Full wwPDB X-ray Structure Validation Report ⓘ

May 23, 2022 – 01:08 pm BST

PDB ID	:	7ZB8
Title	:	Crystal Structure of SARS-CoV-2 Main Protease (Mpro) variant K61A at 2.48 A resolution
Authors	:	Paknia, E.; Rabe von Pappenheim, F.; Funk, L.-M.; Tittmann, K.; Chari, A.
Deposited on	:	2022-03-23
Resolution	:	2.48 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 following versions of software and data (see references ⓘ) were used in the production of this report:

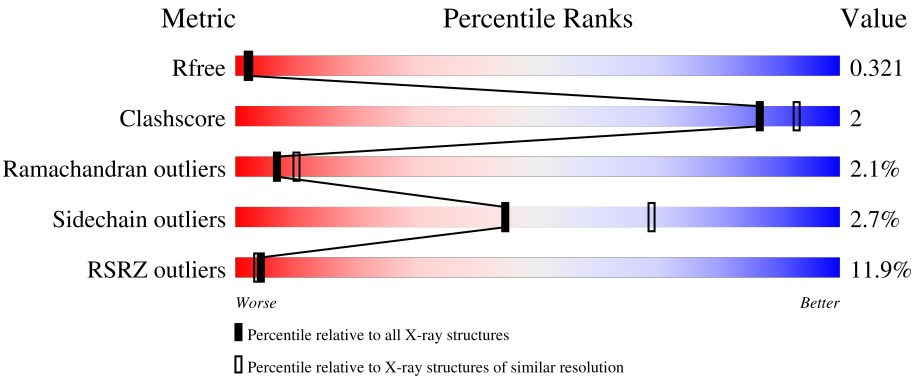
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.28.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.48 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	5857 (2.50-2.46)
Clashscore	141614	6594 (2.50-2.46)
Ramachandran outliers	138981	6469 (2.50-2.46)
Sidechain outliers	138945	6471 (2.50-2.46)
RSRZ outliers	127900	5738 (2.50-2.46)

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	6% 92% 8%
1	C	306	18% 89% 10%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9391 atoms, of which 4613 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.

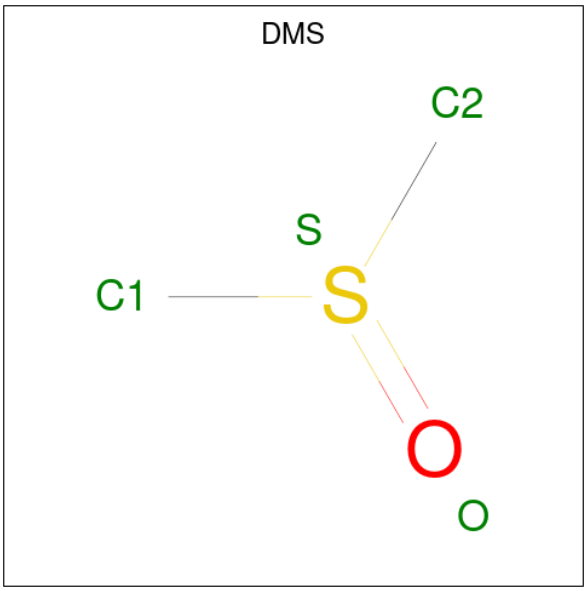
- Molecule 1 is a protein called 3C-like proteinase nsp5.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	306	Total	C	H	N	O	S	0	0	0
			4657	1496	2294	401	444	22			
1	C	306	Total	C	H	N	O	S	0	0	0
			4658	1496	2295	401	444	22			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	61	ALA	LYS	engineered mutation	UNP P0DTD1
C	61	ALA	LYS	engineered mutation	UNP P0DTD1

- Molecule 2 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C_2H_6OS).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	C	1	Total	C	H	O	S	0	0
			10	2	6	1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	C	1	Total	C	H	O	S	0	0
			10	2	6	1	1		
2	C	1	Total	C	H	O	S	0	0
			10	2	6	1	1		
2	C	1	Total	C	H	O	S	0	0
			10	2	6	1	1		

- Molecule 3 is water.

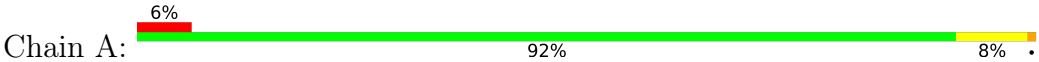
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	22	Total	O	0	0
			22	22		
3	C	14	Total	O	0	0
			14	14		

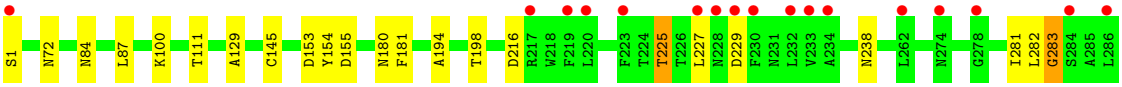
3 Residue-property plots (i)

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 nsp5

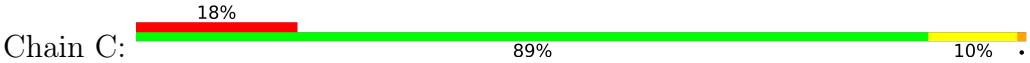
Chain A:

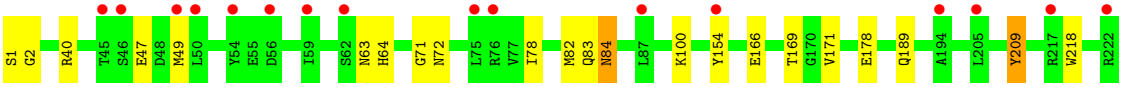


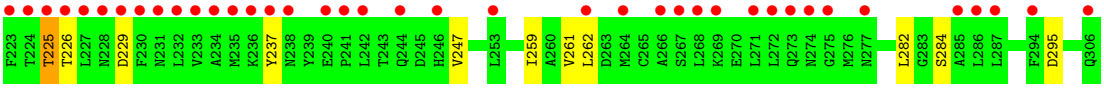


• Molecule 1: 3C-like proteinase nsp5

Chain C:







4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	122.48Å 81.29Å 63.70Å 90.00° 90.22° 90.00°	Depositor
Resolution (Å)	46.45 – 2.48 46.45 – 2.48	Depositor EDS
% Data completeness (in resolution range)	75.0 (46.45-2.48) 75.0 (46.45-2.48)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.56 (at 2.48Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.246 , 0.332 0.245 , 0.321	Depositor DCC
R_{free} test set	885 reflections (5.28%)	wwPDB-VP
Wilson B-factor (Å^2)	61.5	Xtriage
Anisotropy	0.032	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.014 for -h,-k,l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	9391	wwPDB-VP
Average B, all atoms (Å^2)	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.23% 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

Bond lengths and bond angles in the following residue types are not validated in this section: DMS

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	0.25	0/2416	0.47	0/3285
1	C	0.25	0/2416	0.46	0/3285
All	All	0.25	0/4832	0.46	0/6570

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	2363	2294	2306	7	1
1	C	2363	2295	2306	14	0
2	C	16	24	24	0	0
3	A	22	0	0	0	0
3	C	14	0	0	0	0
All	All	4778	4613	4636	20	1

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:A:1:SER:N	1:C:166:GLU:OE2	2.28	0.66
1:C:63:ASN:ND2	1:C:78:ILE:O	2.36	0.59
1:A:84:ASN:ND2	1:A:180:ASN:OD1	2.36	0.57
1:C:261:VAL:HG13	1:C:262:LEU:HD22	1.86	0.56
1:A:111:THR:HG22	1:A:129:ALA:HB2	1.88	0.54
1:A:281:ILE:O	1:A:283:GLY:N	2.40	0.54
1:C:40:ARG:NH2	1:C:82:MET:SD	2.81	0.53
1:C:247:VAL:HG13	1:C:261:VAL:HG11	1.90	0.53
1:A:111:THR:HG23	1:A:292:THR:HG23	1.92	0.50
1:C:49:MET:SD	1:C:189:GLN:NE2	2.86	0.48
1:C:40:ARG:NH2	1:C:84:ASN:O	2.48	0.47
1:C:225:THR:OG1	1:C:229:ASP:OD1	2.34	0.46
1:C:1:SER:OG	1:C:2:GLY:N	2.49	0.45
1:C:83:GLN:NE2	1:C:178:GLU:OE1	2.50	0.44
1:A:153:ASP:O	1:A:155:ASP:N	2.51	0.43
1:C:209:TYR:OH	1:C:259:ILE:O	2.36	0.43
1:C:71:GLY:O	1:C:72:ASN:OD1	2.37	0.42
1:A:225:THR:OG1	1:A:229:ASP:OD2	2.38	0.41
1:C:71:GLY:C	1:C:72:ASN:OD1	2.59	0.41
1:C:169:THR:HG23	1:C:171:VAL:HG22	2.03	0.41

All (1) 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)
1:A:145:CYS:SG	1:A:306:GLN:C[4_545]	1.80	0.40

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	304/306 (99%)	272 (90%)	25 (8%)	7 (2%)	6 9

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	C	304/306 (99%)	264 (87%)	34 (11%)	6 (2%)	7 11
All	All	608/612 (99%)	536 (88%)	59 (10%)	13 (2%)	7 10

All (13) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	154	TYR
1	A	194	ALA
1	A	238	ASN
1	C	47	GLU
1	C	154	TYR
1	C	218	TRP
1	C	284	SER
1	A	198	THR
1	A	282	LEU
1	A	283	GLY
1	C	84	ASN
1	A	216	ASP
1	C	100	LYS

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	262/262 (100%)	255 (97%)	7 (3%)	44 69
1	C	262/262 (100%)	255 (97%)	7 (3%)	44 69
All	All	524/524 (100%)	510 (97%)	14 (3%)	44 69

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	72	ASN
1	A	87	LEU
1	A	100	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	181	PHE
1	A	225	THR
1	A	227	LEU
1	A	304	THR
1	C	64	HIS
1	C	209	TYR
1	C	225	THR
1	C	226	THR
1	C	237	TYR
1	C	282	LEU
1	C	295	ASP

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

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)

4 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	DMS	C	404	-	3,3,3	0.66	0	3,3,3	0.49	0
2	DMS	C	402	-	3,3,3	0.66	0	3,3,3	0.52	0
2	DMS	C	403	-	3,3,3	0.66	0	3,3,3	0.51	0
2	DMS	C	401	-	3,3,3	0.66	0	3,3,3	0.51	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	306/306 (100%)	0.66	18 (5%)	22	22	30, 67, 154, 238	0
1	C	306/306 (100%)	0.90	55 (17%)	1	1	28, 78, 168, 227	0
All	All	612/612 (100%)	0.78	73 (11%)	4	3	28, 73, 165, 238	0

All (73) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	232	LEU	9.5
1	A	233	VAL	9.4
1	C	227	LEU	9.0
1	A	228	ASN	6.9
1	A	278	GLY	6.3
1	C	242	LEU	6.1
1	C	224	THR	6.0
1	C	236	LYS	5.7
1	C	226	THR	5.6
1	C	228	ASN	5.6
1	C	230	PHE	5.6
1	A	220	LEU	5.6
1	C	274	ASN	5.6
1	C	46	SER	5.0
1	C	271	LEU	4.6
1	C	49	MET	4.5
1	C	238	ASN	4.5
1	C	237	TYR	3.9
1	C	253	LEU	3.9
1	A	262	LEU	3.9
1	C	267	SER	3.8
1	C	285	ALA	3.8
1	C	194	ALA	3.6
1	C	286	LEU	3.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	225	THR	3.4
1	C	273	GLN	3.4
1	C	272	LEU	3.2
1	C	45	THR	3.1
1	A	287	LEU	3.1
1	C	266	ALA	3.0
1	C	154	TYR	3.0
1	A	1	SER	3.0
1	A	286	LEU	3.0
1	C	240	GLU	3.0
1	C	287	LEU	2.9
1	C	246	HIS	2.9
1	C	217	ARG	2.9
1	A	217	ARG	2.9
1	C	277	ASN	2.8
1	A	284	SER	2.8
1	C	75	LEU	2.8
1	C	275	GLY	2.7
1	C	54	TYR	2.7
1	C	50	LEU	2.6
1	A	227	LEU	2.6
1	C	235	MET	2.6
1	A	223	PHE	2.6
1	C	223	PHE	2.6
1	C	268	LEU	2.5
1	C	205	LEU	2.5
1	C	262	LEU	2.5
1	C	269	LYS	2.5
1	A	229	ASP	2.5
1	C	232	LEU	2.5
1	C	244	GLN	2.5
1	C	76	ARG	2.4
1	A	274	ASN	2.4
1	C	294	PHE	2.4
1	C	62	SER	2.4
1	C	222	ARG	2.3
1	C	231	ASN	2.3
1	A	234	ALA	2.2
1	C	234	ALA	2.2
1	C	56	ASP	2.2
1	C	264	MET	2.1
1	C	59	ILE	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	87	LEU	2.1
1	C	241	PRO	2.1
1	A	219	PHE	2.1
1	C	229	ASP	2.1
1	C	306	GLN	2.1
1	A	230	PHE	2.1
1	C	233	VAL	2.1

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)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	DMS	C	401	4/4	0.73	0.34	165,198,202,202	0
2	DMS	C	402	4/4	0.91	0.32	81,112,114,114	0
2	DMS	C	404	4/4	0.91	0.20	82,105,113,113	0
2	DMS	C	403	4/4	0.92	0.34	96,118,127,127	0

6.5 Other polymers (i)

There are no such residues in this entry.