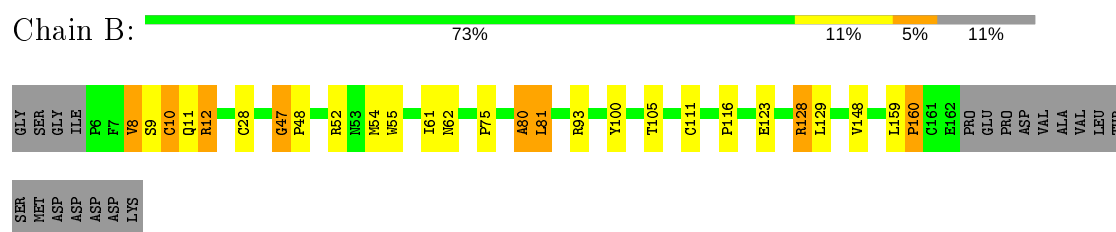
### 3 Residue-property plots [i](#)

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. 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. 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.

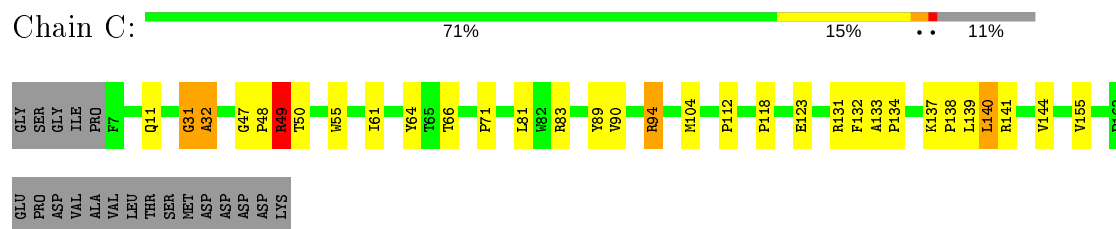
#### • Molecule 1: NON-STRUCTURAL PROTEIN 5A



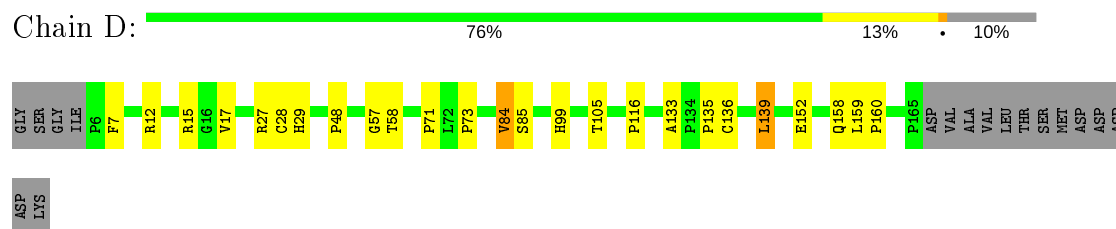
#### • Molecule 1: NON-STRUCTURAL PROTEIN 5A



#### • Molecule 1: NON-STRUCTURAL PROTEIN 5A



#### • Molecule 1: NON-STRUCTURAL PROTEIN 5A



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	100.23Å 101.76Å 149.79Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.12 – 3.50 50.11 – 3.50	Depositor EDS
% Data completeness (in resolution range)	86.0 (50.12-3.50) 86.1 (50.11-3.50)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.51 (at 3.48Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.225 , 0.266 0.271 , 0.305	Depositor DCC
$R_{free}$ test set	866 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	103.9	Xtriage
Anisotropy	0.130	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 108.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.015 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.86	EDS
Total number of atoms	4462	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	90.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.56% 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: ZN, 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	0.68	3/1137 (0.3%)	1.06	10/1561 (0.6%)
1	B	0.71	4/1166 (0.3%)	1.17	12/1599 (0.8%)
1	C	0.60	3/1128 (0.3%)	1.20	14/1553 (0.9%)
1	D	0.66	2/1171 (0.2%)	1.05	10/1607 (0.6%)
All	All	0.67	12/4602 (0.3%)	1.12	46/6320 (0.7%)

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	8
1	B	0	6
1	C	0	4
1	D	1	2
All	All	1	20

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	116	PRO	N-CD	10.43	1.62	1.47
1	B	116	PRO	N-CD	10.25	1.62	1.47
1	B	80	ALA	C-N	8.26	1.53	1.34
1	A	161	CYS	C-N	-7.97	1.15	1.34
1	C	112	PRO	N-CD	7.78	1.58	1.47

The worst 5 of 46 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	137	LYS	C-N-CD	-27.06	61.06	120.60
1	D	84	VAL	C-N-CA	18.53	168.03	121.70
1	B	81	LEU	O-C-N	-18.07	93.80	122.70
1	D	158	GLN	CB-CA-C	13.95	138.29	110.40
1	B	80	ALA	O-C-N	-12.57	102.59	122.70

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	D	84	VAL	CA

5 of 20 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	12	ARG	Sidechain
1	A	15	ARG	Sidechain
1	A	52	ARG	Sidechain
1	A	7	PHE	Peptide
1	A	93	ARG	Sidechain

## 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	1101	0	940	5	0
1	B	1128	0	970	8	3
1	C	1091	0	898	13	1
1	D	1133	0	963	5	4
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	5	0	0	0	0
All	All	4462	0	3771	30	4

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 4.

The worst 5 of 30 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:C:32:ALA:HB1	1:C:48:PRO:HG2	1.58	0.85
1:B:9:SER:CB	1:B:11:GLN:HA	2.16	0.75
1:C:49:ARG:HG3	1:C:49:ARG:O	1.85	0.75
1:C:32:ALA:HB1	1:C:48:PRO:CG	2.21	0.70
1:B:10:CYS:HA	1:B:28:CYS:SG	2.32	0.69

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:132:PHE:CE1	1:D:85:SER:O[4_445]	1.45	0.75
1:B:55:TRP:O	1:D:7:PHE:CZ[3_555]	1.60	0.60
1:B:55:TRP:C	1:D:7:PHE:CZ[3_555]	1.92	0.28
1:B:55:TRP:O	1:D:7:PHE:CE2[3_555]	2.09	0.11

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	149/177 (84%)	141 (95%)	7 (5%)	1 (1%)	22	61
1	B	155/177 (88%)	138 (89%)	17 (11%)	0	100	100
1	C	155/177 (88%)	140 (90%)	13 (8%)	2 (1%)	12	48
1	D	158/177 (89%)	143 (90%)	14 (9%)	1 (1%)	25	64
All	All	617/708 (87%)	562 (91%)	51 (8%)	4 (1%)	25	64

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	162	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	138	PRO
1	C	118	PRO
1	D	84	VAL

### 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	95/153 (62%)	91 (96%)	4 (4%)	30	63
1	B	99/153 (65%)	92 (93%)	7 (7%)	14	46
1	C	92/153 (60%)	86 (94%)	6 (6%)	17	50
1	D	101/153 (66%)	98 (97%)	3 (3%)	41	71
All	All	387/612 (63%)	367 (95%)	20 (5%)	23	56

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

Mol	Chain	Res	Type
1	B	105	THR
1	B	111	CYS
1	C	155	VAL
1	B	62	ASN
1	B	100	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	25	HIS

### 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](#)

Of 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 for Mogul analysis.

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	SO4	A	200	-	4,4,4	0.19	0	6,6,6	0.22	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	161:CYS	C	162:GLU	N	1.15

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.