



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 5, 2026 – 04:58 PM UTC

PDB ID : 7C2Y / pdb_00007c2y
Title : The crystal structure of COVID-2019 main protease in the apo state
Authors : Zhou, X.L.; Zhong, F.L.; Lin, C.; Zhou, H.; Hu, X.H.; Wang, Q.S.; Li, J.; Zhang, J.
Deposited on : 2020-05-10
Resolution : 1.91 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

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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

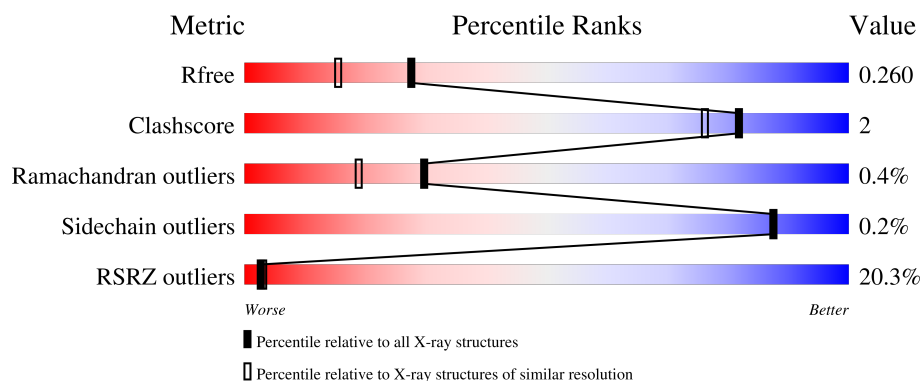
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.91 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}	180053	1188 (1.92-1.92)
Clashscore	190562	1209 (1.92-1.92)
Ramachandran outliers	187476	1195 (1.92-1.92)
Sidechain outliers	187428	1195 (1.92-1.92)
RSRZ outliers	180081	1188 (1.92-1.92)

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 ≥ 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	298	14% 90% 7% .
1	B	298	25% 90% 5% 5%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4383 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
1	A	289	Total	C	N	O	S	0	0	0
			2153	1361	366	405	21			
1	B	283	Total	C	N	O	S	0	0	0
			2045	1294	351	382	18			

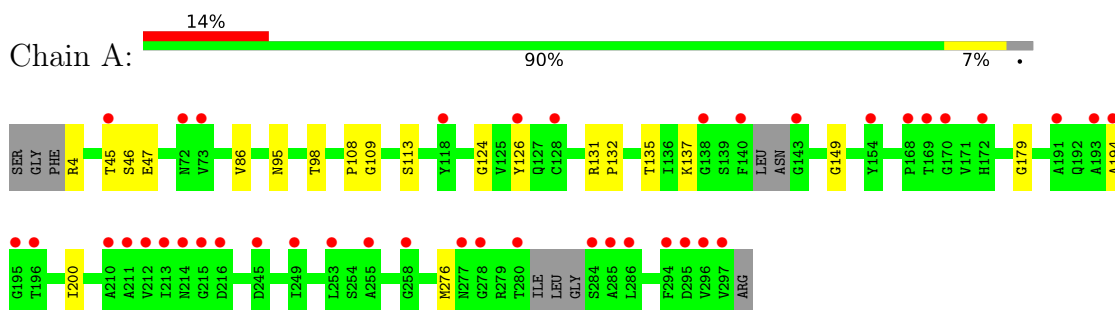
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	107	Total	O	0	0
			107	107		
2	B	78	Total	O	0	0
			78	78		

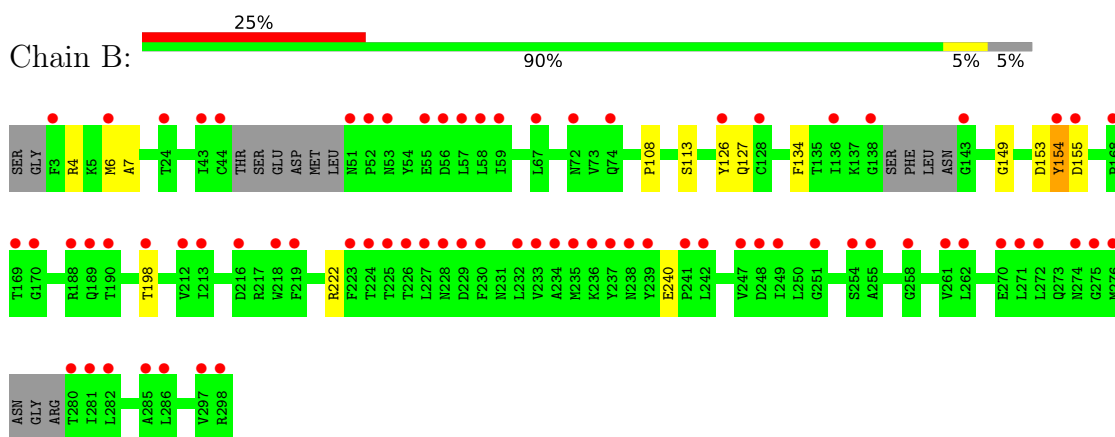
3 Residue-property plots [i](#)

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



- Molecule 1: 3C-like proteinase



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	67.88Å 102.34Å 103.49Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	72.77 – 1.91 72.77 – 1.91	Depositor EDS
% Data completeness (in resolution range)	99.8 (72.77-1.91) 92.7 (72.77-1.91)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.55 (at 1.91Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.220 , 0.262 0.221 , 0.260	Depositor DCC
R_{free} test set	1989 reflections (3.52%)	wwPDB-VP
Wilson B-factor (Å ²)	24.7	Xtriage
Anisotropy	0.525	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 44.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.018 for -h,l,k	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	4383	wwPDB-VP
Average B, all atoms (Å ²)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.90% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

5.1 Standard geometry

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	0.35	0/2200	0.55	0/2997
1	B	0.35	0/2087	0.54	0/2847
All	All	0.35	0/4287	0.55	0/5844

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

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	2153	0	2017	11	0
1	B	2045	0	1871	12	0
2	A	107	0	0	0	0
2	B	78	0	0	1	0
All	All	4383	0	3888	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 2.

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:B:154:TYR:O	1:B:155:ASP:OD1	2.12	0.68

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:86:VAL:HG13	1:A:179:GLY:HA2	1.83	0.61
1:A:45:THR:O	1:A:47:GLU:N	2.35	0.59
1:A:124:GLY:HA3	1:B:6:MET:HG2	1.83	0.59
1:B:198:THR:HG22	1:B:240:GLU:HG2	1.85	0.57
1:A:131:ARG:HD3	1:A:137:LYS:HE3	1.89	0.55
1:B:198:THR:CG2	1:B:240:GLU:HG2	2.39	0.52
1:A:4:ARG:HD2	1:B:126:TYR:CD2	2.46	0.50
1:B:222:ARG:NH1	2:B:302:HOH:O	2.31	0.49
1:A:109:GLY:HA2	1:A:200:ILE:HD13	1.96	0.47
1:A:126:TYR:CD1	1:B:4:ARG:HD2	2.52	0.45
1:B:113:SER:O	1:B:149:GLY:HA2	2.17	0.45
1:A:113:SER:O	1:A:149:GLY:HA2	2.17	0.45
1:B:108:PRO:HG3	1:B:134:PHE:CE1	2.53	0.44
1:B:6:MET:O	1:B:127:GLN:CG	2.67	0.43
1:B:7:ALA:HA	1:B:127:GLN:HG2	2.00	0.43
1:A:135:THR:HG21	1:A:194:ALA:HB2	2.01	0.41
1:A:108:PRO:HB3	1:A:132:PRO:HA	2.03	0.41
1:A:95:ASN:HB3	1:A:98:THR:OG1	2.21	0.41
1:B:153:ASP:O	1:B:154:TYR:C	2.64	0.41

There are no symmetry-related clashes.

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	283/298 (95%)	275 (97%)	7 (2%)	1 (0%)	30	19
1	B	275/298 (92%)	269 (98%)	5 (2%)	1 (0%)	30	19
All	All	558/596 (94%)	544 (98%)	12 (2%)	2 (0%)	30	19

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	154	TYR
1	A	46	SER

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	225/256 (88%)	224 (100%)	1 (0%)	84	83
1	B	202/256 (79%)	202 (100%)	0	100	100
All	All	427/512 (83%)	426 (100%)	1 (0%)	87	87

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	276	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	41	HIS
1	A	107	GLN
1	A	256	GLN
1	A	273	GLN
1	B	163	HIS

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 oligosaccharides 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, 95th 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 ²)	Q<0.9
1	A	289/298 (96%)	0.85	41 (14%) 6 8	19, 29, 47, 52	0
1	B	283/298 (94%)	1.23	75 (26%) 1 2	19, 34, 53, 62	0
All	All	572/596 (95%)	1.04	116 (20%) 3 3	19, 31, 50, 62	0

All (116) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	296	VAL	7.2
1	B	280	THR	6.1
1	A	215	GLY	5.2
1	A	191	ALA	4.9
1	B	281	ILE	4.7
1	B	51	ASN	4.6
1	A	253	LEU	4.6
1	A	284	SER	4.3
1	B	232	LEU	4.3
1	B	227	LEU	4.2
1	B	72	ASN	4.1
1	B	285	ALA	4.1
1	A	213	ILE	4.0
1	A	249	ILE	4.0
1	B	286	LEU	4.0
1	B	233	VAL	4.0
1	B	224	THR	3.9
1	B	189	GLN	3.8
1	B	138	GLY	3.8
1	B	255	ALA	3.8
1	B	241	PRO	3.8
1	A	45	THR	3.7
1	A	297	VAL	3.7
1	B	143	GLY	3.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	118	TYR	3.7
1	A	128	CYS	3.7
1	A	138	GLY	3.7
1	B	59	ILE	3.7
1	B	249	ILE	3.7
1	B	237	TYR	3.6
1	B	154	TYR	3.6
1	A	143	GLY	3.6
1	A	280	THR	3.6
1	B	230	PHE	3.5
1	A	210	ALA	3.5
1	A	294	PHE	3.5
1	A	169	THR	3.4
1	B	44	CYS	3.3
1	A	194	ALA	3.3
1	B	262	LEU	3.3
1	B	128	CYS	3.3
1	B	276	MET	3.2
1	A	140	PHE	3.2
1	A	278	GLY	3.2
1	B	234	ALA	3.1
1	B	155	ASP	3.1
1	B	225	THR	3.1
1	B	282	LEU	3.1
1	B	228	ASN	3.0
1	B	198	THR	3.0
1	B	235	MET	3.0
1	B	53	ASN	3.0
1	B	274	ASN	3.0
1	B	168	PRO	3.0
1	B	271	LEU	2.9
1	B	3	PHE	2.9
1	A	286	LEU	2.9
1	B	58	LEU	2.9
1	B	188	ARG	2.9
1	B	298	ARG	2.9
1	B	219	PHE	2.9
1	A	258	GLY	2.9
1	B	169	THR	2.8
1	B	226	THR	2.8
1	B	223	PHE	2.8
1	A	216	ASP	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	218	TRP	2.8
1	A	154	TYR	2.8
1	B	272	LEU	2.8
1	B	52	PRO	2.7
1	A	193	ALA	2.7
1	A	255	ALA	2.7
1	A	196	THR	2.7
1	A	73	VAL	2.7
1	B	242	LEU	2.6
1	B	297	VAL	2.6
1	B	55	GLU	2.6
1	B	213	ILE	2.6
1	A	72	ASN	2.6
1	B	74	GLN	2.5
1	B	238	ASN	2.5
1	A	277	ASN	2.5
1	B	236	LYS	2.5
1	B	56	ASP	2.5
1	A	168	PRO	2.5
1	B	258	GLY	2.4
1	B	43	ILE	2.4
1	B	190	THR	2.4
1	B	229	ASP	2.4
1	A	211	ALA	2.4
1	A	285	ALA	2.4
1	B	170	GLY	2.3
1	B	247	VAL	2.3
1	B	239	TYR	2.3
1	B	136	ILE	2.3
1	A	214	ASN	2.3
1	B	126	TYR	2.2
1	B	216	ASP	2.2
1	B	57	LEU	2.2
1	B	67	LEU	2.2
1	A	295	ASP	2.2
1	B	275	GLY	2.2
1	A	172	HIS	2.2
1	B	212	VAL	2.1
1	B	251	GLY	2.1
1	B	248	ASP	2.1
1	B	24	THR	2.1
1	B	270	GLU	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	170	GLY	2.1
1	B	6	MET	2.1
1	B	261	VAL	2.1
1	B	254	SER	2.0
1	A	126	TYR	2.0
1	A	245	ASP	2.0
1	A	195	GLY	2.0
1	A	212	VAL	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 oligosaccharides 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.