



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 21, 2023 – 02:29 PM JST

PDB ID : 7VK2  
Title : Crystal Structure of SARS-CoV-2 Mpro at 2.0 Å resolution -9  
Authors : DeMirci, H.; Gul, M.  
Deposited on : 2021-09-29  
Resolution : 2.00 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

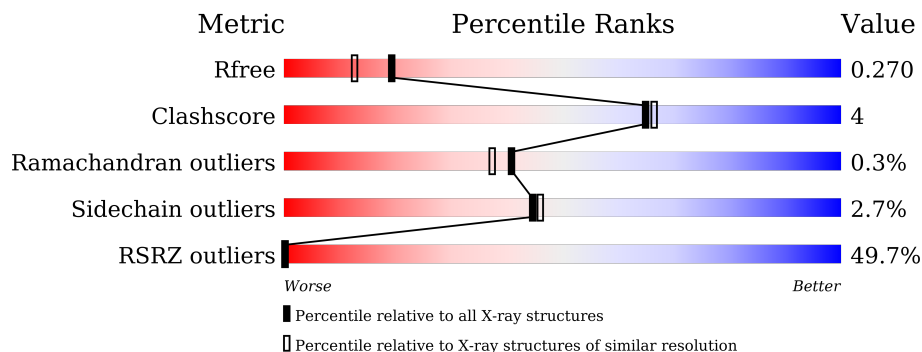
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 2487 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	7	10	0

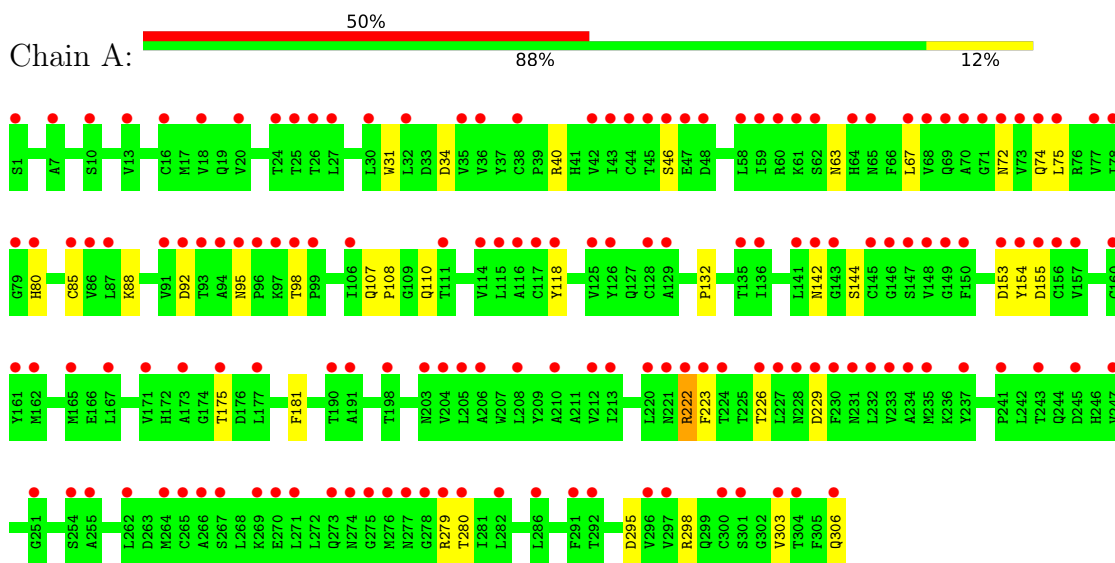
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	40	40	40	0	0

### 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, $\alpha$ , $\beta$ , $\gamma$	115.00Å 54.80Å 45.40Å 90.00° 101.40° 90.00°	Depositor
Resolution (Å)	31.99 – 2.00 34.55 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.9 (31.99-2.00) 100.0 (34.55-2.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.04 (at 2.00Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, $R_{free}$	0.229 , 0.275 0.227 , 0.270	Depositor DCC
$R_{free}$ test set	1887 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	47.2	Xtrriage
Anisotropy	0.101	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 60.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2487	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.58% 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	0.28	0/2500	0.53	0/3395

There are no bond length outliers.

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	2382	20	0
2	A	40	0	0	1	0
All	All	2487	0	2382	20	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 4.

All (20) 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:226:THR:OG1	2:A:402:HOH:O	2.13	0.65
1:A:155:ASP:HA	1:A:306:GLN:HB2	1.85	0.58
1:A:40:ARG:HD3	1:A:85:CYS:HA	1.86	0.57
1:A:279:ARG:HE	1:A:280:THR:H	1.58	0.51

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:175:THR:HG22	1:A:181:PHE:HA	1.93	0.50
1:A:222[A]:ARG:O	1:A:222[A]:ARG:HG3	2.12	0.49
1:A:226:THR:HG23	1:A:229:ASP:H	1.78	0.48
1:A:118:TYR:CE2	1:A:144:SER:HB3	2.48	0.47
1:A:279:ARG:HA	1:A:279:ARG:NE	2.29	0.47
1:A:34:ASP:OD1	1:A:34:ASP:N	2.44	0.47
1:A:107:GLN:O	1:A:110:GLN:HG3	2.15	0.47
1:A:95:ASN:HB3	1:A:98:THR:OG1	2.15	0.46
1:A:67:LEU:HD12	1:A:74:GLN:HE21	1.82	0.45
1:A:92:ASP:OD1	1:A:92:ASP:N	2.49	0.45
1:A:31:TRP:CE2	1:A:75:LEU:HD21	2.51	0.44
1:A:63:ASN:OD1	1:A:80:HIS:ND1	2.48	0.44
1:A:295:ASP:OD1	1:A:298:ARG:NH2	2.52	0.42
1:A:222[B]:ARG:H	1:A:222[B]:ARG:HG3	1.66	0.42
1:A:223:PHE:CD1	1:A:223:PHE:N	2.88	0.41
1:A:108:PRO:HB3	1:A:132:PRO:HA	2.03	0.40

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	314/306 (103%)	306 (98%)	7 (2%)	1 (0%)	41 37

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	154	TYR

### 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%)	264 (97%)	9 (3%)	38 37

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	46[A]	SER
1	A	46[B]	SER
1	A	72	ASN
1	A	88	LYS
1	A	142	ASN
1	A	153	ASP
1	A	222[A]	ARG
1	A	222[B]	ARG
1	A	303	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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%)	2.16	152 (49%) <b>0</b> <b>0</b>	30, 47, 71, 84	0

All (152) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	72	ASN	8.2
1	A	222[A]	ARG	7.8
1	A	79	GLY	6.7
1	A	154	TYR	6.6
1	A	306	GLN	6.4
1	A	24	THR	6.2
1	A	78	ILE	6.1
1	A	155	ASP	5.5
1	A	74	GLN	5.2
1	A	277	ASN	5.1
1	A	77	VAL	4.8
1	A	153	ASP	4.6
1	A	221	ASN	4.6
1	A	226	THR	4.6
1	A	93	THR	4.4
1	A	75	LEU	4.4
1	A	262[A]	LEU	4.4
1	A	303	VAL	4.3
1	A	91	VAL	4.2
1	A	64	HIS	4.1
1	A	234	ALA	4.1
1	A	60	ARG	3.9
1	A	73	VAL	3.9
1	A	223	PHE	3.9
1	A	232	LEU	3.8
1	A	279	ARG	3.8
1	A	274	ASN	3.7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	42	VAL	3.7
1	A	62	SER	3.6
1	A	27	LEU	3.5
1	A	125	VAL	3.5
1	A	233	VAL	3.4
1	A	97	LYS	3.4
1	A	92	ASP	3.4
1	A	300	CYS	3.4
1	A	65[A]	ASN	3.4
1	A	301[A]	SER	3.3
1	A	266	ALA	3.3
1	A	115	LEU	3.3
1	A	45	THR	3.3
1	A	235	MET	3.3
1	A	229	ASP	3.3
1	A	114	VAL	3.3
1	A	230	PHE	3.3
1	A	38	CYS	3.2
1	A	255	ALA	3.2
1	A	220	LEU	3.2
1	A	247	VAL	3.1
1	A	296	VAL	3.1
1	A	265	CYS	3.1
1	A	171	VAL	3.1
1	A	117	CYS	3.1
1	A	156	CYS	3.1
1	A	58	LEU	3.0
1	A	280	THR	3.0
1	A	271	LEU	3.0
1	A	46[A]	SER	3.0
1	A	7	ALA	3.0
1	A	273	GLN	3.0
1	A	96	PRO	3.0
1	A	135[A]	THR	2.9
1	A	269	LYS	2.9
1	A	94	ALA	2.9
1	A	26	THR	2.9
1	A	126	TYR	2.9
1	A	304	THR	2.9
1	A	80	HIS	2.8
1	A	141	LEU	2.8
1	A	136	ILE	2.8

*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	43	ILE	2.8
1	A	267	SER	2.8
1	A	116	ALA	2.7
1	A	148	VAL	2.7
1	A	1	SER	2.7
1	A	227	LEU	2.7
1	A	70	ALA	2.7
1	A	191	ALA	2.7
1	A	67	LEU	2.7
1	A	13	VAL	2.7
1	A	147	SER	2.7
1	A	228	ASN	2.7
1	A	241	PRO	2.6
1	A	106	ILE	2.6
1	A	145	CYS	2.6
1	A	32	LEU	2.6
1	A	118	TYR	2.6
1	A	270	GLU	2.6
1	A	173	ALA	2.6
1	A	157	VAL	2.5
1	A	165[A]	MET	2.5
1	A	16	CYS	2.5
1	A	98	THR	2.5
1	A	99	PRO	2.5
1	A	213	ILE	2.5
1	A	160	CYS	2.5
1	A	276	MET	2.5
1	A	177	LEU	2.4
1	A	44	CYS	2.4
1	A	161	TYR	2.4
1	A	210	ALA	2.4
1	A	167	LEU	2.4
1	A	297	VAL	2.4
1	A	71	GLY	2.4
1	A	59	ILE	2.4
1	A	128	CYS	2.4
1	A	243	THR	2.4
1	A	212	VAL	2.4
1	A	208	LEU	2.4
1	A	275	GLY	2.3
1	A	204	VAL	2.3
1	A	278	GLY	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	224	THR	2.3
1	A	25	THR	2.3
1	A	86	VAL	2.3
1	A	198	THR	2.3
1	A	85	CYS	2.3
1	A	142	ASN	2.3
1	A	87	LEU	2.3
1	A	231	ASN	2.3
1	A	162	MET	2.2
1	A	61	LYS	2.2
1	A	205	LEU	2.2
1	A	95	ASN	2.2
1	A	292	THR	2.2
1	A	20	VAL	2.2
1	A	150	PHE	2.2
1	A	203	ASN	2.2
1	A	254	SER	2.2
1	A	282	LEU	2.2
1	A	69	GLN	2.2
1	A	48	ASP	2.2
1	A	146	GLY	2.2
1	A	175	THR	2.2
1	A	30	LEU	2.1
1	A	206	ALA	2.1
1	A	47[A]	GLU	2.1
1	A	149	GLY	2.1
1	A	264	MET	2.1
1	A	251	GLY	2.1
1	A	129	ALA	2.1
1	A	190	THR	2.1
1	A	35	VAL	2.1
1	A	36	VAL	2.1
1	A	291	PHE	2.1
1	A	237	TYR	2.1
1	A	68	VAL	2.1
1	A	245	ASP	2.0
1	A	286	LEU	2.0
1	A	111	THR	2.0
1	A	18	VAL	2.0
1	A	10	SER	2.0
1	A	143	GLY	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.