wwPDB X-ray Structure Validation Summary Report (i)

Oct 10, 2023 – 12:12 AM EDT

PDB ID	:	7MB4
Title	:	SARS-CoV-2 Main Protease (Mpro) C145A in Complex with Cleavage Site Nsp4/5 (P6-P1)
Authors	:	Lockbaum, G.J.; Schiffer, C.A.
Deposited on	:	2021-03-31
Resolution	:	1.83 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 (i) 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 (1)) were used in the production of this report:

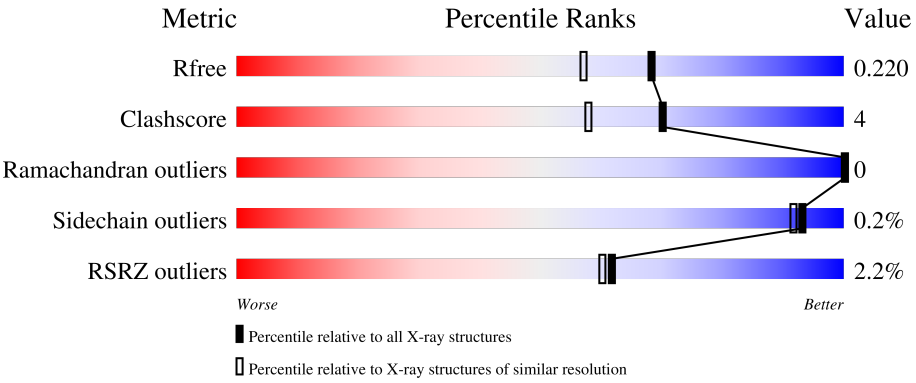
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.83 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.



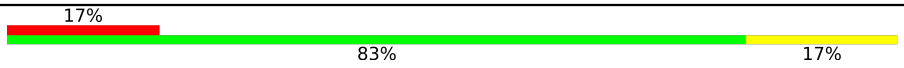
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4003 (1.86-1.82)
Clashscore	141614	4233 (1.86-1.82)
Ramachandran outliers	138981	4185 (1.86-1.82)
Sidechain outliers	138945	4186 (1.86-1.82)
RSRZ outliers	127900	3957 (1.86-1.82)

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 <=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% 90% 10%
1	B	306	3% 89% 10%
1	C	306	% 89% 11%
1	D	306	2% 91% 8%
2	E	6	67% 33%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	6	
2	G	6	
2	H	6	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 19767 atoms, of which 9182 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	0	0
			4509	1460	2207	389	432	21			
1	D	305	Total	C	H	N	O	S	0	0	0
			4489	1456	2195	387	430	21			
1	C	305	Total	C	H	N	O	S	0	4	0
			4655	1495	2294	401	444	21			
1	B	305	Total	C	H	N	O	S	0	2	0
			4656	1494	2298	400	443	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
D	145	ALA	CYS	engineered mutation	UNP P0DTD1
C	145	ALA	CYS	engineered mutation	UNP P0DTD1
B	145	ALA	CYS	engineered mutation	UNP P0DTD1

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	6	Total	C	H	N	O	0	0	0
			90	26	47	7	10			
2	H	6	Total	C	H	N	O	0	0	0
			90	26	47	7	10			
2	G	6	Total	C	H	N	O	0	0	0
			90	26	47	7	10			
2	F	6	Total	C	H	N	O	0	0	0
			90	26	47	7	10			

- Molecule 3 is water.

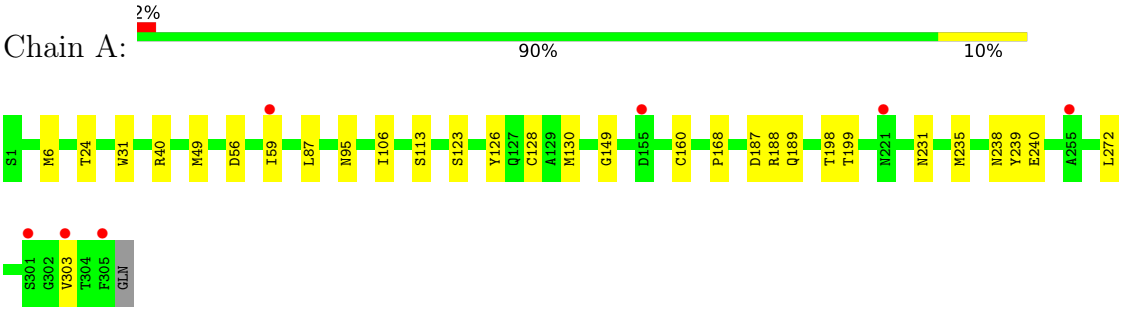
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	218	Total O 218 218	0	0
3	D	262	Total O 262 262	0	0
3	C	296	Total O 296 296	0	0
3	B	301	Total O 301 301	0	0
3	E	6	Total O 6 6	0	0
3	H	6	Total O 6 6	0	0
3	G	5	Total O 5 5	0	0
3	F	4	Total O 4 4	0	0

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.

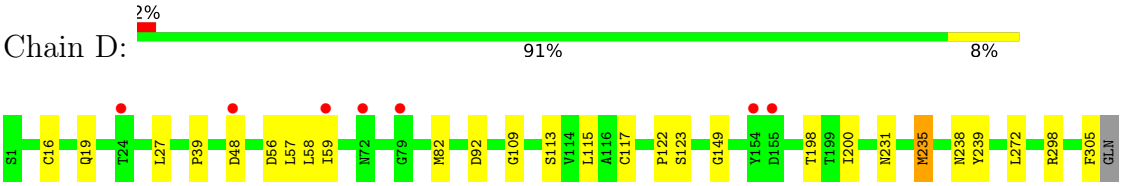
• Molecule 1: 3C-like proteinase

Chain A:



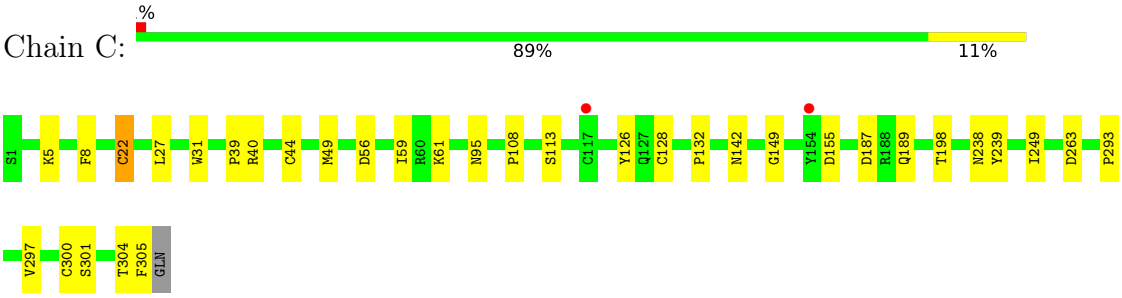
• Molecule 1: 3C-like proteinase

Chain D:



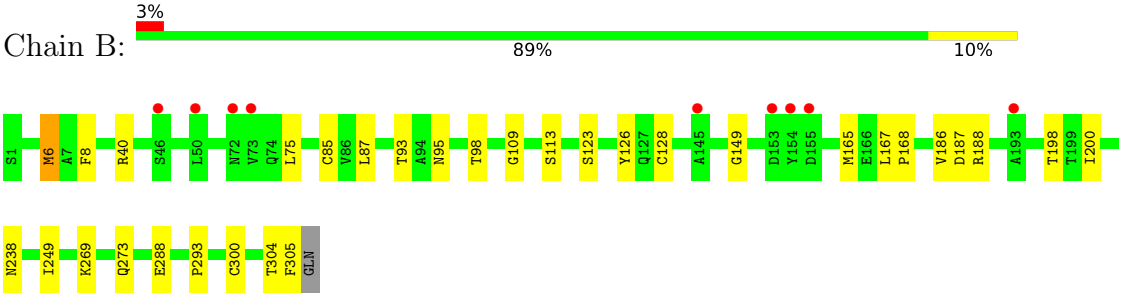
• Molecule 1: 3C-like proteinase

Chain C:



• Molecule 1: 3C-like proteinase

Chain B:



• Molecule 2: THR-SER-ALA-VAL-LEU-GLN

Chain E:

• Molecule 2: THR-SER-ALA-VAL-LEU-GLN

Chain H:

• Molecule 2: THR-SER-ALA-VAL-LEU-GLN

Chain G:

• Molecule 2: THR-SER-ALA-VAL-LEU-GLN

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	69.72Å 100.57Å 99.28Å 90.00° 104.14° 90.00°	Depositor
Resolution (Å)	29.60 – 1.83 29.60 – 1.83	Depositor EDS
% Data completeness (in resolution range)	98.3 (29.60-1.83) 94.8 (29.60-1.83)	Depositor EDS
R_{merge}	0.03	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.53 (at 1.83Å)	Xtriage
Refinement program	PHENIX 1.19.1_4122	Depositor
R, R_{free}	0.177 , 0.220 0.177 , 0.220	Depositor DCC
R_{free} test set	2000 reflections (1.73%)	wwPDB-VP
Wilson B-factor (Å^2)	23.3	Xtriage
Anisotropy	0.340	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 46.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	19767	wwPDB-VP
Average B, all atoms (Å^2)	32.0	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 80.50 % 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 4.4113e-07. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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	0.87	1/2354 (0.0%)	0.81	0/3209
1	B	0.91	3/2419 (0.1%)	0.87	3/3290 (0.1%)
1	C	0.98	6/2429 (0.2%)	0.88	3/3304 (0.1%)
1	D	0.89	2/2346 (0.1%)	0.84	1/3199 (0.0%)
2	E	1.16	0/42	0.88	0/55
2	F	0.79	0/42	1.00	0/55
2	G	1.20	1/42 (2.4%)	0.95	0/55
2	H	1.32	1/42 (2.4%)	0.95	0/55
All	All	0.92	14/9716 (0.1%)	0.85	7/13222 (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	G	0	1

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	126	TYR	CE2-CZ	-7.96	1.28	1.38
1	D	117	CYS	CB-SG	-7.95	1.68	1.82
1	C	22	CYS	CB-SG	-7.81	1.69	1.82
1	C	128	CYS	CB-SG	-7.26	1.70	1.82
1	C	300	CYS	CB-SG	-7.17	1.70	1.82

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	6	MET	CG-SD-CE	5.95	109.72	100.20
1	C	187	ASP	CB-CG-OD1	5.76	123.48	118.30
1	B	187	ASP	CB-CG-OD1	5.27	123.04	118.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	235	MET	CG-SD-CE	-5.26	91.79	100.20
1	B	165	MET	CG-SD-CE	5.19	108.51	100.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	G	-6	THR	Mainchain

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	2302	2207	2208	22	0
1	B	2358	2298	2286	17	0
1	C	2361	2294	2281	15	0
1	D	2294	2195	2196	15	0
2	E	43	47	47	2	0
2	F	43	47	47	1	0
2	G	43	47	47	2	0
2	H	43	47	47	0	0
3	A	218	0	0	1	1
3	B	301	0	0	0	1
3	C	296	0	0	1	1
3	D	262	0	0	1	2
3	E	6	0	0	0	0
3	F	4	0	0	0	0
3	G	5	0	0	0	0
3	H	6	0	0	0	0
All	All	10585	9182	9159	67	3

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.

The worst 5 of 67 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:106:ILE:HD11	1:A:130:MET:HB2	1.57	0.86
1:A:123:SER:HB2	1:B:304:THR:HG22	1.74	0.70
1:D:198:THR:HG22	1:D:238:ASN:HD21	1.58	0.69
1:D:198:THR:HG22	1:D:238:ASN:ND2	2.10	0.66
1:A:106:ILE:HD11	1:A:130:MET:CB	2.27	0.65

All (3) 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:532:HOH:O	3:D:535:HOH:O[2_757]	2.08	0.12
3:D:630:HOH:O	3:B:672:HOH:O[1_656]	2.15	0.05
3:A:516:HOH:O	3:C:615:HOH:O[2_656]	2.18	0.02

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%)	291 (96%)	12 (4%)	0	100	100
1	B	305/306 (100%)	300 (98%)	5 (2%)	0	100	100
1	C	307/306 (100%)	303 (99%)	4 (1%)	0	100	100
1	D	303/306 (99%)	291 (96%)	12 (4%)	0	100	100
2	E	4/6 (67%)	4 (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	1234/1248 (99%)	1201 (97%)	33 (3%)	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	248/262 (95%)	248 (100%)	0	100 100
1	B	261/262 (100%)	261 (100%)	0	100 100
1	C	261/262 (100%)	259 (99%)	2 (1%)	81 75
1	D	246/262 (94%)	246 (100%)	0	100 100
2	E	5/5 (100%)	5 (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%)	5 (100%)	0	100 100
All	All	1036/1068 (97%)	1034 (100%)	2 (0%)	93 92

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	5	LYS
1	C	142	ASN

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

Mol	Chain	Res	Type
1	D	180	ASN
1	D	238	ASN

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^{th} 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%)	-0.07	7 (2%) 60 58	18, 32, 48, 56	0
1	B	305/306 (99%)	-0.19	9 (2%) 50 47	12, 27, 42, 60	0
1	C	305/306 (99%)	-0.31	2 (0%) 87 87	14, 25, 39, 49	0
1	D	305/306 (99%)	-0.15	7 (2%) 60 58	16, 28, 46, 59	0
2	E	6/6 (100%)	0.08	0 100 100	22, 23, 40, 49	0
2	F	6/6 (100%)	0.59	1 (16%) 1 1	30, 31, 41, 53	0
2	G	6/6 (100%)	0.13	1 (16%) 1 1	22, 24, 35, 51	0
2	H	6/6 (100%)	-0.02	0 100 100	24, 26, 34, 46	0
All	All	1244/1248 (99%)	-0.17	27 (2%) 62 60	12, 28, 46, 60	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	154	TYR	6.3
1	D	59	ILE	4.9
1	B	72	ASN	4.2
1	A	255	ALA	4.1
1	B	154	TYR	4.1

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.