



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 15, 2016 – 05:13 PM EDT

PDB ID : 5CFC
Title : Crystal Structure of Human Coronavirus SAFV-3
Authors : Mullapudi, E.; Plevka, P.
Deposited on : 2015-07-08
Resolution : 2.50 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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

MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

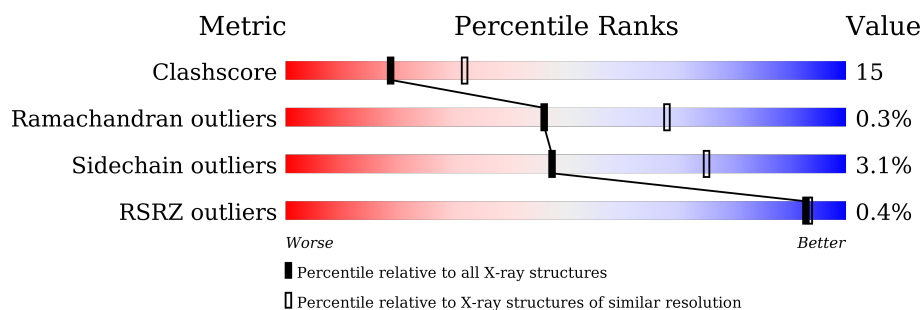
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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))
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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. 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	252	
2	B	232	
3	C	258	
4	D	24	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6352 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 Genome polyprotein,VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	252	Total	C	N	O	S	0	0	0
			1996	1273	332	384	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	36	PHE	VAL	conflict	UNP C0MHL9

- Molecule 2 is a protein called Genome polyprotein, VP3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	232	Total	C	N	O	S	0	0	0
			1804	1166	285	343	10			

- Molecule 3 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	258	Total	C	N	O	S	0	0	0
			2023	1274	364	378	7			

- Molecule 4 is a protein called VP4.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	D	24	Total	C	N	O	0	0	0
			189	115	32	42			

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	126	Total	O	0	0
			126	126		

Continued on next page...

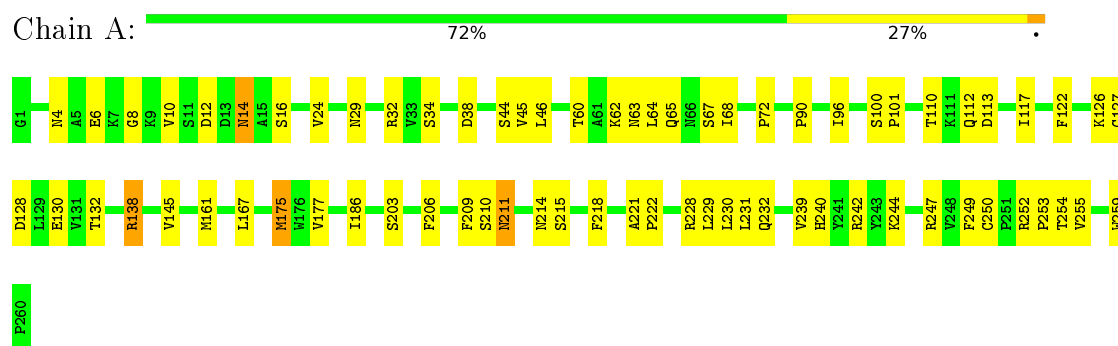
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	103	Total 103	O 103	0	0
5	C	103	Total 103	O 103	0	0
5	D	8	Total 8	O 8	0	0

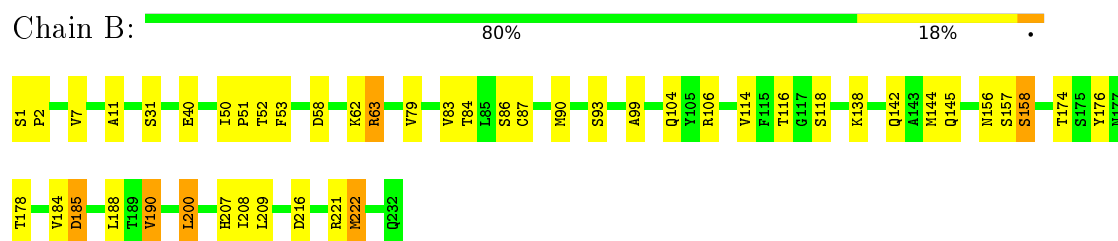
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

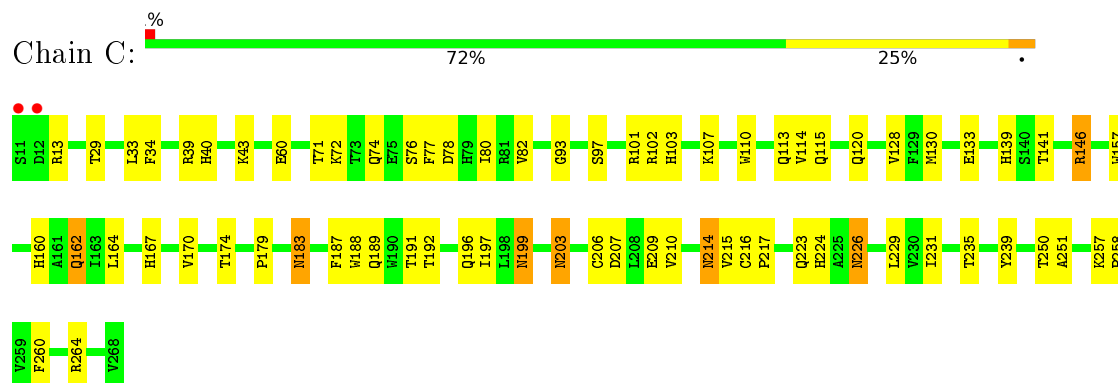
- Molecule 1: Genome polyprotein, VP1



- Molecule 2: Genome polyprotein, VP3

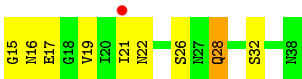


- Molecule 3: Genome polyprotein



- Molecule 4: VP4





4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α , β , γ	300.50Å 300.50Å 722.10Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	70.00 – 2.50 69.36 – 2.50	Depositor EDS
% Data completeness (in resolution range)	62.4 (70.00-2.50) 62.4 (69.36-2.50)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.36 (at 2.51Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.220 , (Not available) 0.225 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å ²)	27.5	Xtriage
Anisotropy	0.869	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.67 , 13.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.24$	Xtriage
Estimated twinning fraction	0.064 for -h,-k,l	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	6352	wwPDB-VP
Average B, all atoms (Å ²)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.09% 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	0.32	0/2055	0.62	0/2808
2	B	0.34	0/1860	0.63	0/2553
3	C	0.30	0/2078	0.63	0/2843
4	D	0.39	0/191	0.51	0/259
All	All	0.32	0/6184	0.62	0/8463

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	1996	0	1938	70	0
2	B	1804	0	1759	61	0
3	C	2023	0	1961	78	0
4	D	189	0	167	8	0
5	A	126	0	0	6	0
5	B	103	0	0	6	0
5	C	103	0	0	2	0
5	D	8	0	0	0	0
All	All	6352	0	5825	182	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 15.

The worst 5 of 182 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:46:LEU:HD13	1:A:68:ILE:HG12	1.34	1.05
3:C:203:ASN:HD22	3:C:203:ASN:H	1.10	0.94
2:B:138:LYS:H	2:B:142:GLN:NE2	1.71	0.87
3:C:226:ASN:HD22	3:C:226:ASN:H	1.18	0.87
1:A:29:ASN:HB2	1:A:32:ARG:HH11	1.42	0.85

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	250/252 (99%)	237 (95%)	12 (5%)	1 (0%)	39	61
2	B	230/232 (99%)	220 (96%)	9 (4%)	1 (0%)	39	61
3	C	256/258 (99%)	244 (95%)	12 (5%)	0	100	100
4	D	22/24 (92%)	20 (91%)	2 (9%)	0	100	100
All	All	758/766 (99%)	721 (95%)	35 (5%)	2 (0%)	46	68

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	250	CYS
2	B	222	MET

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	233/233 (100%)	229 (98%)	4 (2%)	68	89
2	B	206/206 (100%)	199 (97%)	7 (3%)	44	72
3	C	222/222 (100%)	214 (96%)	8 (4%)	42	69
4	D	21/21 (100%)	19 (90%)	2 (10%)	11	20
All	All	682/682 (100%)	661 (97%)	21 (3%)	47	75

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	190	VAL
3	C	141	THR
3	C	214	ASN
2	B	185	ASP
3	C	226	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	38	GLN
3	C	139	HIS
4	D	16	ASN
3	C	74	GLN
3	C	86	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates 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 [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	252/252 (100%)	-0.13	0 100 100	24, 31, 46, 55	0
2	B	232/232 (100%)	-0.09	0 100 100	22, 30, 42, 57	0
3	C	258/258 (100%)	-0.11	2 (0%) 87 89	24, 29, 52, 77	0
4	D	24/24 (100%)	0.37	1 (4%) 40 45	30, 44, 68, 70	0
All	All	766/766 (100%)	-0.10	3 (0%) 93 93	22, 30, 50, 77	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	11	SER	3.7
4	D	21	ILE	2.6
3	C	12	ASP	2.2

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.