



Full wwPDB X-ray Structure Validation Report ⓘ

Nov 23, 2023 – 02:38 AM JST

PDB ID : 8GQC
Title : Crystal structure of the SARS-unique domain (SUD) of SARS-CoV-2 (1.35 angstrom resolution)
Authors : Qin, B.; Li, Z.; Aumonier, S.; Wang, M.; Cui, S.
Deposited on : 2022-08-30
Resolution : 1.35 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 : 2.36
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) : 2.36

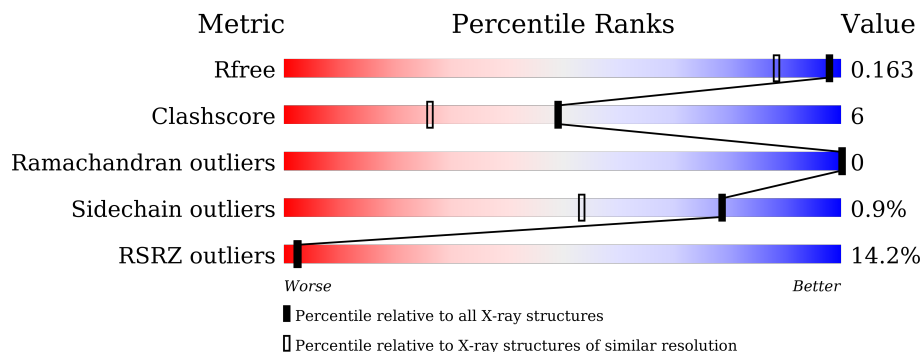
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.35 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	1509 (1.38-1.34)
Clashscore	141614	1551 (1.38-1.34)
Ramachandran outliers	138981	1530 (1.38-1.34)
Sidechain outliers	138945	1530 (1.38-1.34)
RSRZ outliers	127900	1487 (1.38-1.34)

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	264	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4203 atoms, of which 1956 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 Papain-like protease nsp3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	246	3889	1236	1956	308	379	10	0	7	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	516	CYS	LEU	engineered mutation	UNP P0DTC1
A	647	CYS	TYR	engineered mutation	UNP P0DTC1

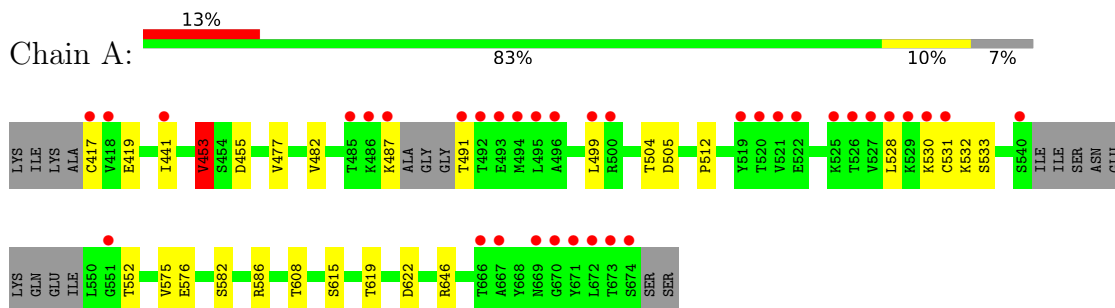
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	314	Total	O	0	0
			314	314		

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: Papain-like protease nsp3



4 Data and refinement statistics i

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	87.01Å 87.01Å 76.77Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	37.85 – 1.35 37.85 – 1.35	Depositor EDS
% Data completeness (in resolution range)	75.4 (37.85-1.35) 75.4 (37.85-1.35)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.97 (at 1.35Å)	Xtrriage
Refinement program	PHENIX 1.20_4459	Depositor
R, R_{free}	0.140 , 0.162 0.140 , 0.163	Depositor DCC
R_{free} test set	2713 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å ²)	22.4	Xtrriage
Anisotropy	0.035	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 51.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.020 for -h,-k,l	Xtrriage
F_o, F_c correlation	0.98	EDS
Total number of atoms	4203	wwPDB-VP
Average B, all atoms (Å ²)	33.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.88% 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	0.52	1/1990 (0.1%)	0.68	1/2708 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	419	GLU	CD-OE1	-5.41	1.19	1.25

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	453	VAL	CG1-CB-CG2	5.16	119.16	110.90

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	1933	1956	1977	23	0
2	A	314	0	0	10	0
All	All	2247	1956	1977	23	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 6.

All (23) 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:528:LEU:O	2:A:701:HOH:O	1.86	0.91
1:A:622[B]:ASP:OD2	2:A:702:HOH:O	2.00	0.80
1:A:455:ASP:OD1	2:A:703:HOH:O	2.02	0.77
1:A:615:SER:O	1:A:619[A]:THR:HG23	1.83	0.76
1:A:530:LYS:O	1:A:531:CYS:SG	2.44	0.75
1:A:552:THR:OG1	2:A:704:HOH:O	2.05	0.73
1:A:582:SER:OG	1:A:586:ARG:NH1	2.22	0.72
1:A:417:CYS:SG	2:A:701:HOH:O	2.48	0.71
1:A:487:LYS:HD2	2:A:711:HOH:O	1.98	0.64
1:A:619[A]:THR:HG21	2:A:816:HOH:O	1.98	0.64
1:A:499:LEU:HD21	1:A:512:PRO:HG3	1.85	0.58
1:A:530:LYS:C	1:A:531:CYS:SG	2.87	0.53
1:A:417:CYS:SG	1:A:531:CYS:HB3	2.51	0.51
1:A:505:ASP:O	1:A:533:SER:HB3	2.14	0.48
1:A:417:CYS:SG	1:A:531:CYS:HA	2.53	0.48
1:A:575:VAL:CG2	1:A:608[A]:THR:HG22	2.44	0.47
1:A:491:THR:N	2:A:711:HOH:O	2.47	0.47
1:A:441:ILE:HD13	1:A:482:VAL:HG11	1.97	0.46
1:A:646:ARG:NH1	2:A:707:HOH:O	2.27	0.46
1:A:441:ILE:HG12	2:A:908:HOH:O	2.17	0.44
1:A:576[B]:GLU:OE1	1:A:608[B]:THR:HA	2.19	0.43
1:A:504:THR:HG23	1:A:532:LYS:O	2.20	0.42
1:A:453:VAL:HG13	1:A:453:VAL:O	2.20	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	247/264 (94%)	243 (98%)	4 (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	223/230 (97%)	220 (99%)	3 (1%)	69 37

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	453	VAL
1	A	477[A]	VAL
1	A	477[B]	VAL

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](#)

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

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	246/264 (93%)	0.59	35 (14%) 2 2	17, 26, 53, 77	0

All (35) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	531	CYS	8.9
1	A	673	THR	8.6
1	A	529	LYS	7.7
1	A	672	LEU	7.5
1	A	491	THR	7.4
1	A	674	SER	6.8
1	A	530	LYS	6.0
1	A	487	LYS	6.0
1	A	492	THR	5.4
1	A	528	LEU	4.9
1	A	671	TYR	4.2
1	A	527	VAL	4.1
1	A	493	GLU	4.0
1	A	417	CYS	3.8
1	A	494	MET	3.5
1	A	526	THR	3.4
1	A	486	LYS	3.2
1	A	495	LEU	3.1
1	A	500	ARG	2.9
1	A	670	GLY	2.8
1	A	521	VAL	2.7
1	A	551	GLY	2.7
1	A	485	THR	2.6
1	A	499	LEU	2.6
1	A	441	ILE	2.6
1	A	519	TYR	2.6
1	A	666	THR	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	540	SER	2.4
1	A	418	VAL	2.4
1	A	525	LYS	2.3
1	A	496	ALA	2.3
1	A	522	GLU	2.2
1	A	520	THR	2.2
1	A	667	ALA	2.2
1	A	669	ASN	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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.