wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 11:14 pm BST

PDB ID	:	2GT7
Title	:	Crystal structure of SARS coronavirus main peptidase at pH 6.0 in the space group P21
Authors	:	Lee, T.-W.; Cherney, M.M.; Huitema, C.; Liu, J.; James, K.E.; Powers, J.C.; Eltis, L.D.; James, M.N.G.
Deposited on	:	2006-04-27
Resolution	:	1.82 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

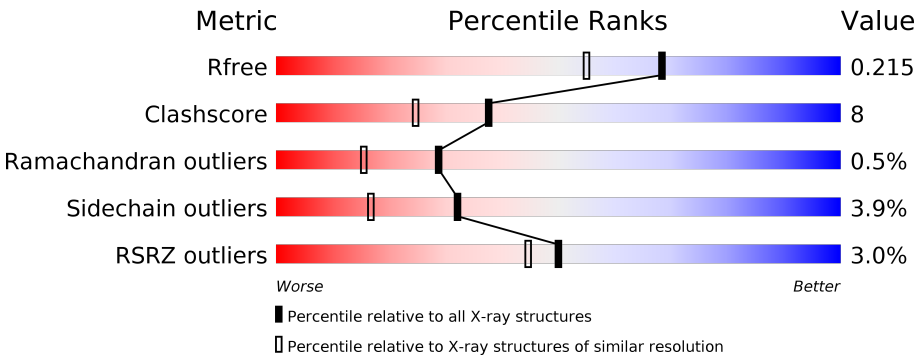
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.82 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	7484 (1.84-1.80)
Clashscore	141614	8401 (1.84-1.80)
Ramachandran outliers	138981	8290 (1.84-1.80)
Sidechain outliers	138945	8290 (1.84-1.80)
RSRZ outliers	127900	7371 (1.84-1.80)

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 on 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	3% 83% 13% • •
1	B	306	3% 84% 11% • •

2 Entry composition (i)

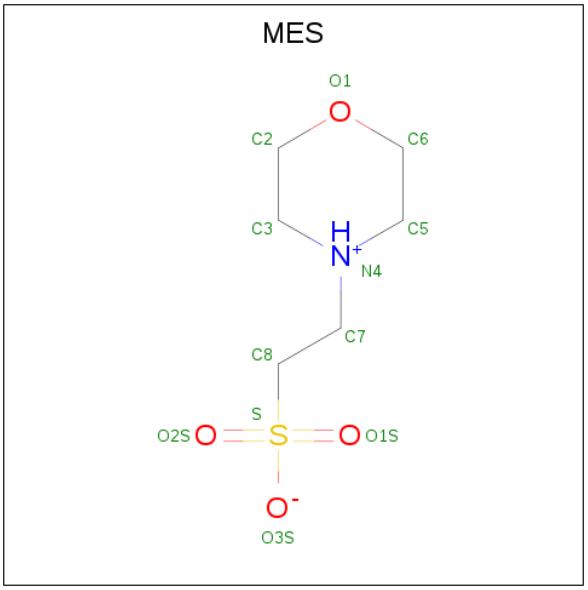
There are 3 unique types of molecules in this entry. The entry contains 5354 atoms, of which 0 are hydrogens and 0 are deuteriums.

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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	300	2374	1504	404	442	24	0	12	0
1	B	302	2377	1502	406	444	25	0	11	0

- Molecule 2 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	B	1	12	6	1	4	1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	292	292	292	0	0

Continued on next page...

Continued from previous page...

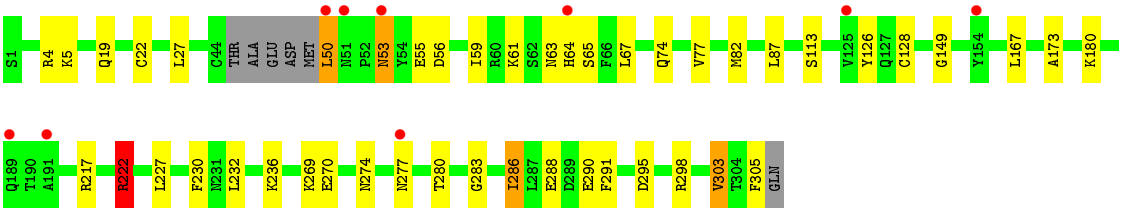
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	B	299	299	299	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

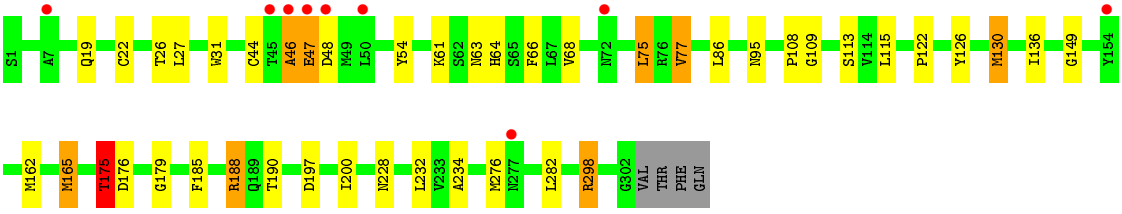
- Molecule 1: 3C-like proteinase

Chain A:



- Molecule 1: 3C-like proteinase

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	52.39Å 96.19Å 67.91Å 90.00° 102.91° 90.00°	Depositor
Resolution (Å)	40.00 – 1.82 41.25 – 1.82	Depositor EDS
% Data completeness (in resolution range)	96.4 (40.00-1.82) 96.3 (41.25-1.82)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.68 (at 1.82Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.170 , 0.217 0.169 , 0.215	Depositor DCC
R_{free} test set	2883 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	21.6	Xtriage
Anisotropy	0.306	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 42.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5354	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.51% 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 (i)

5.1 Standard geometry (i)

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

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	0.91	2/2466 (0.1%)	0.88	4/3349 (0.1%)
1	B	0.89	0/2474	0.89	6/3360 (0.2%)
All	All	0.90	2/4940 (0.0%)	0.89	10/6709 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	222	ARG	CB-CG	-7.01	1.33	1.52
1	A	290	GLU	CG-CD	5.74	1.60	1.51

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	217	ARG	NE-CZ-NH1	8.14	124.37	120.30
1	B	298	ARG	NE-CZ-NH1	7.97	124.29	120.30
1	A	222	ARG	N-CA-CB	-7.18	97.67	110.60
1	A	222	ARG	NE-CZ-NH2	-7.13	116.73	120.30
1	A	217	ARG	NE-CZ-NH2	-6.56	117.02	120.30

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	2374	0	2333	38	0
1	B	2377	0	2335	34	0
2	B	12	0	12	3	0
3	A	292	0	0	8	0
3	B	299	0	0	5	0
All	All	5354	0	4680	73	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 8.

The worst 5 of 73 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:22:CYS:SG	1:A:61:LYS:NZ	2.24	1.11
1:A:77[A]:VAL:HG12	3:A:385:HOH:O	1.55	1.05
1:B:22:CYS:SG	1:B:61:LYS:HE3	2.07	0.95
1:B:175:THR:HG23	1:B:176:ASP:O	1.71	0.89
1:B:19:GLN:HE21	1:B:26:THR:HG21	1.40	0.87

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	308/306 (101%)	302 (98%)	6 (2%)	0	100 100
1	B	311/306 (102%)	302 (97%)	6 (2%)	3 (1%)	15 5
All	All	619/612 (101%)	604 (98%)	12 (2%)	3 (0%)	29 15

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	47	GLU
1	B	46	ALA
1	B	48	ASP

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	270/263 (103%)	258 (96%)	12 (4%)	28 13
1	B	270/263 (103%)	258 (96%)	12 (4%)	28 13
All	All	540/526 (103%)	516 (96%)	24 (4%)	32 13

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	286	ILE
1	B	75	LEU
1	B	232	LEU
1	A	303	VAL
1	B	27	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	277	ASN
1	B	228	ASN
1	B	19	GLN
1	A	256	GLN
1	B	119	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

1 ligand is 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	MES	B	601	-	12,12,12	2.19	1 (8%)	14,16,16	3.29	8 (57%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	MES	B	601	-	-	4/6/14/14	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	601	MES	C8-S	-7.38	1.67	1.77

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	601	MES	C5-N4-C3	8.09	127.05	108.83
2	B	601	MES	O2S-S-C8	4.82	112.72	106.92
2	B	601	MES	O3S-S-O2S	-3.44	102.88	111.27

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	601	MES	C2-C3-N4	3.15	114.88	110.10
2	B	601	MES	C6-O1-C2	2.99	119.88	109.89

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	601	MES	C8-C7-N4-C3
2	B	601	MES	C7-C8-S-O3S
2	B	601	MES	C7-C8-S-O1S
2	B	601	MES	C7-C8-S-O2S

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	601	MES	3	0

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 (i)

6.1 Protein, DNA and RNA chains (i)

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	300/306 (98%)	0.13	9 (3%)	50	44	20, 28, 42, 49	0
1	B	302/306 (98%)	0.11	9 (2%)	50	44	20, 29, 38, 50	0
All	All	602/612 (98%)	0.12	18 (2%)	50	44	20, 29, 40, 50	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	46	ALA	5.5
1	A	277	ASN	4.2
1	B	154	TYR	4.1
1	B	72	ASN	3.5
1	A	189	GLN	3.5

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 carbohydrates 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, 95th 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	MES	B	601	12/12	0.78	0.18	16,21,33,34	12

6.5 Other polymers (i)

There are no such residues in this entry.