Full wwPDB X-ray Structure Validation Report ⓘ

Sep 24, 2024 – 08:11 PM JST

PDB ID	:	8WSK
Title	:	Crystal structure of SARS-Cov-2 main protease, pH=8.5
Authors	:	Jiang, H.H.; Zhou, X.L.; Zhang, J.; Li, J.
Deposited on	:	2023-10-17
Resolution	:	1.88 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:

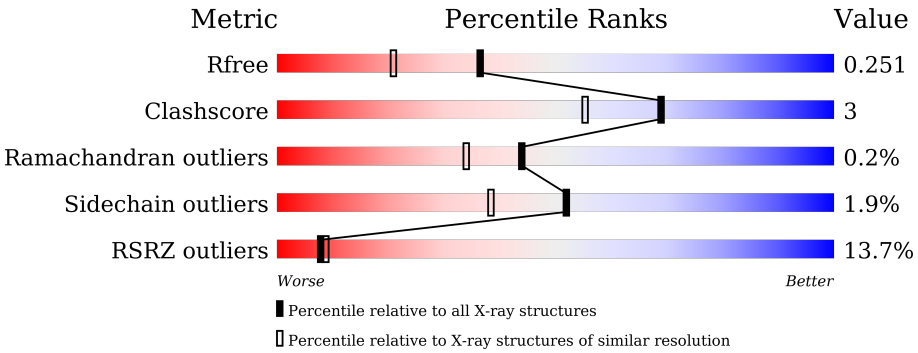
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.002 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.38.2

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.88 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	1090 (1.88-1.88)
Clashscore	180529	1144 (1.88-1.88)
Ramachandran outliers	177936	1135 (1.88-1.88)
Sidechain outliers	177891	1135 (1.88-1.88)
RSRZ outliers	164620	1090 (1.88-1.88)

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	296	12% 91% 8%
1	B	296	15% 89% 10%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4575 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	293	Total	C	N	O	S	0	0	0
			2219	1412	376	410	21			
1	B	292	Total	C	N	O	S	0	0	0
			2189	1389	370	409	21			

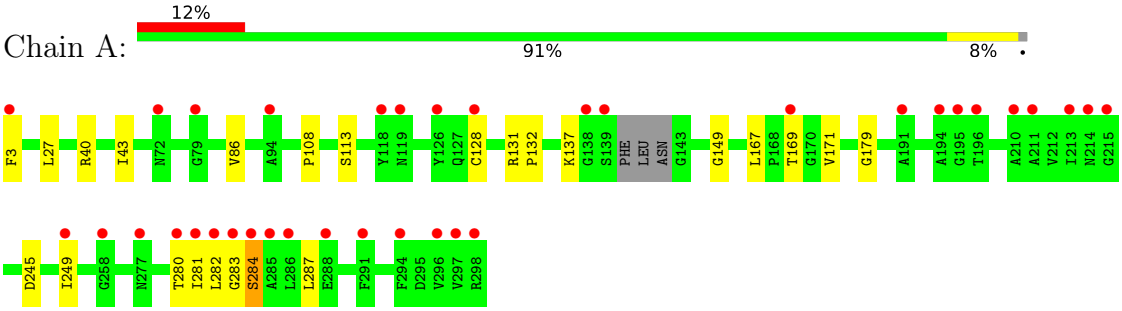
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	110	Total	O	0	0
			110	110		
2	B	57	Total	O	0	0
			57	57		

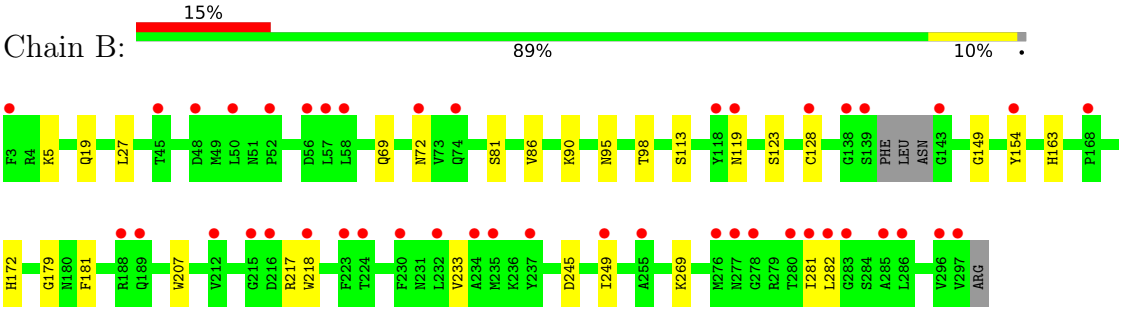
3 Residue-property plots (i)

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 1: 3C-like proteinase nsp5



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	68.01Å 102.90Å 103.47Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.48 – 1.88 36.48 – 1.89	Depositor EDS
% Data completeness (in resolution range)	98.9 (36.48-1.88) 98.9 (36.48-1.89)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.26 (at 1.88Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.222 , 0.251 0.222 , 0.251	Depositor DCC
R_{free} test set	2903 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å^2)	31.1	Xtriage
Anisotropy	0.674	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 47.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.016 for -h,l,k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4575	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.01% 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	0.46	1/2268 (0.0%)	0.63	1/3087 (0.0%)
1	B	0.40	1/2237 (0.0%)	0.58	1/3049 (0.0%)
All	All	0.43	2/4505 (0.0%)	0.61	2/6136 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	128	CYS	CB-SG	-9.94	1.65	1.82
1	B	128	CYS	CB-SG	-7.12	1.70	1.82

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	282	LEU	CA-CB-CG	7.21	131.89	115.30
1	B	128	CYS	CB-CA-C	-5.20	99.99	110.40

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	2219	0	2151	12	0
1	B	2189	0	2097	17	0
2	A	110	0	0	0	0
2	B	57	0	0	0	0
All	All	4575	0	4248	29	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 (29) 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:281:ILE:H	1:A:284:SER:H	1.39	0.71
1:B:218:TRP:CZ2	1:B:281:ILE:HD11	2.30	0.67
1:A:280:THR:HB	1:A:284:SER:N	2.13	0.64
1:A:86:VAL:HG13	1:A:179:GLY:HA2	1.81	0.63
1:B:282:LEU:O	1:B:282:LEU:HD12	1.98	0.62
1:A:245:ASP:O	1:A:249:ILE:HG12	1.99	0.62
1:A:169:THR:HG23	1:A:171:VAL:HG22	1.81	0.62
1:B:218:TRP:CZ2	1:B:281:ILE:CD1	2.87	0.57
1:A:281:ILE:N	1:A:284:SER:H	2.02	0.55
1:B:233:VAL:HG21	1:B:269:LYS:HE2	1.90	0.54
1:B:207:TRP:HE1	1:B:282:LEU:HD11	1.75	0.50
1:B:233:VAL:HG11	1:B:269:LYS:HG3	1.93	0.50
1:B:19:GLN:OE1	1:B:119:ASN:HB3	2.12	0.50
1:B:245:ASP:O	1:B:249:ILE:HG12	2.12	0.50
1:B:69:GLN:NE2	1:B:72:ASN:O	2.47	0.48
1:A:108:PRO:HB3	1:A:132:PRO:HA	1.98	0.44
1:A:287:LEU:HD23	1:A:287:LEU:HA	1.76	0.44
1:B:217:ARG:HE	1:B:217:ARG:HB3	1.66	0.44
1:B:207:TRP:NE1	1:B:282:LEU:HD11	2.33	0.44
1:B:86:VAL:HG13	1:B:179:GLY:HA2	1.98	0.43
1:A:40:ARG:O	1:A:43:ILE:HG12	2.18	0.43
1:B:5:LYS:HE3	1:B:5:LYS:HB2	1.87	0.43
1:B:95:ASN:HB3	1:B:98:THR:OG1	2.19	0.42
1:B:207:TRP:NE1	1:B:282:LEU:HD21	2.35	0.42
1:A:131:ARG:HD3	1:A:137:LYS:HE3	2.01	0.42
1:B:113:SER:O	1:B:149:GLY:HA2	2.20	0.41
1:A:113:SER:O	1:A:149:GLY:HA2	2.20	0.41
1:A:280:THR:HB	1:A:283:GLY:C	2.41	0.41
1:B:163:HIS:CD2	1:B:172:HIS:HB3	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	289/296 (98%)	281 (97%)	8 (3%)	0	100 100
1	B	288/296 (97%)	280 (97%)	7 (2%)	1 (0%)	37 26
All	All	577/592 (98%)	561 (97%)	15 (3%)	1 (0%)	44 34

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
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	239/255 (94%)	235 (98%)	4 (2%)	56 43
1	B	234/255 (92%)	229 (98%)	5 (2%)	48 34
All	All	473/510 (93%)	464 (98%)	9 (2%)	52 38

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	3	PHE
1	A	27	LEU
1	A	167	LEU
1	A	284	SER
1	B	27	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	81	SER
1	B	90	LYS
1	B	123	SER
1	B	181	PHE

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

Mol	Chain	Res	Type
1	B	273	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 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, 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	293/296 (98%)	0.65	36 (12%)	9	10	20, 30, 50, 61	0
1	B	292/296 (98%)	0.95	44 (15%)	6	7	21, 39, 60, 68	0
All	All	585/592 (98%)	0.80	80 (13%)	8	9	20, 34, 56, 68	0

All (80) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	282	LEU	5.6
1	A	283	GLY	5.1
1	A	296	VAL	5.0
1	B	118	TYR	4.9
1	B	215	GLY	4.9
1	A	118	TYR	4.4
1	A	282	LEU	4.0
1	B	223	PHE	4.0
1	B	139	SER	3.9
1	B	297	VAL	3.8
1	A	297	VAL	3.8
1	A	210	ALA	3.8
1	A	215	GLY	3.8
1	B	285	ALA	3.6
1	A	213	ILE	3.5
1	B	283	GLY	3.5
1	B	255	ALA	3.5
1	A	214	ASN	3.4
1	A	286	LEU	3.4
1	A	191	ALA	3.3
1	A	3	PHE	3.3
1	B	72	ASN	3.3
1	B	168	PRO	3.2
1	B	235	MET	3.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	281	ILE	3.2
1	A	280	THR	3.1
1	B	280	THR	3.1
1	B	276	MET	3.1
1	A	138	GLY	3.0
1	B	278	GLY	2.9
1	B	128	CYS	2.8
1	A	258	GLY	2.8
1	B	277	ASN	2.8
1	B	286	LEU	2.7
1	A	291	PHE	2.7
1	A	298	ARG	2.7
1	B	189	GLN	2.7
1	B	45	THR	2.7
1	A	281	ILE	2.7
1	A	139	SER	2.7
1	B	143	GLY	2.7
1	B	50	LEU	2.7
1	A	284	SER	2.6
1	B	224	THR	2.6
1	B	230	PHE	2.6
1	B	56	ASP	2.6
1	A	277	ASN	2.6
1	B	138	GLY	2.5
1	B	48	ASP	2.5
1	B	216	ASP	2.5
1	B	232	LEU	2.5
1	A	128	CYS	2.4
1	B	57	LEU	2.4
1	B	234	ALA	2.4
1	B	154	TYR	2.4
1	B	119	ASN	2.4
1	A	169	THR	2.4
1	A	196	THR	2.4
1	B	237	TYR	2.4
1	B	249	ILE	2.3
1	A	94	ALA	2.3
1	A	195	GLY	2.3
1	A	119	ASN	2.3
1	A	211	ALA	2.2
1	B	3	PHE	2.2
1	B	58	LEU	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	72	ASN	2.2
1	A	79	GLY	2.2
1	A	294	PHE	2.2
1	B	218	TRP	2.2
1	B	188	ARG	2.1
1	A	288	GLU	2.1
1	A	194	ALA	2.1
1	A	285	ALA	2.1
1	B	74	GLN	2.1
1	B	296	VAL	2.1
1	B	52	PRO	2.1
1	A	126	TYR	2.1
1	A	249	ILE	2.0
1	B	212	VAL	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.