



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 20, 2023 – 04:30 PM JST

PDB ID : 7CWB  
Title : Ambient-Temperature Serial Femtosecond X-ray Crystal structure of SARS-CoV-2 Main Protease at 1.9 Å Resolution (C121)  
Authors : DeMirici, H.  
Deposited on : 2020-08-27  
Resolution : 1.90 Å (reported)

This is a Full wwPDB X-ray Structure Validation 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
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) : 2.36

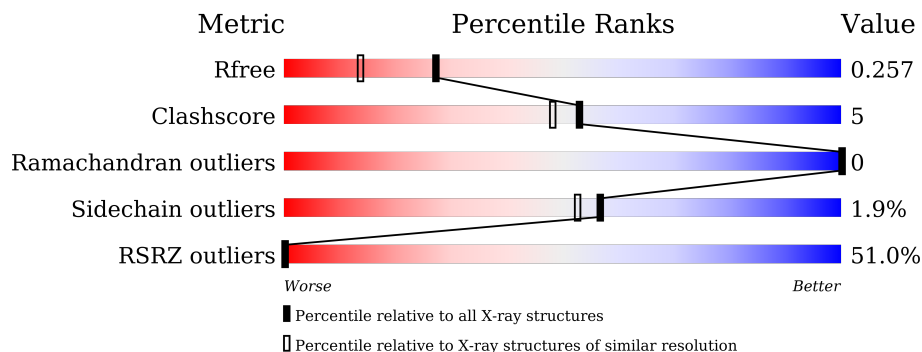
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 of 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	306	

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 2560 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 3C-like proteinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	306	2447	1545	416	462	24	0	10	0

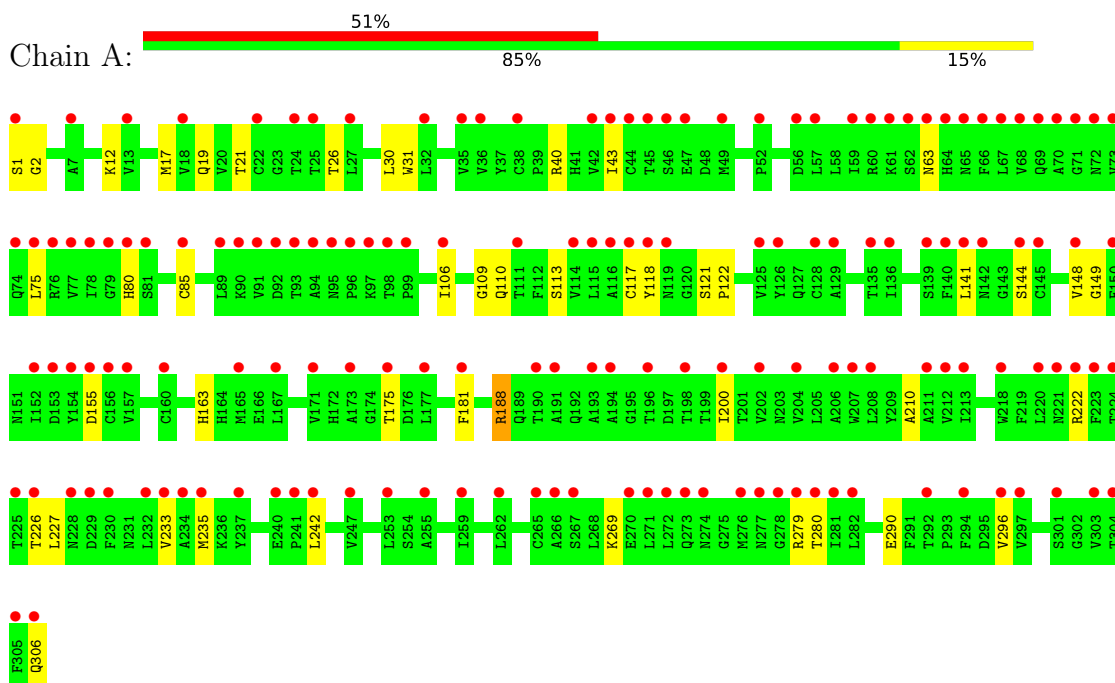
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	110	113	113	0	3

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

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

- Molecule 1: 3C-like proteinase



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	114.00Å 53.50Å 45.00Å 90.00° 102.00° 90.00°	Depositor
Resolution (Å)	34.08 – 1.90 34.08 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.9 (34.08-1.90) 100.0 (34.08-1.90)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.06 (at 1.89Å)	Xtrriage
Refinement program	PHENIX phenix-dev-3318	Depositor
R, $R_{free}$	0.218 , 0.257 0.217 , 0.257	Depositor DCC
$R_{free}$ test set	2000 reflections (9.50%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	35.4	Xtrriage
Anisotropy	0.162	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 69.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.39$ , $\langle L^2 \rangle = 0.23$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2560	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.76% 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](#)

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	0.42	1/2500 (0.0%)	0.58	0/3395

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	106	ILE	C-N	-5.79	1.20	1.34

There are no bond angle outliers.

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	2447	0	2381	26	0
2	A	113	0	0	4	2
All	All	2560	0	2381	26	2

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 5.

All (26) 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:12:LYS:NZ	1:A:306:GLN:O	2.23	0.71

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:290:GLU:OE1	2:A:401:HOH:O	2.14	0.64
1:A:279:ARG:HE	1:A:280:THR:H	1.46	0.62
1:A:109:GLY:HA2	1:A:200:ILE:HD13	1.88	0.56
1:A:233:VAL:HG21	1:A:269:LYS:HE3	1.90	0.54
1:A:21:THR:OG1	1:A:26:THR:OG1	2.21	0.53
1:A:31:TRP:CE2	1:A:75:LEU:HD21	2.44	0.52
1:A:1:SER:OG	1:A:2:GLY:N	2.43	0.52
1:A:118:TYR:CE2	1:A:144:SER:HB3	2.46	0.51
1:A:163:HIS:HE1	2:A:483:HOH:O	1.93	0.51
1:A:19:GLN:HE21	1:A:26:THR:HG21	1.76	0.51
1:A:227:LEU:HD11	1:A:242:LEU:HB3	1.92	0.51
1:A:63:ASN:OD1	1:A:80:HIS:ND1	2.41	0.49
1:A:17:MET:HG3	1:A:117:CYS:SG	2.54	0.47
1:A:118:TYR:CE2	1:A:141:LEU:HB2	2.51	0.46
1:A:175:THR:HG22	1:A:181:PHE:HA	1.97	0.46
1:A:279:ARG:HE	1:A:280:THR:N	2.13	0.46
1:A:113:SER:O	1:A:149:GLY:HA2	2.16	0.45
1:A:188:ARG:NH1	2:A:406:HOH:O	2.39	0.44
1:A:40:ARG:HD3	1:A:85:CYS:HA	2.00	0.43
1:A:30:LEU:HD22	1:A:148:VAL:HG11	2.00	0.42
1:A:110:GLN:NE2	2:A:413:HOH:O	2.51	0.42
1:A:117:CYS:O	1:A:144:SER:HA	2.19	0.42
1:A:210:ALA:HB2	1:A:296:VAL:HG13	2.02	0.42
1:A:40:ARG:O	1:A:43:ILE:HG12	2.20	0.41
1:A:121:SER:HA	1:A:122:PRO:HD3	1.87	0.41

All (2) 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)
2:A:479:HOH:O	2:A:491:HOH:O[4_546]	2.03	0.17
2:A:484:HOH:O	2:A:497:HOH:O[2_555]	2.14	0.06

## 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	314/306 (103%)	308 (98%)	6 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	273/263 (104%)	267 (98%)	6 (2%)	52	47

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	155	ASP
1	A	188	ARG
1	A	222[A]	ARG
1	A	222[B]	ARG
1	A	226	THR
1	A	235	MET

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

Mol	Chain	Res	Type
1	A	19	GLN

### 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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 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

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	306/306 (100%)	2.32	156 (50%) <b>0</b> <b>0</b>	21, 39, 70, 113	0

All (156) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	72	ASN	10.7
1	A	306	GLN	9.8
1	A	222[A]	ARG	8.6
1	A	154	TYR	8.2
1	A	153	ASP	8.2
1	A	277	ASN	7.0
1	A	78	ILE	6.7
1	A	79	GLY	6.4
1	A	73	VAL	6.2
1	A	24	THR	5.8
1	A	303	VAL	5.8
1	A	223	PHE	5.3
1	A	62	SER	5.1
1	A	75	LEU	5.1
1	A	74	GLN	5.1
1	A	77	VAL	5.0
1	A	226	THR	4.9
1	A	221	ASN	4.7
1	A	91	VAL	4.7
1	A	232	LEU	4.5
1	A	93	THR	4.4
1	A	46[A]	SER	4.4
1	A	42	VAL	4.2
1	A	64	HIS	4.2
1	A	68	VAL	4.1
1	A	233	VAL	4.1
1	A	262[A]	LEU	4.1

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	70	ALA	4.0
1	A	155	ASP	4.0
1	A	125	VAL	4.0
1	A	274	ASN	4.0
1	A	224	THR	4.0
1	A	213	ILE	3.9
1	A	220	LEU	3.9
1	A	65[A]	ASN	3.9
1	A	60	ARG	3.8
1	A	80	HIS	3.8
1	A	279	ARG	3.7
1	A	305	PHE	3.7
1	A	71	GLY	3.6
1	A	234	ALA	3.6
1	A	141	LEU	3.6
1	A	67	LEU	3.5
1	A	167	LEU	3.5
1	A	235	MET	3.5
1	A	272	LEU	3.4
1	A	193	ALA	3.4
1	A	247	VAL	3.4
1	A	90	LYS	3.4
1	A	98	THR	3.4
1	A	92	ASP	3.4
1	A	229	ASP	3.3
1	A	25	THR	3.3
1	A	117	CYS	3.3
1	A	115	LEU	3.2
1	A	99	PRO	3.2
1	A	135[A]	THR	3.2
1	A	255	ALA	3.2
1	A	278	GLY	3.2
1	A	259	ILE	3.2
1	A	22	CYS	3.2
1	A	230	PHE	3.2
1	A	271	LEU	3.1
1	A	44	CYS	3.1
1	A	63	ASN	3.1
1	A	212	VAL	3.1
1	A	304	THR	3.1
1	A	59	ILE	3.0
1	A	265	CYS	3.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	114	VAL	3.0
1	A	211	ALA	3.0
1	A	136	ILE	3.0
1	A	27	LEU	2.9
1	A	171	VAL	2.9
1	A	76	ARG	2.9
1	A	85	CYS	2.9
1	A	47[A]	GLU	2.8
1	A	194	ALA	2.8
1	A	45	THR	2.8
1	A	270	GLU	2.8
1	A	116	ALA	2.8
1	A	198	THR	2.8
1	A	165[A]	MET	2.8
1	A	94	ALA	2.8
1	A	13	VAL	2.8
1	A	35	VAL	2.8
1	A	267	SER	2.7
1	A	191	ALA	2.7
1	A	118	TYR	2.7
1	A	296	VAL	2.7
1	A	160	CYS	2.7
1	A	152	ILE	2.7
1	A	7	ALA	2.6
1	A	280	THR	2.6
1	A	140	PHE	2.6
1	A	294	PHE	2.6
1	A	126	TYR	2.6
1	A	200	ILE	2.6
1	A	148	VAL	2.6
1	A	276	MET	2.5
1	A	32	LEU	2.5
1	A	282	LEU	2.5
1	A	119	ASN	2.5
1	A	266	ALA	2.5
1	A	38	CYS	2.5
1	A	228	ASN	2.5
1	A	301[A]	SER	2.5
1	A	297	VAL	2.5
1	A	96	PRO	2.4
1	A	69	GLN	2.4
1	A	57	LEU	2.4

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	156	CYS	2.4
1	A	1	SER	2.4
1	A	97	LYS	2.4
1	A	190	THR	2.4
1	A	129	ALA	2.4
1	A	49	MET	2.3
1	A	218	TRP	2.3
1	A	150	PHE	2.3
1	A	43	ILE	2.3
1	A	177	LEU	2.3
1	A	225	THR	2.3
1	A	242	LEU	2.3
1	A	66	PHE	2.3
1	A	157	VAL	2.3
1	A	128	CYS	2.3
1	A	175	THR	2.3
1	A	208	LEU	2.3
1	A	142	ASN	2.3
1	A	207	TRP	2.3
1	A	81	SER	2.3
1	A	273	GLN	2.2
1	A	36	VAL	2.2
1	A	61	LYS	2.2
1	A	144	SER	2.2
1	A	181	PHE	2.2
1	A	18	VAL	2.2
1	A	106	ILE	2.2
1	A	95	ASN	2.2
1	A	52	PRO	2.2
1	A	111	THR	2.2
1	A	139	SER	2.2
1	A	241	PRO	2.1
1	A	202	VAL	2.1
1	A	173	ALA	2.1
1	A	237	TYR	2.1
1	A	196	THR	2.1
1	A	292	THR	2.1
1	A	145	CYS	2.1
1	A	253	LEU	2.1
1	A	56	ASP	2.1
1	A	204	VAL	2.0
1	A	206	ALA	2.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	240	GLU	2.0
1	A	89	LEU	2.0
1	A	281	ILE	2.0

## 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 monosaccharides in this entry.

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.