



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 16, 2022 – 12:14 PM EDT

PDB ID : 7MB8
Title : SARS-CoV-2 Main Protease (Mpro) C145A in Complex with Cleavage Site
Nsp8/9 (P6-P1)
Authors : Lockbaum, G.J.; Schiffer, C.A.
Deposited on : 2021-03-31
Resolution : 1.62 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

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

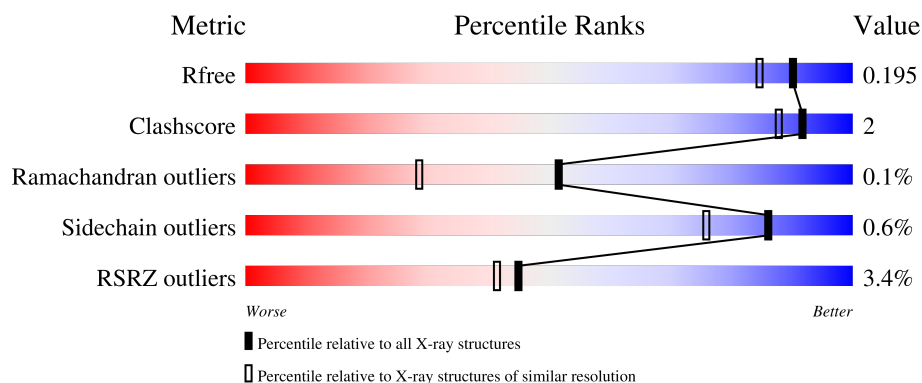
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.62 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	4693 (1.64-1.60)
Clashscore	141614	5002 (1.64-1.60)
Ramachandran outliers	138981	4888 (1.64-1.60)
Sidechain outliers	138945	4887 (1.64-1.60)
RSRZ outliers	127900	4609 (1.64-1.60)

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 ≥ 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> <div></div> <div>94%</div> <div>6%</div> </div> </div>
1	B	306	<div> <div>4%</div> <div> <div></div> <div>93%</div> <div>6%</div> </div> </div>
1	C	306	<div> <div>2%</div> <div> <div></div> <div>96%</div> <div></div> </div> </div>
1	D	306	<div> <div>2%</div> <div> <div></div> <div>93%</div> <div>7%</div> </div> </div>
2	E	6	<div> <div>17%</div> <div> <div></div> <div>100%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	6	<div><div></div><div>100%</div></div>
2	G	6	<div><div></div><div>67%</div><div>33%</div></div>
2	H	6	<div><div></div><div>83%</div><div>17%</div></div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 20551 atoms, of which 9359 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	305	Total	C	H	N	O	S	0	4	0
			4597	1479	2261	399	438	20			
1	C	305	Total	C	H	N	O	S	0	2	0
			4594	1479	2263	394	437	21			
1	B	304	Total	C	H	N	O	S	0	0	0
			4618	1477	2282	399	440	20			
1	D	306	Total	C	H	N	O	S	0	3	0
			4705	1508	2327	404	445	21			

There are 4 discrepancies between the modelled and reference sequences:

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

- Molecule 2 is a protein called SER-ALA-VAL-LYS-LEU-GLN.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	E	6	Total	C	H	N	O		0	2	0
			121	34	67	10	10				
2	G	6	Total	C	H	N	O		0	0	0
			98	28	53	8	9				
2	F	6	Total	C	H	N	O		0	0	0
			98	28	53	8	9				
2	H	6	Total	C	H	N	O		0	0	0
			98	28	53	8	9				

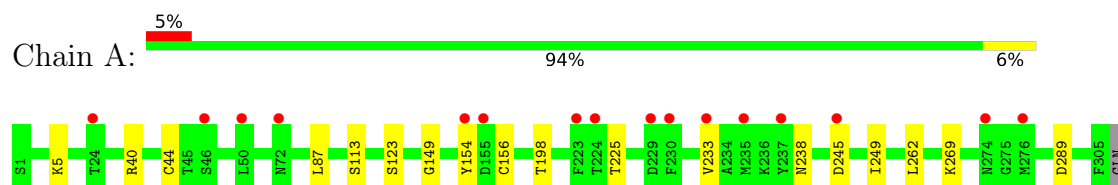
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	312	Total 312	O 312	0	0
3	C	393	Total 393	O 393	0	0
3	B	424	Total 424	O 424	0	0
3	D	463	Total 463	O 463	0	0
3	E	5	Total 5	O 5	0	0
3	G	11	Total 11	O 11	0	0
3	F	6	Total 6	O 6	0	0
3	H	8	Total 8	O 8	0	0

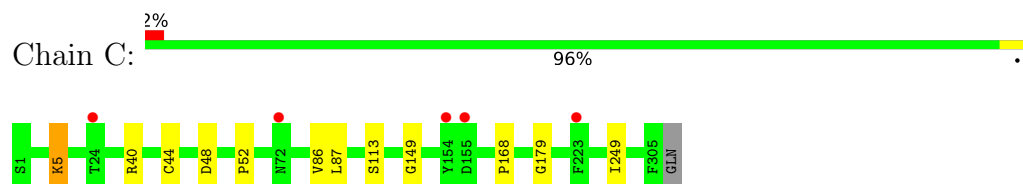
3 Residue-property plots [i](#)

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.

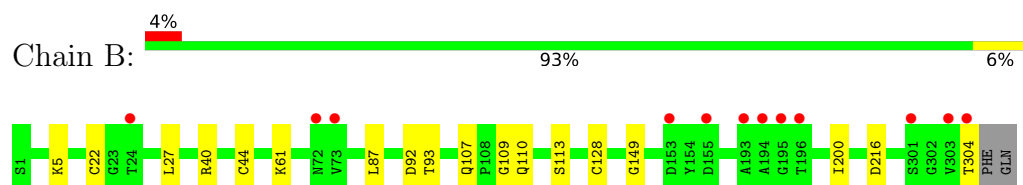
- Molecule 1: 3C-like proteinase



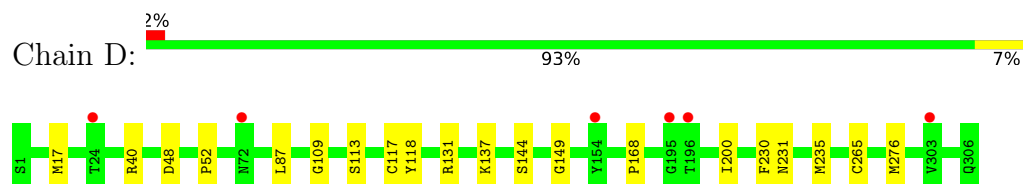
- Molecule 1: 3C-like proteinase



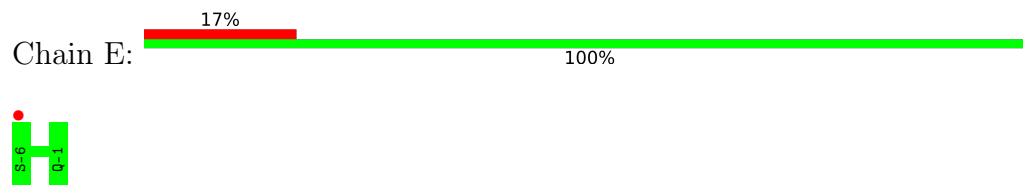
- Molecule 1: 3C-like proteinase



- Molecule 1: 3C-like proteinase



- Molecule 2: SER-ALA-VAL-LYS-LEU-GLN



- Molecule 2: SER-ALA-VAL-LYS-LEU-GLN

Chain G:  67% 33%




• Molecule 2: SER-ALA-VAL-LYS-LEU-GLN

Chain F:  17% 100%



• Molecule 2: SER-ALA-VAL-LYS-LEU-GLN

Chain H:  17% 83% 17%



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	67.81Å 102.97Å 104.29Å 90.00° 101.36° 90.00°	Depositor
Resolution (Å)	29.49 – 1.62 29.49 – 1.62	Depositor EDS
% Data completeness (in resolution range)	98.7 (29.49-1.62) 93.4 (29.49-1.62)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.14 (at 1.62Å)	Xtriage
Refinement program	PHENIX 1.19.1_4122	Depositor
R, R_{free}	0.169 , 0.195 0.169 , 0.195	Depositor DCC
R_{free} test set	2000 reflections (1.15%)	wwPDB-VP
Wilson B-factor (Å ²)	16.0	Xtriage
Anisotropy	0.606	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 47.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	20551	wwPDB-VP
Average B, all atoms (Å ²)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 49.86 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 7.0283e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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 [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	0.61	2/2399 (0.1%)	0.67	1/3266 (0.0%)
1	B	0.66	2/2387 (0.1%)	0.70	0/3246
1	C	0.64	1/2390 (0.0%)	0.69	0/3252
1	D	0.63	1/2438 (0.0%)	0.75	0/3314
2	E	0.70	0/61	0.83	0/78
2	F	0.75	0/44	0.79	0/56
2	G	0.69	0/44	0.89	0/56
2	H	0.79	0/44	0.65	0/56
All	All	0.64	6/9807 (0.1%)	0.71	1/13324 (0.0%)

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	G	0	1

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	44	CYS	CB-SG	-7.70	1.69	1.82
1	B	128	CYS	CB-SG	-7.27	1.69	1.82
1	B	44	CYS	CB-SG	-7.15	1.70	1.82
1	A	44	CYS	CB-SG	-6.81	1.70	1.82
1	A	156	CYS	CB-SG	-6.77	1.70	1.82

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	289	ASP	CB-CG-OD1	5.79	123.51	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	G	-2	LEU	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	2336	2261	2249	7	0
1	B	2336	2282	2283	7	0
1	C	2331	2263	2255	7	0
1	D	2378	2327	2318	11	0
2	E	54	67	55	0	0
2	F	45	53	53	0	0
2	G	45	53	53	1	0
2	H	45	53	53	1	0
3	A	312	0	0	0	1
3	B	424	0	0	0	1
3	C	393	0	0	1	6
3	D	463	0	0	0	5
3	E	5	0	0	0	0
3	F	6	0	0	0	0
3	G	11	0	0	0	0
3	H	8	0	0	0	0
All	All	11192	9359	9319	31	7

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.

The worst 5 of 31 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:HE3	2.45	0.56
1:D:168:PRO:HB3	2:H:-6:SER:HB2	1.87	0.56
1:B:92:ASP:OD1	1:B:93:THR:HG23	2.06	0.56
1:D:231:ASN:HB3	1:D:235:MET:HE1	1.88	0.55
1:B:109:GLY:HA2	1:B:200:ILE:HD13	1.89	0.54

The worst 5 of 7 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:D:435:HOH:O	3:D:543:HOH:O[1_455]	2.00	0.20
3:C:559:HOH:O	3:D:445:HOH:O[2_646]	2.03	0.17
3:C:531:HOH:O	3:D:708:HOH:O[1_455]	2.06	0.14
3:A:693:HOH:O	3:C:721:HOH:O[2_555]	2.13	0.07
3:C:670:HOH:O	3:D:723:HOH:O[1_455]	2.15	0.05

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	307/306 (100%)	303 (99%)	3 (1%)	1 (0%)	41	21
1	B	302/306 (99%)	295 (98%)	7 (2%)	0	100	100
1	C	305/306 (100%)	302 (99%)	3 (1%)	0	100	100
1	D	307/306 (100%)	301 (98%)	6 (2%)	0	100	100
2	E	6/6 (100%)	6 (100%)	0	0	100	100
2	F	4/6 (67%)	4 (100%)	0	0	100	100
2	G	4/6 (67%)	4 (100%)	0	0	100	100
2	H	4/6 (67%)	4 (100%)	0	0	100	100
All	All	1239/1248 (99%)	1219 (98%)	19 (2%)	1 (0%)	51	28

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	154	TYR

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	255/262 (97%)	254 (100%)	1 (0%)	91	84
1	B	258/262 (98%)	255 (99%)	3 (1%)	71	52
1	C	255/262 (97%)	254 (100%)	1 (0%)	91	84
1	D	263/262 (100%)	263 (100%)	0	100	100
2	E	7/5 (140%)	7 (100%)	0	100	100
2	F	5/5 (100%)	5 (100%)	0	100	100
2	G	5/5 (100%)	5 (100%)	0	100	100
2	H	5/5 (100%)	4 (80%)	1 (20%)	1	0
All	All	1053/1068 (99%)	1047 (99%)	6 (1%)	86	76

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	27	LEU
1	B	216	ASP
2	H	-6	SER
1	C	5	LYS
1	A	5	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	273	GLN
1	C	238	ASN
1	D	69	GLN
1	D	74	GLN

5.3.3 RNA ⓘ

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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	305/306 (99%)	-0.04	16 (5%) 27 24	12, 22, 43, 52	0
1	B	304/306 (99%)	-0.31	12 (3%) 39 36	12, 19, 34, 53	0
1	C	305/306 (99%)	-0.36	5 (1%) 72 71	12, 20, 36, 46	0
1	D	306/306 (100%)	-0.41	6 (1%) 65 63	12, 18, 31, 52	0
2	E	6/6 (100%)	-0.30	1 (16%) 1 1	17, 19, 24, 44	0
2	F	6/6 (100%)	-0.02	1 (16%) 1 1	17, 21, 30, 44	0
2	G	6/6 (100%)	-0.24	0 100 100	17, 18, 26, 44	0
2	H	6/6 (100%)	0.71	1 (16%) 1 1	16, 20, 33, 52	0
All	All	1244/1248 (99%)	-0.27	42 (3%) 45 41	12, 20, 39, 53	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	154	TYR	7.2
2	H	-6	SER	6.3
1	B	303	VAL	5.6
1	D	72	ASN	5.0
1	C	154	TYR	4.9

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.