



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 2, 2017 – 02:40 PM EDT

PDB ID : 1P9S  
Title : Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of anti-SARS Drugs  
Authors : Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J.R.; Hilgenfeld, R.  
Deposited on : unknown  
Resolution : 2.54 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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20030345

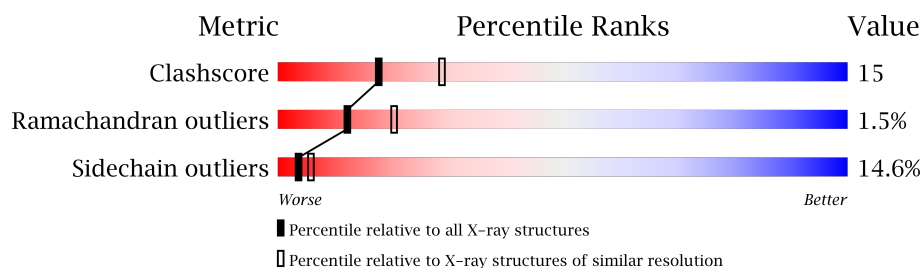
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.54 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	112137	5755 (2.58-2.50)
Ramachandran outliers	110173	5652 (2.58-2.50)
Sidechain outliers	110143	5654 (2.58-2.50)

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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	300	
1	B	300	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4633 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 Replicase polypeptide 1ab.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	300	Total	C	N	O	S	Se	0	0	0
			2299	1461	394	426	8	10			
1	B	299	Total	C	N	O	S	Se	0	0	0
			2290	1457	392	423	8	10			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	6	MSE	MET	MODIFIED RESIDUE	UNP Q05002
A	60	MSE	MET	MODIFIED RESIDUE	UNP Q05002
A	81	MSE	MET	MODIFIED RESIDUE	UNP Q05002
A	95	MSE	MET	MODIFIED RESIDUE	UNP Q05002
A	129	MSE	MET	MODIFIED RESIDUE	UNP Q05002
A	161	MSE	MET	MODIFIED RESIDUE	UNP Q05002
A	180	MSE	MET	MODIFIED RESIDUE	UNP Q05002
A	198	MSE	MET	MODIFIED RESIDUE	UNP Q05002
A	241	MSE	MET	MODIFIED RESIDUE	UNP Q05002
A	296	MSE	MET	MODIFIED RESIDUE	UNP Q05002
B	6	MSE	MET	MODIFIED RESIDUE	UNP Q05002
B	60	MSE	MET	MODIFIED RESIDUE	UNP Q05002
B	81	MSE	MET	MODIFIED RESIDUE	UNP Q05002
B	95	MSE	MET	MODIFIED RESIDUE	UNP Q05002
B	129	MSE	MET	MODIFIED RESIDUE	UNP Q05002
B	161	MSE	MET	MODIFIED RESIDUE	UNP Q05002
B	180	MSE	MET	MODIFIED RESIDUE	UNP Q05002
B	198	MSE	MET	MODIFIED RESIDUE	UNP Q05002
B	241	MSE	MET	MODIFIED RESIDUE	UNP Q05002
B	296	MSE	MET	MODIFIED RESIDUE	UNP Q05002

- Molecule 2 is 1,4-DIETHYLENE DIOXIDE (three-letter code: DIO) (formula: C<sub>4</sub>H<sub>8</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			6	4	2		
2	B	1	Total	C	O	0	0
			6	4	2		

- Molecule 3 is water.

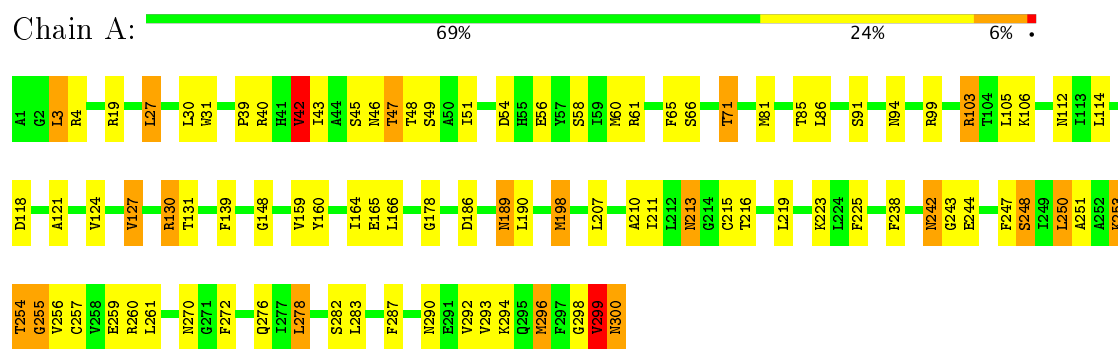
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	16	Total	O	0	0
			16	16		
3	B	16	Total	O	0	0
			16	16		

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

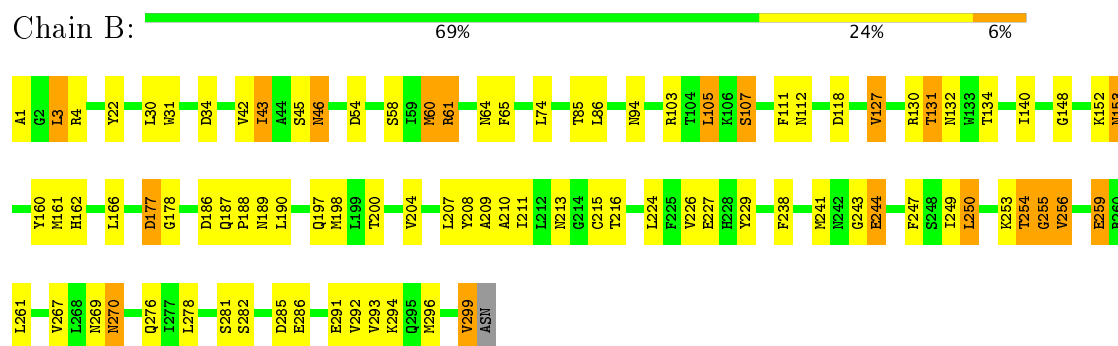
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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.

Note EDS was not executed.

#### • Molecule 1: Replicase polyprotein 1ab



#### • Molecule 1: Replicase polyprotein 1ab



## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	53.36 Å   76.24 Å   73.48 Å 90.00°   103.70°   90.00°	Depositor
Resolution (Å)	70.71 – 2.54	Depositor
% Data completeness (in resolution range)	98.2 (70.71-2.54)	Depositor
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	REFMAC 5.1.24	Depositor
R, $R_{free}$	0.196 , 0.279	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	4633	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

## 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: DIO

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	0.97	1/2341 (0.0%)	0.96	6/3153 (0.2%)
1	B	0.94	2/2332 (0.1%)	1.00	7/3141 (0.2%)
All	All	0.96	3/4673 (0.1%)	0.98	13/6294 (0.2%)

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

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	291	GLU	CD-OE2	5.24	1.31	1.25
1	B	60	MSE	SE-CE	-5.11	1.65	1.95
1	A	19	ARG	NE-CZ	5.06	1.39	1.33

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	186	ASP	CB-CG-OD2	10.46	127.71	118.30
1	B	54	ASP	CB-CG-OD2	9.12	126.50	118.30
1	A	186	ASP	CB-CG-OD2	8.42	125.88	118.30
1	B	285	ASP	CB-CG-OD2	8.31	125.78	118.30
1	A	54	ASP	CB-CG-OD2	7.84	125.36	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	299	VAL	Peptide

## 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	2299	0	2227	77	0
1	B	2290	0	2221	68	0
2	A	6	0	8	2	0
2	B	6	0	8	0	0
3	A	16	0	0	2	0
3	B	16	0	0	2	0
All	All	4633	0	4464	139	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 15.

The worst 5 of 139 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:254:THR:O	1:B:256:VAL:N	1.79	1.16
1:A:254:THR:O	1:A:256:VAL:N	1.79	1.16
1:A:299:VAL:CG1	1:A:300:ASN:H	1.60	1.12
1:B:43:ILE:HG21	1:B:60:MSE:HE3	1.43	1.00
1:A:299:VAL:HG13	1:A:300:ASN:H	1.27	0.99

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	298/300 (99%)	278 (93%)	15 (5%)	5 (2%)	11	17
1	B	297/300 (99%)	281 (95%)	12 (4%)	4 (1%)	14	24
All	All	595/600 (99%)	559 (94%)	27 (4%)	9 (2%)	12	20

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	254	THR
1	A	255	GLY
1	A	299	VAL
1	B	244	GLU
1	B	254	THR

### 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	244/235 (104%)	206 (84%)	38 (16%)	3	5
1	B	243/235 (103%)	210 (86%)	33 (14%)	4	7
All	All	487/470 (104%)	416 (85%)	71 (15%)	3	6

5 of 71 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	270	ASN
1	B	30	LEU
1	B	276	GLN
1	A	276	GLN
1	A	299	VAL

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	46	ASN
1	B	63	HIS
1	B	196	ASN
1	A	290	ASN
1	B	189	ASN

### 5.3.3 RNA [i](#)

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](#)

2 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	DIO	A	301	-	6,6,6	0.64	0	6,6,6	0.59	0
2	DIO	B	301	-	6,6,6	0.72	0	6,6,6	0.57	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	DIO	A	301	-	-	0/0/6/6	0/1/1/1
2	DIO	B	301	-	-	0/0/6/6	0/1/1/1

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.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	301	DIO	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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

EDS was not executed - this section is therefore empty.

### 6.4 Ligands ⓘ

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers ⓘ

EDS was not executed - this section is therefore empty.