



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 16, 2022 – 12:08 PM EDT

PDB ID : 7MB5  
Title : SARS-CoV-2 Main Protease (Mpro) C145A in Complex with Cleavage Site Nsp5/6 (P6-P1)  
Authors : Lockbaum, G.J.; Schiffer, C.A.  
Deposited on : 2021-03-31  
Resolution : 1.60 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 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 : 2.28.1  
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) : 2.28.1

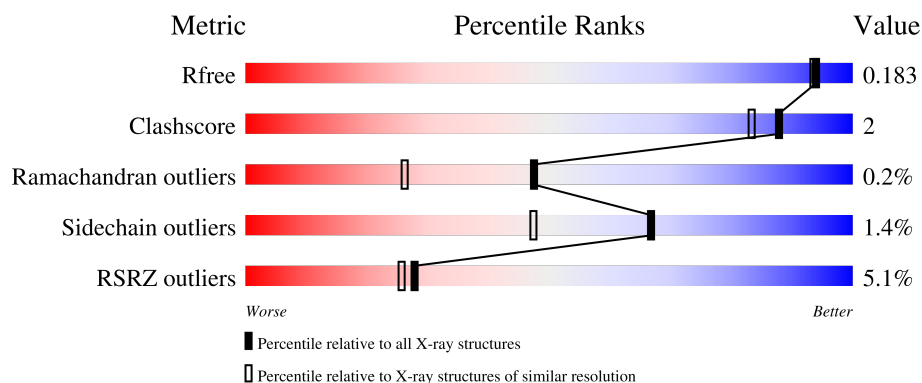
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.60 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	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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	<div> <div>5%</div> <div>92%</div> <div>8%</div> </div>
1	B	306	<div> <div>4%</div> <div>94%</div> <div>6%</div> </div>
2	C	6	<div> <div>33%</div> <div>50%</div> <div>33%</div> <div>17%</div> </div>
2	D	6	<div> <div>17%</div> <div>67%</div> <div>17%</div> <div>17%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9999 atoms, of which 4589 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	B	306	Total	C	H	N	O	S	0	0	0
			4623	1488	2275	397	442	21			
1	A	305	Total	C	H	N	O	S	0	0	0
			4540	1465	2228	393	433	21			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	145	ALA	CYS	engineered mutation	UNP P0DTD1
A	145	ALA	CYS	engineered mutation	UNP P0DTD1

- Molecule 2 is a protein called SER-GLY-VAL-THR-PHE-GLN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	6	Total	C	H	N	O	0	0	0
			88	28	43	7	10			
2	C	6	Total	C	H	N	O	0	0	0
			88	28	43	7	10			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	384	Total	O	0	0
			384	384		
3	A	259	Total	O	0	0
			259	259		
3	D	10	Total	O	0	0
			10	10		
3	C	7	Total	O	0	0
			7	7		

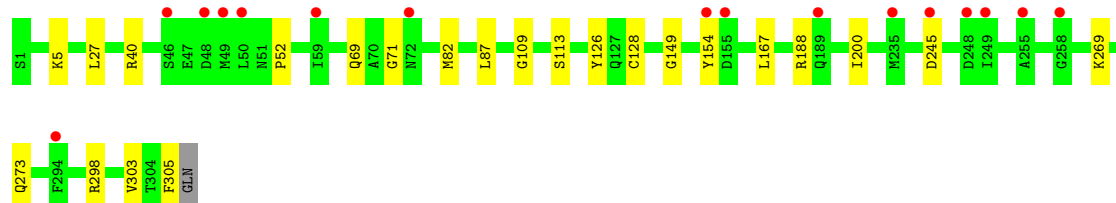
### 3 Residue-property plots [i](#)

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.

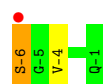
- Molecule 1: 3C-like proteinase



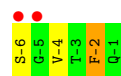
- Molecule 1: 3C-like proteinase



- Molecule 2: SER-GLY-VAL-THR-PHE-GLN



- Molecule 2: SER-GLY-VAL-THR-PHE-GLN



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	67.92Å 100.11Å 103.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.32 – 1.60 29.32 – 1.60	Depositor EDS
% Data completeness (in resolution range)	96.6 (29.32-1.60) 93.3 (29.32-1.60)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.31 (at 1.60Å)	Xtriage
Refinement program	PHENIX 1.19.1_4122	Depositor
R, $R_{free}$	0.164 , 0.184 0.164 , 0.183	Depositor DCC
$R_{free}$ test set	1998 reflections (2.21%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	17.1	Xtriage
Anisotropy	0.316	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.43 , 56.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.000 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	9999	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.72% 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	0.45	0/2364	0.71	0/3219
1	B	0.46	0/2401	0.76	3/3266 (0.1%)
2	C	0.75	0/45	1.21	1/58 (1.7%)
2	D	0.66	0/45	0.86	0/58
All	All	0.46	0/4855	0.74	4/6601 (0.1%)

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
2	C	0	1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	C	-2	PHE	O-C-N	-6.99	111.52	122.70
1	B	187	ASP	CB-CG-OD1	6.37	124.03	118.30
1	B	130	MET	CG-SD-CE	-5.24	91.82	100.20
1	B	130	MET	CA-CB-CG	5.01	121.82	113.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	-2	PHE	Mainchain

## 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	2312	2228	2228	10	0
1	B	2348	2275	2276	9	0
2	C	45	43	43	1	0
2	D	45	43	43	2	0
3	A	259	0	0	0	2
3	B	384	0	0	2	2
3	C	7	0	0	0	0
3	D	10	0	0	0	0
All	All	5410	4589	4590	19	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 2.

All (19) 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:22:CYS:SG	1:B:61:LYS:HE2	2.32	0.69
1:A:40:ARG:HA	1:A:87:LEU:HG	1.81	0.62
1:A:52:PRO:HD2	1:A:188:ARG:HG3	1.85	0.57
1:B:231:ASN:O	1:B:235:MET:HG3	2.07	0.55
1:B:40:ARG:HA	1:B:87:LEU:HG	1.92	0.51
1:B:113:SER:O	1:B:149:GLY:HA2	2.11	0.50
1:B:168:PRO:HB3	2:D:-6:SER:HB3	1.93	0.50
1:A:298:ARG:HG3	1:A:303:VAL:HB	1.95	0.49
1:B:48:ASP:O	1:B:52:PRO:HG3	2.15	0.47
1:A:167:LEU:HD23	2:C:-4:VAL:HG22	1.96	0.47
1:A:298:ARG:HD2	1:A:305:PHE:HZ	1.80	0.46
1:A:69:GLN:HG2	1:A:71:GLY:O	2.16	0.45
1:B:256:GLN:OE1	3:B:401:HOH:O	2.21	0.45
1:A:269:LYS:O	1:A:273:GLN:HG3	2.18	0.44
1:B:167:LEU:HD23	2:D:-4:VAL:HG22	2.00	0.42
1:A:126:TYR:HE1	1:A:128:CYS:HG	1.67	0.42
1:A:113:SER:O	1:A:149:GLY:HA2	2.20	0.42
1:B:24:THR:HG22	3:B:580:HOH:O	2.21	0.41
1:A:109:GLY:HA2	1:A:200:ILE:HD13	2.03	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)
3:B:637:HOH:O	3:A:460:HOH:O[2_454]	1.96	0.24
3:B:674:HOH:O	3:A:623:HOH:O[2_554]	2.11	0.09

## 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	303/306 (99%)	297 (98%)	5 (2%)	1 (0%)	41	21
1	B	304/306 (99%)	299 (98%)	5 (2%)	0	100	100
2	C	4/6 (67%)	4 (100%)	0	0	100	100
2	D	4/6 (67%)	4 (100%)	0	0	100	100
All	All	615/624 (99%)	604 (98%)	10 (2%)	1 (0%)	47	26

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	250/262 (95%)	246 (98%)	4 (2%)	62	41

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	257/262 (98%)	256 (100%)	1 (0%)	91	84
2	C	5/5 (100%)	4 (80%)	1 (20%)	1	0
2	D	5/5 (100%)	4 (80%)	1 (20%)	1	0
All	All	517/534 (97%)	510 (99%)	7 (1%)	67	47

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	27	LEU
1	A	5	LYS
1	A	27	LEU
1	A	82	MET
1	A	245	ASP
2	D	-6	SER
2	C	-6	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	19	GLN
1	A	277	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

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	305/306 (99%)	0.04	16 (5%) 27 24	12, 21, 40, 50	0
1	B	306/306 (100%)	-0.27	13 (4%) 36 33	12, 18, 31, 42	0
2	C	6/6 (100%)	2.51	2 (33%) 0 0	21, 28, 42, 48	0
2	D	6/6 (100%)	1.44	1 (16%) 1 1	15, 20, 34, 52	0
All	All	623/624 (99%)	-0.08	32 (5%) 28 26	12, 19, 39, 52	0

All (32) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	-6	SER	9.7
2	D	-6	SER	8.6
1	A	255	ALA	5.8
1	B	154	TYR	5.2
1	A	59	ILE	5.2
1	A	154	TYR	4.7
1	B	194	ALA	4.4
1	A	258	GLY	4.2
2	C	-5	GLY	4.1
1	A	155	ASP	4.0
1	B	196	THR	3.8
1	B	303	VAL	3.6
1	A	72	ASN	3.5
1	B	24	THR	3.5
1	B	46	SER	3.2
1	A	50	LEU	3.1
1	B	73	VAL	3.1
1	A	245	ASP	2.8
1	B	305	PHE	2.8
1	A	294	PHE	2.7
1	A	249	ILE	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	46	SER	2.5
1	B	304	THR	2.4
1	A	49	MET	2.3
1	B	48	ASP	2.3
1	B	302	GLY	2.2
1	B	45	THR	2.2
1	A	189	GLN	2.2
1	A	235	MET	2.1
1	B	193	ALA	2.0
1	A	248	ASP	2.0
1	A	48	ASP	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.