Full wwPDB X-ray Structure Validation Report (i)

Oct 30, 2024 – 12:08 PM EDT

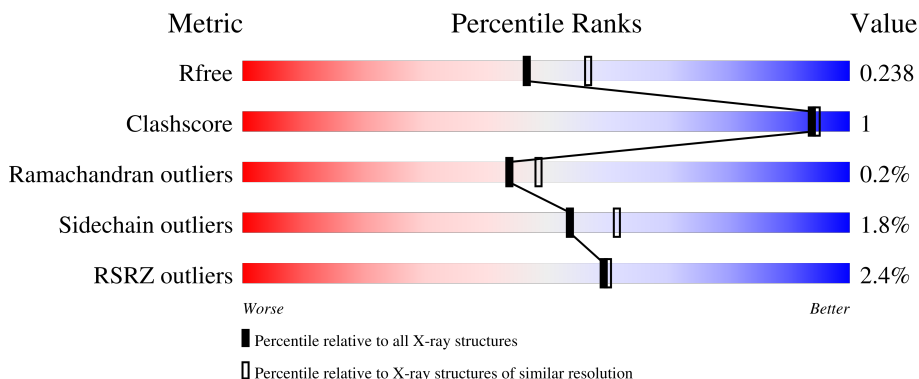|		
PDB ID	:	8U25
Title	:	Crystal Structure of SARS-CoV-2 Main Protease (Mpro) L50F/E166A/L167F Triple Mutant
Authors	:	Kohaal, N.; Lewandowski, E.M.; Wang, J.; Chen, Y.
Deposited on	:	2023-09-05
Resolution	:	2.23 Å(reported)

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 2.23 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	Percentile Ranks	Value
Rfree		0.238
Clashscore		1
Ramachandran outliers		0.2%
Sidechain outliers		1.8%
RSRZ outliers		2.4%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	3139 (2.26-2.22)
Clashscore	180529	3381 (2.26-2.22)
Ramachandran outliers	177936	3334 (2.26-2.22)
Sidechain outliers	177891	3335 (2.26-2.22)
RSRZ outliers	164620	3138 (2.26-2.22)

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	96% •
1	B	306	4% / 94% / 6%
1	C	306	2% / 96% •
1	D	306	4% / 93% / 6% •

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 10006 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	306	Total	C	N	O	S	0	1	0
			2378	1507	403	446	22			
1	B	306	Total	C	N	O	S	0	1	0
			2376	1506	403	444	23			
1	C	306	Total	C	N	O	S	0	1	0
			2376	1506	403	444	23			
1	D	306	Total	C	N	O	S	0	0	0
			2370	1503	402	443	22			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	50	PHE	LEU	engineered mutation	UNP P0DTC1
A	166	ALA	GLU	engineered mutation	UNP P0DTC1
A	167	PHE	LEU	engineered mutation	UNP P0DTC1
B	50	PHE	LEU	engineered mutation	UNP P0DTC1
B	166	ALA	GLU	engineered mutation	UNP P0DTC1
B	167	PHE	LEU	engineered mutation	UNP P0DTC1
C	50	PHE	LEU	engineered mutation	UNP P0DTC1
C	166	ALA	GLU	engineered mutation	UNP P0DTC1
C	167	PHE	LEU	engineered mutation	UNP P0DTC1
D	50	PHE	LEU	engineered mutation	UNP P0DTC1
D	166	ALA	GLU	engineered mutation	UNP P0DTC1
D	167	PHE	LEU	engineered mutation	UNP P0DTC1

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	151	Total	O	0	0
			151	151		
2	B	123	Total	O	0	0
			123	123		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	123	Total O 123 123		0	0
2	D	109	Total O 109 109		0	0

3 Residue-property plots ⓘ

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.
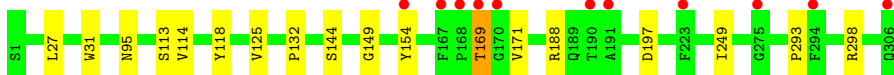
• Molecule 1: 3C-like proteinase

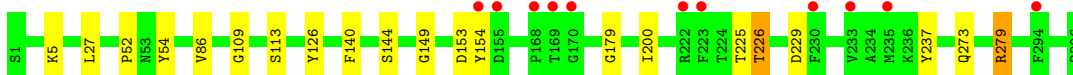Chain A:
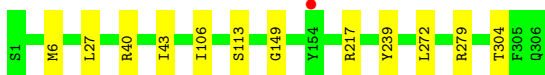
• Molecule 1: 3C-like proteinase

Chain B:

• Molecule 1: 3C-like proteinase

Chain C:

• Molecule 1: 3C-like proteinase

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	71.12Å 87.23Å 107.56Å 90.00° 104.49° 90.00°	Depositor
Resolution (Å)	47.71 – 2.23 47.71 – 2.23	Depositor EDS
% Data completeness (in resolution range)	97.5 (47.71-2.23) 97.6 (47.71-2.23)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.03 (at 2.22Å)	Xtriage
Refinement program	REFMAC 5.8.0419	Depositor
R, R_{free}	0.180 , 0.231 0.187 , 0.238	Depositor DCC
R_{free} test set	3155 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	27.1	Xtriage
Anisotropy	0.648	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 27.3	EDS
L-test for twinning[2]	$< |L| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	10006	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 4.49% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[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	0.37	0/2433	0.69	0/3305
1	B	0.37	0/2431	0.70	0/3302
1	C	0.36	0/2431	0.66	0/3302
1	D	0.36	0/2425	0.70	0/3294
All	All	0.37	0/9720	0.69	0/13203

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	2378	0	2312	5	0
1	B	2376	0	2313	7	0
1	C	2376	0	2313	3	0
1	D	2370	0	2309	11	0
2	A	151	0	0	0	0
2	B	123	0	0	0	0
2	C	123	0	0	0	0
2	D	109	0	0	1	0
All	All	10006	0	9247	25	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 1.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

All (25) 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:D:279:ARG:HH21	1:D:279:ARG:HB3	1.48	0.79
1:D:226:THR:HG23	1:D:229:ASP:HB2	1.93	0.51
1:B:169:THR:HB	1:B:171:VAL:HG23	1.92	0.50
1:D:113:SER:O	1:D:149:GLY:HA2	2.12	0.50
1:A:279:ARG:HH11	1:A:279:ARG:HG3	1.77	0.49
1:B:118:TYR:CE1	1:B:144:SER:HB3	2.46	0.49
1:A:6:MET:HG2	1:D:126:TYR:CD1	2.48	0.48
1:C:86:VAL:HG23	1:C:179:GLY:HA2	1.96	0.48
1:D:109:GLY:HA2	1:D:200:ILE:HD13	1.94	0.48
1:B:132:PRO:HD2	1:B:197:ASP:OD1	2.13	0.47
1:C:31:TRP:CD2	1:C:95:ASN:HB2	2.49	0.47
1:D:226:THR:CG2	1:D:229:ASP:OD2	2.63	0.47
1:B:31:TRP:CD2	1:B:95:ASN:HB2	2.51	0.45
1:A:40:ARG:O	1:A:43:ILE:HG12	2.16	0.45
1:D:86:VAL:HG23	1:D:179:GLY:HA2	1.98	0.45
1:D:140:PHE:HB3	1:D:144:SER:OG	2.17	0.45
1:A:113:SER:O	1:A:149:GLY:HA2	2.18	0.43
1:B:249:ILE:HG22	1:B:293:PRO:HG2	2.02	0.42
1:C:113:SER:O	1:C:149:GLY:HA2	2.19	0.42
1:B:114:VAL:O	1:B:125:VAL:HA	2.20	0.42
1:D:52:PRO:HG2	1:D:54:TYR:CE2	2.55	0.41
1:D:237:TYR:OH	1:D:273:GLN:HA	2.20	0.41
1:D:225:THR:HG23	2:D:506:HOH:O	2.20	0.41
1:A:239:TYR:CZ	1:A:272:LEU:HD21	2.55	0.40
1:B:113:SER:O	1:B:149:GLY:HA2	2.20	0.40

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	305/306 (100%)	302 (99%)	3 (1%)	0	100	100
1	B	305/306 (100%)	302 (99%)	2 (1%)	1 (0%)	37	39
1	C	305/306 (100%)	298 (98%)	7 (2%)	0	100	100
1	D	304/306 (99%)	294 (97%)	9 (3%)	1 (0%)	37	39
All	All	1219/1224 (100%)	1196 (98%)	21 (2%)	2 (0%)	44	49

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	154	TYR
1	D	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	263/262 (100%)	259 (98%)	4 (2%)	60	69
1	B	263/262 (100%)	259 (98%)	4 (2%)	60	69
1	C	263/262 (100%)	257 (98%)	6 (2%)	45	52
1	D	262/262 (100%)	257 (98%)	5 (2%)	52	60
All	All	1051/1048 (100%)	1032 (98%)	19 (2%)	54	61

All (19) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	27	LEU
1	A	106	ILE
1	A	217	ARG
1	A	304	THR
1	B	27	LEU
1	B	169	THR
1	B	188	ARG
1	B	298	ARG
1	C	51	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	216	ASP
1	C	217	ARG
1	C	222	ARG
1	C	245	ASP
1	C	304	THR
1	D	5	LYS
1	D	27	LEU
1	D	153	ASP
1	D	226	THR
1	D	279	ARG

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

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 oligosaccharides in this entry.

5.6   Ligand geometry ⓘ

There are no ligands in this entry.

5.7   Other polymers ⓘ

There are no such residues in this entry.

5.8   Polymer linkage issues ⓘ

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	306/306 (100%)	-0.41	1 (0%)	90	91	9, 24, 41, 69	3 (0%)
1	B	306/306 (100%)	-0.06	11 (3%)	46	46	13, 30, 61, 87	3 (0%)
1	C	306/306 (100%)	-0.25	6 (1%)	64	65	10, 28, 49, 86	3 (0%)
1	D	306/306 (100%)	0.06	11 (3%)	46	46	19, 32, 69, 95	2 (0%)
All	All	1224/1224 (100%)	-0.17	29 (2%)	59	60	9, 28, 58, 95	11 (0%)

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	168	PRO	3.9
1	D	168	PRO	3.8
1	D	155	ASP	3.8
1	D	170	GLY	3.5
1	B	306	GLN	3.2
1	C	222	ARG	3.2
1	A	154	TYR	3.1
1	D	154	TYR	3.0
1	D	233	VAL	3.0
1	B	191	ALA	3.0
1	B	170	GLY	3.0
1	D	169	THR	2.9
1	C	223	PHE	2.9
1	B	169	THR	2.8
1	B	294	PHE	2.6
1	B	154	TYR	2.5
1	C	154	TYR	2.4
1	C	72	ASN	2.4
1	D	222	ARG	2.4
1	B	223	PHE	2.3
1	B	275	GLY	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	224	THR	2.2
1	B	167	PHE	2.2
1	D	294	PHE	2.2
1	C	155	ASP	2.2
1	D	223	PHE	2.1
1	D	235	MET	2.1
1	B	190	THR	2.1
1	D	230	PHE	2.1

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.