Full wwPDB X-ray Structure Validation Report ⓘ

Jul 29, 2024 – 02:54 PM EDT

PDB ID	:	9AUO
Title	:	Structure of SARS-CoV-2 Mpro mutant (L50F,T304I)
Authors	:	Gajiwala, K.S.; Greasley, S.E.; Ferre, R.A.; Liu, W.; Stewart, A.E.
Deposited on	:	2024-02-29
Resolution	:	2.42 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

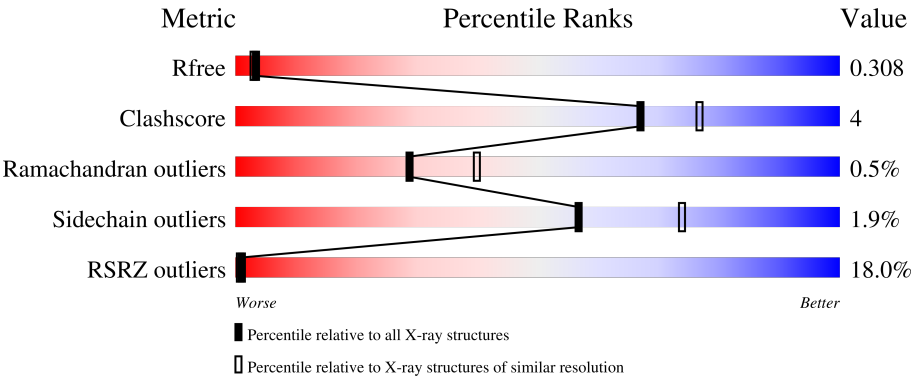
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.42 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	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)
RSRZ outliers	127900	4543 (2.44-2.40)

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	16% 88% 12%
1	B	306	20% 88% 10%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4704 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 nsp5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	306	Total	C	N	O	S	0	0	0
			2371	1504	402	443	22			
1	B	301	Total	C	N	O	S	0	0	0
			2332	1477	396	437	22			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	50	PHE	LEU	engineered mutation	UNP P0DTD1
A	304	ILE	THR	engineered mutation	UNP P0DTD1
B	50	PHE	LEU	engineered mutation	UNP P0DTD1
B	304	ILE	THR	engineered mutation	UNP P0DTD1

- Molecule 2 is water.

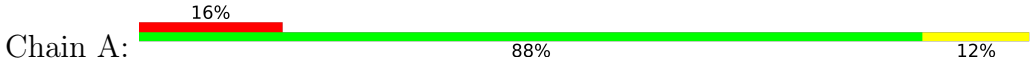
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	O	0	0
			1	1		

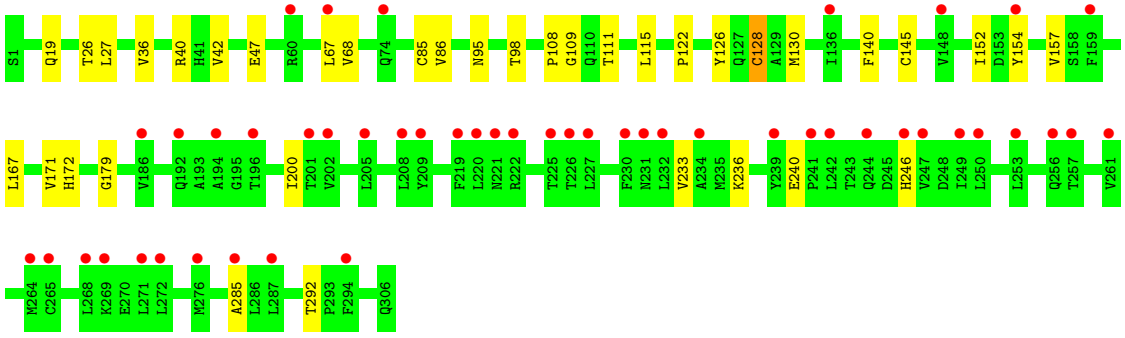
3 Residue-property plots (i)

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

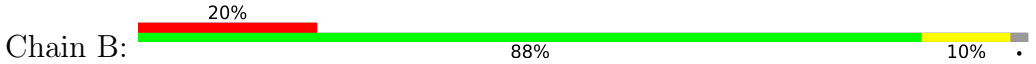
Chain A:

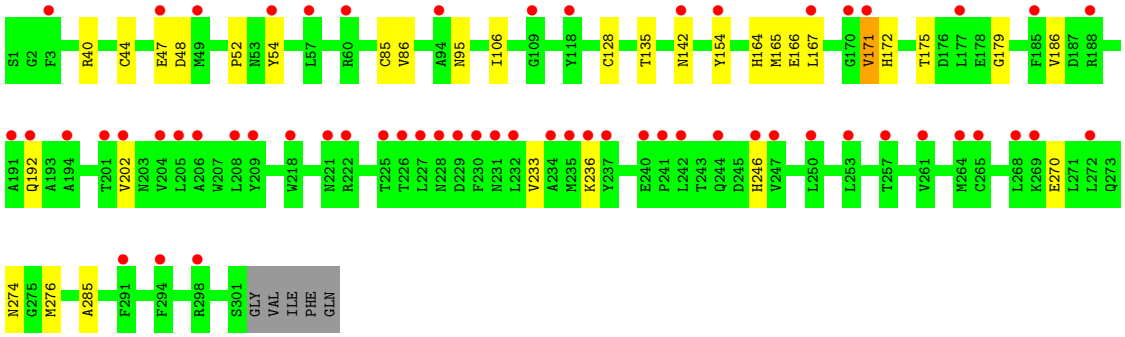




• Molecule 1: 3C-like proteinase nsp5

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	121.87Å 82.70Å 64.23Å 90.00° 90.38° 90.00°	Depositor
Resolution (Å)	68.43 – 2.42 68.43 – 2.42	Depositor EDS
% Data completeness (in resolution range)	83.4 (68.43-2.42) 83.4 (68.43-2.42)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.33 (at 2.42Å)	Xtriage
Refinement program	BUSTER 2.11.8 (8-JUN-2022)	Depositor
R, R_{free}	0.276 , 0.315 0.269 , 0.308	Depositor DCC
R_{free} test set	1036 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	83.5	Xtriage
Anisotropy	0.009	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 52.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.019 for -h,-k,l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	4704	wwPDB-VP
Average B, all atoms (Å^2)	98.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.10% 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	0.33	0/2425	0.56	0/3295
1	B	0.30	0/2385	0.49	0/3241
All	All	0.32	0/4810	0.52	0/6536

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	2371	0	2316	20	0
1	B	2332	0	2276	17	0
2	A	1	0	0	0	0
All	All	4704	0	4592	36	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 4.

All (36) 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:233:VAL:HA	1:A:236:LYS:NZ	1.88	0.89

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:233:VAL:HA	1:B:236:LYS:NZ	1.98	0.78
1:A:233:VAL:HA	1:A:236:LYS:HZ3	1.48	0.77
1:B:233:VAL:HA	1:B:236:LYS:HZ1	1.58	0.69
1:B:186:VAL:H	1:B:192:GLN:HE22	1.41	0.68
1:A:233:VAL:HA	1:A:236:LYS:HZ1	1.62	0.60
1:B:48:ASP:HB3	1:B:52:PRO:HB3	1.85	0.59
1:A:109:GLY:HA2	1:A:200:ILE:HD13	1.89	0.55
1:B:166:GLU:OE2	1:B:172:HIS:HD2	1.88	0.55
1:B:40:ARG:HD3	1:B:85:CYS:HA	1.90	0.53
1:B:233:VAL:HA	1:B:236:LYS:HZ3	1.72	0.53
1:A:86:VAL:HG13	1:A:179:GLY:HA2	1.92	0.51
1:B:86:VAL:HG13	1:B:179:GLY:HA2	1.91	0.51
1:A:115:LEU:HD11	1:A:122:PRO:HB3	1.93	0.50
1:B:44:CYS:SG	1:B:54:TYR:CE1	3.06	0.49
1:A:140:PHE:HD2	1:A:172:HIS:CG	2.31	0.48
1:A:200:ILE:HA	1:A:240:GLU:HG3	1.95	0.48
1:B:270:GLU:OE1	1:B:274:ASN:ND2	2.47	0.48
1:B:164:HIS:CE1	1:B:175:THR:HG23	2.49	0.48
1:A:36:VAL:HG21	1:A:68:VAL:HG11	1.96	0.47
1:A:152:ILE:HG12	1:A:157:VAL:HG22	1.97	0.46
1:A:19:GLN:HE21	1:A:26:THR:HG21	1.80	0.46
1:A:111:THR:HG23	1:A:292:THR:HG23	1.98	0.46
1:B:166:GLU:OE2	1:B:172:HIS:CD2	2.69	0.45
1:A:167:LEU:HD12	1:A:171:VAL:HG23	1.99	0.44
1:A:40:ARG:HD3	1:A:85:CYS:HA	2.00	0.44
1:B:202:VAL:HG11	1:B:246:HIS:HD2	1.83	0.44
1:B:135:THR:HB	1:B:171:VAL:HG11	1.99	0.44
1:A:108:PRO:HA	1:A:130:MET:CG	2.48	0.43
1:A:95:ASN:HB3	1:A:98:THR:OG1	2.19	0.42
1:A:27:LEU:HB2	1:A:145:CYS:O	2.19	0.42
1:A:126:TYR:HE2	1:A:128:CYS:HG	1.68	0.42
1:B:167:LEU:HB2	1:B:171:VAL:C	2.41	0.41
1:A:285:ALA:HB3	1:B:285:ALA:HB3	2.04	0.40
1:A:27:LEU:HD21	1:A:42:VAL:HB	2.03	0.40
1:B:276:MET:SD	1:B:285:ALA:O	2.80	0.40

There are no symmetry-related clashes.

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%)	291 (96%)	12 (4%)	1 (0%)	41 54
1	B	299/306 (98%)	283 (95%)	14 (5%)	2 (1%)	22 31
All	All	603/612 (98%)	574 (95%)	26 (4%)	3 (0%)	29 40

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	154	TYR
1	B	95	ASN
1	B	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	263/263 (100%)	259 (98%)	4 (2%)	65 79
1	B	259/263 (98%)	253 (98%)	6 (2%)	50 68
All	All	522/526 (99%)	512 (98%)	10 (2%)	57 74

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	47	GLU
1	A	67	LEU
1	A	128	CYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	246	HIS
1	B	47	GLU
1	B	106	ILE
1	B	128	CYS
1	B	142	ASN
1	B	165	MET
1	B	171	VAL

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	19	GLN
1	B	119	ASN
1	B	172	HIS
1	B	192	GLN

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	306/306 (100%)	0.94	49 (16%)	1	1	58, 80, 134, 143	0
1	B	301/306 (98%)	1.28	60 (19%)	1	0	61, 102, 161, 167	0
All	All	607/612 (99%)	1.11	109 (17%)	1	1	58, 89, 154, 167	0

All (109) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	246	HIS	14.0
1	B	205	LEU	9.4
1	A	230	PHE	8.7
1	B	242	LEU	8.5
1	B	171	VAL	8.4
1	B	227	LEU	8.4
1	A	232	LEU	8.1
1	A	242	LEU	6.9
1	B	247	VAL	6.5
1	B	230	PHE	6.4
1	B	154	TYR	6.3
1	A	249	ILE	6.2
1	B	261	VAL	6.0
1	B	209	TYR	5.7
1	B	232	LEU	5.6
1	B	268	LEU	5.3
1	A	220	LEU	5.1
1	B	167	LEU	5.0
1	B	228	ASN	5.0
1	B	191	ALA	5.0
1	B	188	ARG	4.9
1	B	142	ASN	4.9
1	B	170	GLY	4.9
1	A	246	HIS	4.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	235	MET	4.8
1	A	285	ALA	4.7
1	B	253	LEU	4.7
1	B	204	VAL	4.6
1	A	276	MET	4.5
1	B	202	VAL	4.4
1	B	47	GLU	4.4
1	B	229	ASP	4.3
1	B	236	LYS	4.2
1	A	268	LEU	4.1
1	A	60	ARG	4.0
1	A	209	TYR	3.9
1	A	264	MET	3.8
1	A	201	THR	3.8
1	A	287	LEU	3.8
1	B	265	CYS	3.7
1	A	194	ALA	3.7
1	B	269	LYS	3.6
1	B	240	GLU	3.5
1	A	67	LEU	3.4
1	B	294	PHE	3.4
1	B	226	THR	3.4
1	B	231	ASN	3.4
1	B	241	PRO	3.4
1	B	264	MET	3.4
1	B	3	PHE	3.3
1	A	250	LEU	3.3
1	B	208	LEU	3.2
1	A	269	LYS	3.2
1	A	225	THR	3.2
1	A	222	ARG	3.2
1	B	49	MET	3.1
1	A	154	TYR	3.0
1	A	227	LEU	3.0
1	A	261	VAL	3.0
1	B	185	PHE	3.0
1	B	60	ARG	3.0
1	A	234	ALA	3.0
1	B	222	ARG	2.9
1	B	194	ALA	2.9
1	A	244	GLN	2.9
1	A	186	VAL	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	218	TRP	2.8
1	A	192	GLN	2.7
1	B	272	LEU	2.7
1	A	257	THR	2.7
1	B	94	ALA	2.7
1	A	226	THR	2.6
1	A	239	TYR	2.6
1	B	291	PHE	2.6
1	A	219	PHE	2.6
1	B	221	ASN	2.5
1	B	118	TYR	2.5
1	A	202	VAL	2.5
1	A	196	THR	2.5
1	B	298	ARG	2.5
1	A	221	ASN	2.5
1	A	294	PHE	2.5
1	A	271	LEU	2.5
1	B	257	THR	2.5
1	A	231	ASN	2.4
1	A	241	PRO	2.4
1	B	244	GLN	2.4
1	A	253	LEU	2.4
1	A	272	LEU	2.4
1	B	57	LEU	2.4
1	B	237	TYR	2.3
1	B	201	THR	2.3
1	B	234	ALA	2.3
1	A	256	GLN	2.3
1	B	109	GLY	2.3
1	A	247	VAL	2.2
1	A	148	VAL	2.2
1	A	205	LEU	2.2
1	A	208	LEU	2.2
1	B	206	ALA	2.2
1	A	265	CYS	2.2
1	B	250	LEU	2.1
1	B	192	GLN	2.1
1	B	225	THR	2.1
1	A	159	PHE	2.1
1	B	177	LEU	2.1
1	A	136	ILE	2.0
1	B	54	TYR	2.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	74	GLN	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.