



Full wwPDB X-ray Structure Validation Report ⓘ

Sep 24, 2024 – 02:09 PM JST

PDB ID : 8WSJ
Title : Crystal structure of SARS-Cov-2 main protease, pH=6.5
Authors : Jiang, H.H.; Zhou, X.L.; Zhang, J.; Li, J.
Deposited on : 2023-10-17
Resolution : 1.74 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 : 3.0
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.002 (Gargrove)
Density-Fitness : 1.0.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.38.2

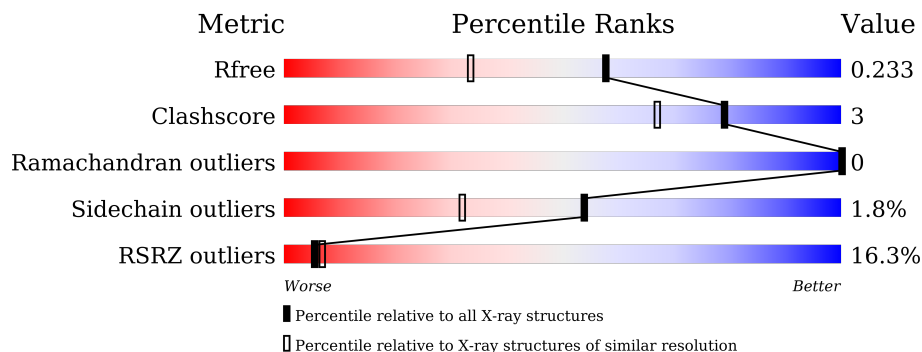
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.74 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}	164625	1043 (1.74-1.74)
Clashscore	180529	1119 (1.74-1.74)
Ramachandran outliers	177936	1112 (1.74-1.74)
Sidechain outliers	177891	1112 (1.74-1.74)
RSRZ outliers	164620	1043 (1.74-1.74)

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	295	 12% 94% 6%
2	B	299	 21% 91% 9%

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4617 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
			Total	C	N	O	S			
1	A	295	2217	1403	378	415	21	0	0	0

- Molecule 2 is a protein called 3C-like proteinase nsp5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	299	2176	1383	372	402	19	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	140	LEU	PHE	conflict	UNP P0DTC1
B	141	PHE	LEU	conflict	UNP P0DTC1

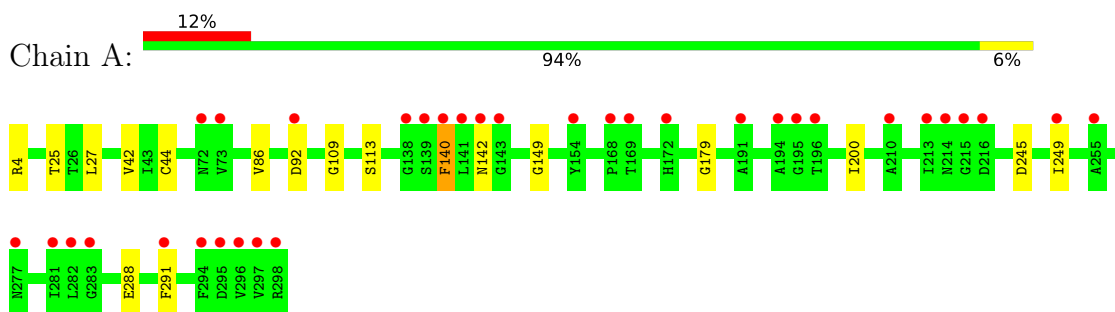
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	137	Total 137	O 137	0	0
3	B	87	Total 87	O 87	0	0

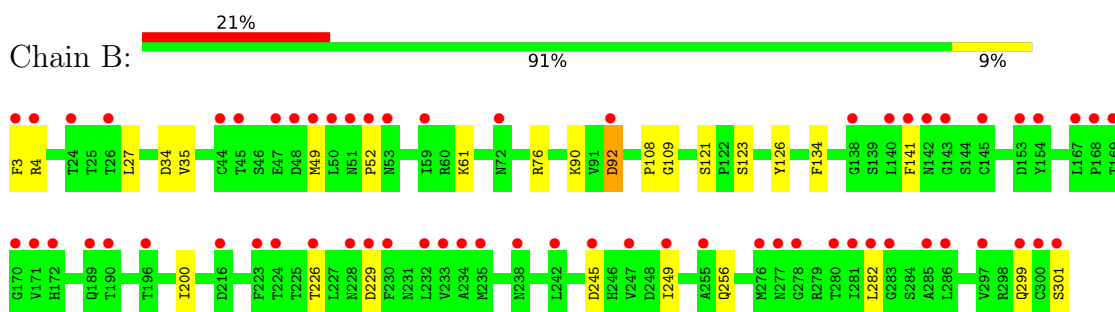
3 Residue-property plots [i](#)

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 nsp5



- Molecule 2: 3C-like proteinase nsp5



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	67.84Å 102.73Å 102.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.49 – 1.74 32.49 – 1.74	Depositor EDS
% Data completeness (in resolution range)	97.4 (32.49-1.74) 97.3 (32.49-1.74)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.24 (at 1.75Å)	Xtrriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.209 , 0.233 0.209 , 0.233	Depositor DCC
R_{free} test set	3622 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å ²)	27.2	Xtrriage
Anisotropy	0.517	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 48.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.011 for -h,l,k	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	4617	wwPDB-VP
Average B, all atoms (Å ²)	34.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.52% of the height of the origin peak. No significant pseudotranslation is detected.*

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	0.37	0/2266	0.58	0/3089
2	B	0.37	0/2226	0.56	0/3044
All	All	0.37	0/4492	0.57	0/6133

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 [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	2217	0	2113	9	0
2	B	2176	0	2022	15	0
3	A	137	0	0	0	0
3	B	87	0	0	1	0
All	All	4617	0	4135	22	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 3.

All (22) 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:25:THR:HG21	1:A:44:CYS:O	1.92	0.69

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:226:THR:HG23	2:B:229:ASP:H	1.61	0.66
2:B:245:ASP:O	2:B:249:ILE:HG13	1.97	0.63
2:B:3:PHE:HB2	2:B:282:LEU:HD22	1.84	0.59
2:B:108:PRO:HG3	2:B:134:PHE:CE1	2.44	0.53
2:B:61:LYS:NZ	3:B:402:HOH:O	2.41	0.53
1:A:245:ASP:O	1:A:249:ILE:HD13	2.08	0.52
1:A:109:GLY:HA2	1:A:200:ILE:HD13	1.92	0.52
2:B:35:VAL:HG22	2:B:90:LYS:HE2	1.91	0.52
2:B:109:GLY:HA2	2:B:200:ILE:HD13	1.93	0.50
2:B:76:ARG:CB	2:B:92:ASP:OD2	2.60	0.50
2:B:226:THR:CG2	2:B:229:ASP:H	2.25	0.49
1:A:25:THR:OG1	1:A:42:VAL:O	2.31	0.47
2:B:256:GLN:OE1	2:B:301:SER:HB3	2.16	0.46
2:B:34:ASP:O	2:B:90:LYS:HD3	2.15	0.46
1:A:86:VAL:HG23	1:A:179:GLY:HA2	1.97	0.46
2:B:123:SER:OG	2:B:141:PHE:HZ	1.99	0.45
2:B:49:MET:HA	2:B:52:PRO:HG3	1.99	0.45
1:A:288:GLU:HG2	1:A:291:PHE:CE1	2.52	0.44
1:A:140:PHE:CD2	2:B:299:GLN:O	2.70	0.44
1:A:113:SER:O	1:A:149:GLY:HA2	2.19	0.42
1:A:4:ARG:HD2	2:B:126:TYR:CD2	2.56	0.41

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	293/295 (99%)	285 (97%)	8 (3%)	0	100	100
2	B	297/299 (99%)	292 (98%)	5 (2%)	0	100	100
All	All	590/594 (99%)	577 (98%)	13 (2%)	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	236/254 (93%)	232 (98%)	4 (2%)	56	34
2	B	220/258 (85%)	216 (98%)	4 (2%)	54	32
All	All	456/512 (89%)	448 (98%)	8 (2%)	54	32

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	27	LEU
1	A	92	ASP
1	A	140	PHE
1	A	142	ASN
2	B	4	ARG
2	B	27	LEU
2	B	92	ASP
2	B	121	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	142	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 oligosaccharides 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, 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	295/295 (100%)	0.65	34 (11%) 11 15	21, 30, 49, 65	0
2	B	299/299 (100%)	0.98	63 (21%) 3 3	22, 36, 55, 68	0
All	All	594/594 (100%)	0.82	97 (16%) 5 7	21, 33, 54, 68	0

All (97) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	140	PHE	6.3
2	B	141	PHE	5.8
1	A	282	LEU	5.7
1	A	281	ILE	5.7
1	A	297	VAL	5.6
2	B	140	LEU	5.2
1	A	141	LEU	4.7
2	B	50	LEU	4.5
2	B	281	ILE	4.5
2	B	3	PHE	4.3
2	B	45	THR	4.1
1	A	296	VAL	3.9
2	B	234	ALA	3.9
2	B	282	LEU	3.8
2	B	300	CYS	3.8
2	B	277	ASN	3.6
2	B	242	LEU	3.5
2	B	286	LEU	3.5
1	A	72	ASN	3.3
2	B	226	THR	3.3
1	A	298	ARG	3.2
2	B	168	PRO	3.2
2	B	255	ALA	3.2
2	B	92	ASP	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	51	ASN	3.1
1	A	213	ILE	3.0
2	B	172	HIS	3.0
2	B	49	MET	3.0
2	B	72	ASN	3.0
2	B	233	VAL	2.9
2	B	24	THR	2.9
1	A	216	ASP	2.9
1	A	142	ASN	2.9
1	A	92	ASP	2.9
2	B	245	ASP	2.8
2	B	142	ASN	2.8
2	B	189	GLN	2.8
1	A	196	THR	2.8
2	B	230	PHE	2.8
2	B	229	ASP	2.7
1	A	194	ALA	2.7
1	A	210	ALA	2.7
2	B	154	TYR	2.7
2	B	224	THR	2.7
2	B	47	GLU	2.7
1	A	154	TYR	2.7
2	B	228	ASN	2.6
1	A	195	GLY	2.6
2	B	278	GLY	2.6
2	B	48	ASP	2.6
1	A	169	THR	2.6
2	B	59	ILE	2.6
2	B	280	THR	2.6
2	B	44	CYS	2.5
2	B	285	ALA	2.5
2	B	216	ASP	2.5
2	B	196	THR	2.5
2	B	283	GLY	2.5
2	B	223	PHE	2.5
2	B	190	THR	2.4
1	A	283	GLY	2.4
2	B	235	MET	2.4
2	B	232	LEU	2.4
1	A	214	ASN	2.4
2	B	171	VAL	2.4
2	B	247	VAL	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	143	GLY	2.4
1	A	295	ASP	2.4
2	B	138	GLY	2.4
1	A	168	PRO	2.3
2	B	52	PRO	2.3
2	B	153	ASP	2.3
2	B	276	MET	2.3
2	B	145	CYS	2.3
2	B	238	ASN	2.3
1	A	215	GLY	2.3
2	B	299	GLN	2.2
1	A	191	ALA	2.2
1	A	277	ASN	2.2
1	A	255	ALA	2.2
2	B	297	VAL	2.2
1	A	172	HIS	2.2
2	B	143	GLY	2.2
1	A	138	GLY	2.2
1	A	294	PHE	2.1
2	B	169	THR	2.1
2	B	170	GLY	2.1
1	A	249	ILE	2.1
2	B	167	LEU	2.1
2	B	301	SER	2.1
2	B	4	ARG	2.1
1	A	139	SER	2.1
1	A	291	PHE	2.0
2	B	53	ASN	2.0
1	A	73	VAL	2.0
2	B	26	THR	2.0
2	B	249	ILE	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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.