



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 11:31 pm BST

PDB ID : 5MSF  
Title : MS2 PROTEIN CAPSID/RNA COMPLEX  
Authors : Rowsell, S.; Stonehouse, N.J.; Convery, M.A.; Adams, C.J.; Ellington, A.D.;  
Hirao, I.; Peabody, D.S.; Stockley, P.G.; Phillips, S.E.V.  
Deposited on : 1998-05-15  
Resolution : 2.80 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](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

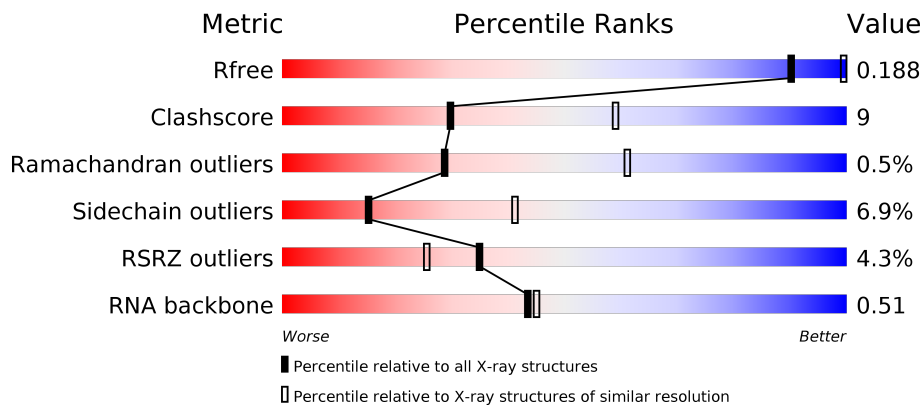
MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)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

# 1 Overall quality at a glance

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

The reported resolution of this entry is 2.80 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.



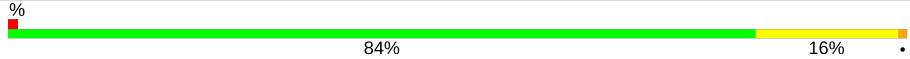
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)
RNA backbone	3102	1227 (3.10-2.50)

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  $\geq 3$ , 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	R	18	
1	S	18	
2	A	129	
2	B	129	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	C	129	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into two segments: a green segment on the left representing 84% and a yellow segment on the right representing 16%. A small red square is at the beginning of the bar, and a small black dot is at the end. The percentage values '84%' and '16%' are printed below the bar. A '%' symbol is located at the top left of the bar area.</p>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 3737 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 RNA chain called 5'-R(\*CP\*CP\*GP\*GP\*AP\*GP\*GP\*AP\*UP\*CP\*AP\*CP\*CP\*AP\*CP\*GP\*GP\*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	R	18	Total 386	C 173	N 75	O 121	P 17	0	0	0
1	S	16	Total 342	C 154	N 67	O 106	P 15	0	0	0

- Molecule 2 is a protein called MS2 PROTEIN CAPSID.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	129	Total 972	C 610	N 167	O 191	S 4	0	2	0
2	B	129	Total 968	C 607	N 166	O 191	S 4	0	1	0
2	C	129	Total 968	C 607	N 166	O 191	S 4	0	1	0

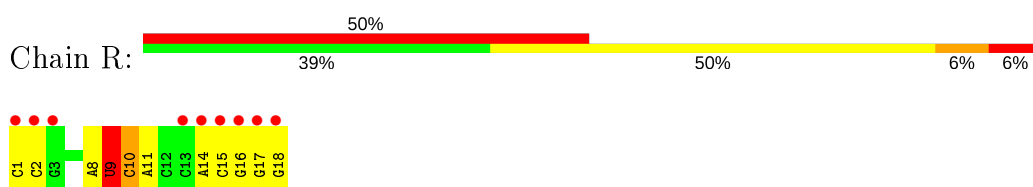
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	R	2	Total 2	O 2	0	0
3	A	30	Total 30	O 30	0	0
3	B	43	Total 43	O 43	0	0
3	C	26	Total 26	O 26	0	0

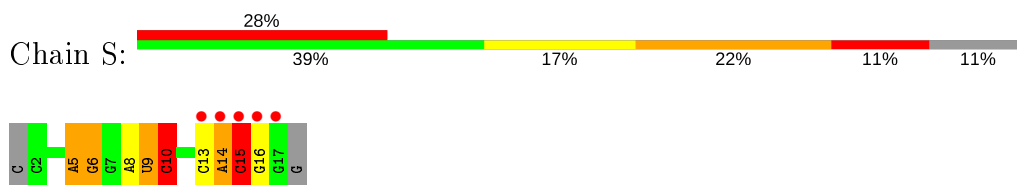
### 3 Residue-property plots [i](#)

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.

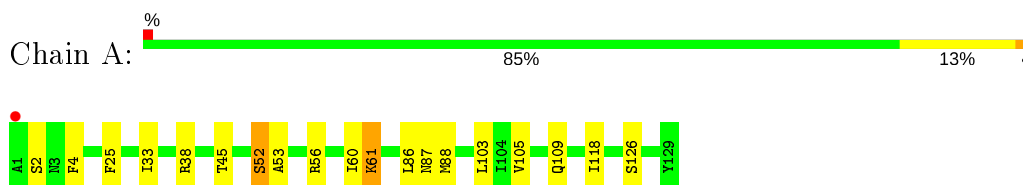
- Molecule 1: 5'-R(\*CP\*CP\*GP\*GP\*AP\*GP\*GP\*AP\*UP\*CP\*AP\*CP\*CP\*AP\*CP\*GP\*GP\*G)-3'



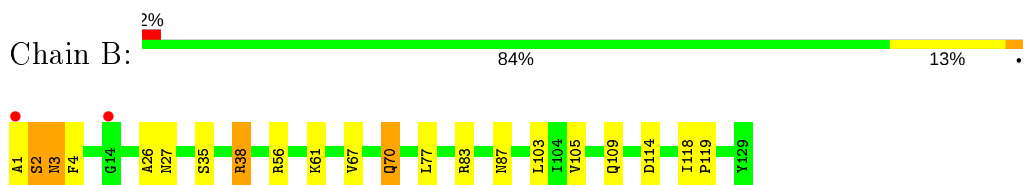
- Molecule 1: 5'-R(\*CP\*CP\*GP\*GP\*AP\*GP\*GP\*AP\*UP\*CP\*AP\*CP\*CP\*AP\*CP\*GP\*GP\*G)-3'



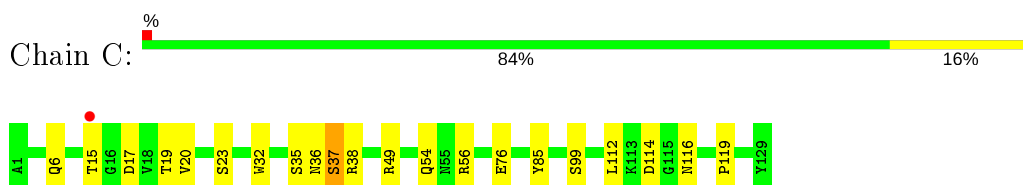
- Molecule 2: MS2 PROTEIN CAPSID



- Molecule 2: MS2 PROTEIN CAPSID



- Molecule 2: MS2 PROTEIN CAPSID



## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	287.30Å 287.30Å 651.60Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 2.80 29.99 – 2.80	Depositor EDS
% Data completeness (in resolution range)	68.3 (30.00-2.80) 68.4 (29.99-2.80)	Depositor EDS
$R_{merge}$	0.18	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.61 (at 2.80Å)	Xtrriage
Refinement program	X-PLOR 3.860	Depositor
R, $R_{free}$	0.188 , 0.197 0.186 , 0.188	Depositor DCC
$R_{free}$ test set	6927 reflections (4.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	35.4	Xtrriage
Anisotropy	0.228	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 53.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	3737	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.39% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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	R	0.38	0/432	0.86	2/673 (0.3%)
1	S	2.09	2/383 (0.5%)	1.59	7/596 (1.2%)
2	A	0.37	0/999	0.65	0/1359
2	B	0.37	0/990	0.68	0/1348
2	C	0.36	0/990	0.63	0/1348
All	All	0.75	2/3794 (0.1%)	0.84	9/5324 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	S	14	A	O3'-P	38.66	2.07	1.61
1	S	15	C	O3'-P	8.60	1.71	1.61

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	S	15	C	O5'-P-OP1	-19.87	86.86	110.70
1	S	15	C	OP1-P-OP2	16.95	145.03	119.60
1	S	15	C	P-O3'-C3'	-14.18	102.68	119.70
1	S	5	A	O3'-P-O5'	8.44	120.04	104.00
1	S	14	A	P-O3'-C3'	-6.90	111.42	119.70

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	R	386	0	199	9	0
1	S	342	0	176	22	0
2	A	972	0	975	13	0
2	B	968	0	966	13	0
2	C	968	0	966	11	0
3	A	30	0	0	0	0
3	B	43	0	0	2	0
3	C	26	0	0	2	0
3	R	2	0	0	0	0
All	All	3737	0	3282	65	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 9.

The worst 5 of 65 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:S:15:C:O2'	1:S:16:G:H5'	1.18	1.27
1:S:14:A:O3'	1:S:15:C:P	2.07	1.10
1:S:15:C:O2'	1:S:16:G:C5'	2.05	1.03
1:S:15:C:H2'	1:S:16:G:O5'	1.68	0.93
1:S:14:A:H2'	1:S:15:C:H5'	1.52	0.92

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	
2	A	129/129 (100%)	124 (96%)	4 (3%)	1 (1%)	19	49
2	B	128/129 (99%)	122 (95%)	5 (4%)	1 (1%)	19	49
2	C	128/129 (99%)	119 (93%)	9 (7%)	0	100	100

*Continued on next page...*



Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	385/387 (100%)	365 (95%)	18 (5%)	2 (0%)	29 61

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	2	SER
2	A	2	SER

### 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
2	A	108/106 (102%)	101 (94%)	7 (6%)	17 44
2	B	107/106 (101%)	99 (92%)	8 (8%)	13 37
2	C	107/106 (101%)	99 (92%)	8 (8%)	13 37
All	All	322/318 (101%)	299 (93%)	23 (7%)	15 39

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	38	ARG
2	B	77	LEU
2	C	99	SER
2	B	70	GLN
2	B	103	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
2	A	55	ASN
2	B	3	ASN
2	B	36	ASN
2	B	70	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	R	17/18 (94%)	2 (11%)	0
1	S	14/18 (77%)	4 (28%)	0
All	All	31/36 (86%)	6 (19%)	0

5 of 6 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	R	9	U
1	R	10	C
1	S	6	G
1	S	9	U
1	S	10	C

There are no RNA pucker outliers to report.

### 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](#)

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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	S	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	S	14:A	O3'	15:C	P	2.07

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q < 0.9
1	R	18/18 (100%)	2.07	9 (50%) 0 0	33, 63, 99, 99	18 (100%)
1	S	16/18 (88%)	1.69	5 (31%) 0 0	32, 64, 95, 99	16 (100%)
2	A	129/129 (100%)	-0.86	1 (0%) 86 81	14, 26, 61, 92	0
2	B	129/129 (100%)	-0.94	2 (1%) 72 66	14, 22, 54, 93	0
2	C	129/129 (100%)	-0.83	1 (0%) 86 81	18, 28, 57, 100	0
All	All	421/423 (99%)	-0.65	18 (4%) 35 25	14, 26, 76, 100	34 (8%)

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	R	18	G	7.7
1	R	17	G	5.1
1	R	1	C	4.7
1	S	17	G	4.6
1	R	16	G	4.3

### 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

There are no such residues in this entry.