



wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Feb 12, 2018 – 02:08 PM EST

PDB ID : 5WK1
EMDB ID: : EMD-8854
Title : Structure of the major capsid protein and the capsid stabilizing protein of the marine siphovirus TW1
Authors : Wang, Z.; Rossmann, M.G.
Deposited on : 2017-07-24
Resolution : 3.60 Å(reported)

This is a wwPDB/EMDatabank EM Map/Model Validation Summary Report
for a publicly released PDB/EMDB entry.

We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030736

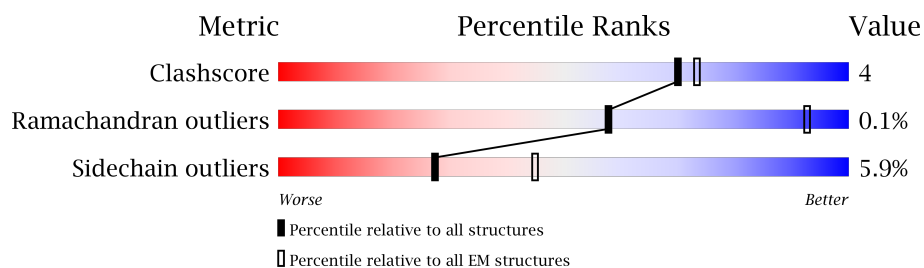
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.60 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)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	K	148	84% 16%
1	L	148	84% 16%
1	M	148	85% 15%
1	S	148	86% 14%
1	X	148	85% 15%
1	Y	148	84% 16%
1	Z	148	84% 16%
2	A	352	70% 13% 16%
2	B	352	72% 11% 16%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	C	352	 70% 13% 16%
2	D	352	 72% 11% • 16%
2	E	352	 72% 11% • 16%
2	F	352	 73% 11% 16%
2	G	352	 69% 14% • 16%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 23303 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Capsid Stabilizing Protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	X	148	Total	C	N	O	S	0	0
			1053	655	180	215	3		
1	L	148	Total	C	N	O	S	0	0
			1053	655	180	215	3		
1	Y	148	Total	C	N	O	S	0	0
			1053	655	180	215	3		
1	M	148	Total	C	N	O	S	0	0
			1053	655	180	215	3		
1	S	148	Total	C	N	O	S	0	0
			1053	655	180	215	3		
1	Z	148	Total	C	N	O	S	0	0
			1053	655	180	215	3		
1	K	148	Total	C	N	O	S	0	0
			1053	655	180	215	3		

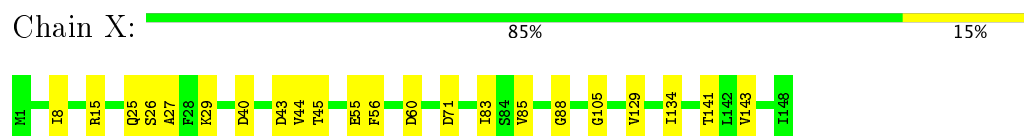
- Molecule 2 is a protein called Major Capsid Protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	A	294	Total	C	N	O	S	0	0
			2276	1427	389	445	15		
2	B	294	Total	C	N	O	S	0	0
			2276	1427	389	445	15		
2	C	294	Total	C	N	O	S	0	0
			2276	1427	389	445	15		
2	D	294	Total	C	N	O	S	0	0
			2276	1427	389	445	15		
2	E	294	Total	C	N	O	S	0	0
			2276	1427	389	445	15		
2	F	294	Total	C	N	O	S	0	0
			2276	1427	389	445	15		
2	G	294	Total	C	N	O	S	0	0
			2276	1427	389	445	15		

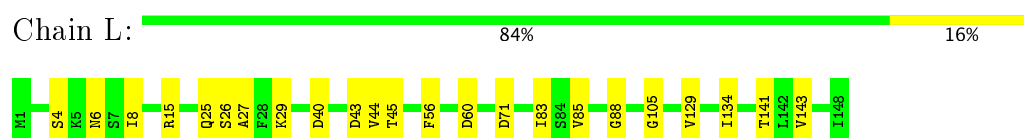
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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.

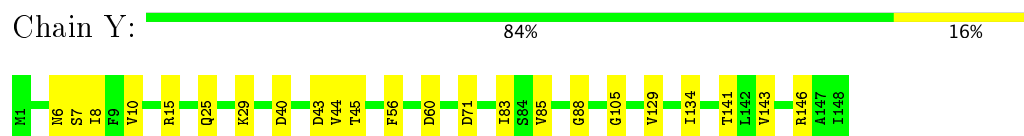
- Molecule 1: Capsid Stabilizing Protein



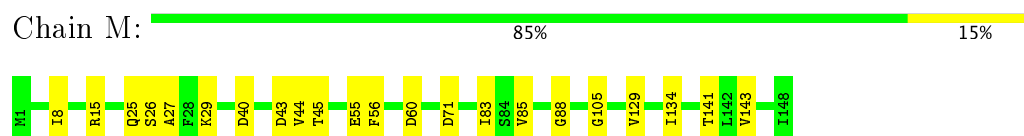
- Molecule 1: Capsid Stabilizing Protein



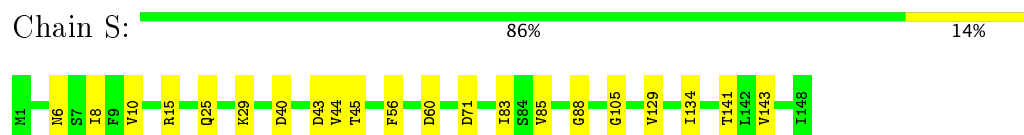
- Molecule 1: Capsid Stabilizing Protein



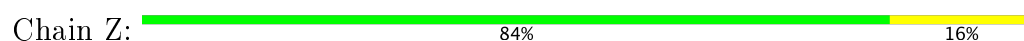
- Molecule 1: Capsid Stabilizing Protein



- Molecule 1: Capsid Stabilizing Protein



- Molecule 1: Capsid Stabilizing Protein





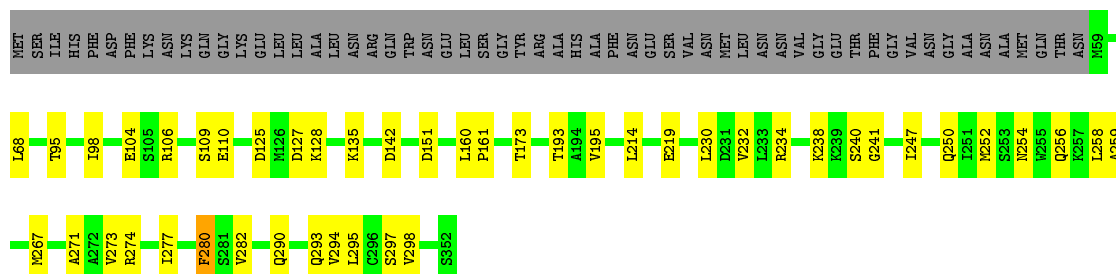
- Molecule 1: Capsid Stabilizing Protein

Chain K: 84% 16%



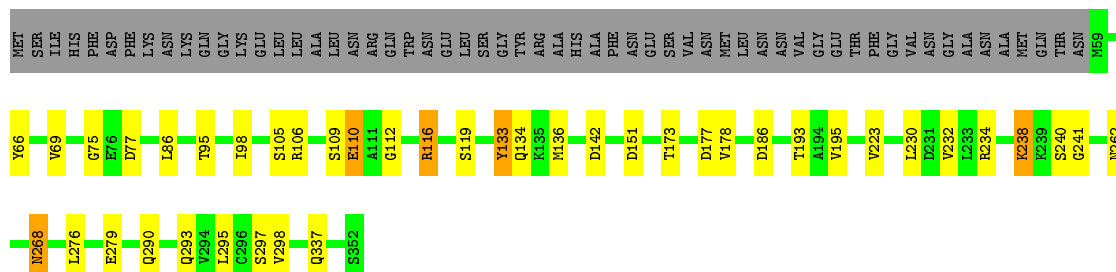
- Molecule 2: Major Capsid Protein

Chain A: 70% 13% 16%



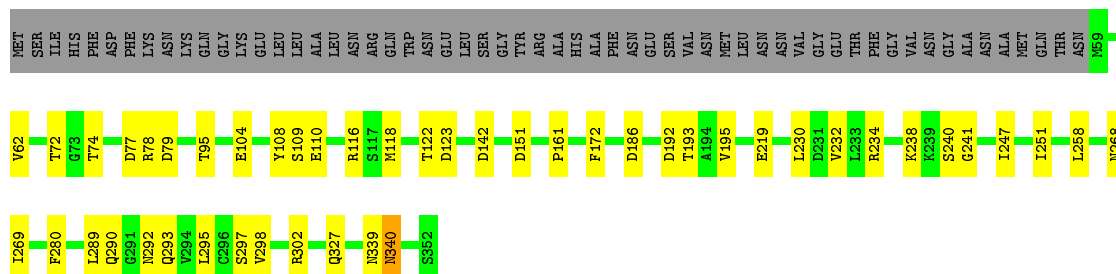
- Molecule 2: Major Capsid Protein

Chain B: 72% 11% 16%



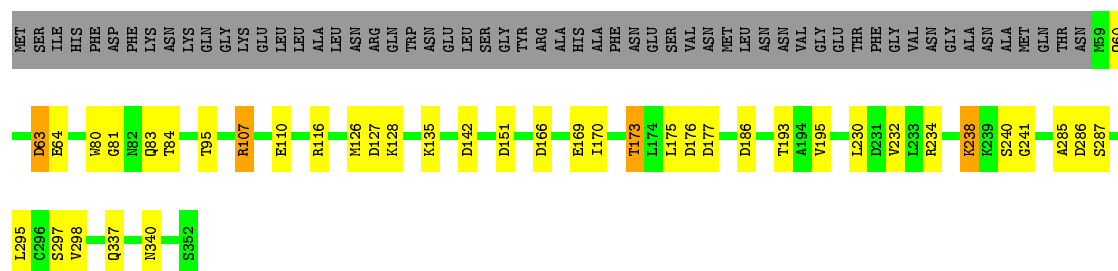
- Molecule 2: Major Capsid Protein

Chain C: 70% 13% 16%



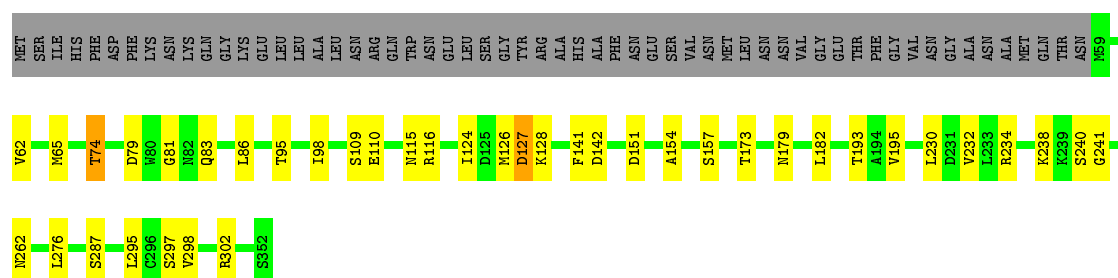
- Molecule 2: Major Capsid Protein

Chain D:  72% 11% 16%



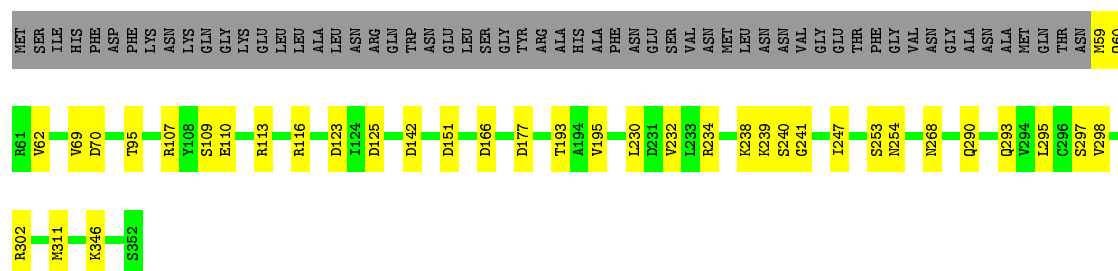
• Molecule 2: Major Capsid Protein

Chain E:  72% 11% 16%



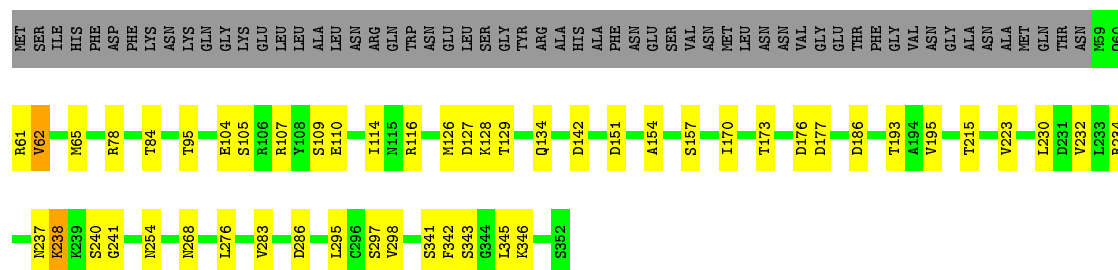
• Molecule 2: Major Capsid Protein

Chain F:  73% 11% 16%



• Molecule 2: Major Capsid Protein

Chain G:  69% 14% 16%



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of particles used	8663	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	20	Depositor
Minimum defocus (nm)	500	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	22500	Depositor
Image detector	DIRECT ELECTRON DE-16 (4k x 4k)	Depositor

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 >2	RMSZ	# Z >2
1	K	0.30	0/1064	0.48	0/1448
1	L	0.30	0/1064	0.48	0/1448
1	M	0.30	0/1064	0.48	0/1448
1	S	0.30	0/1064	0.48	0/1448
1	X	0.30	0/1064	0.48	0/1448
1	Y	0.30	0/1064	0.48	0/1448
1	Z	0.30	0/1064	0.48	0/1448
2	A	0.32	0/2315	0.50	0/3128
2	B	0.33	0/2315	0.49	0/3128
2	C	0.31	0/2315	0.49	0/3128
2	D	0.31	0/2315	0.50	0/3128
2	E	0.31	0/2315	0.48	0/3128
2	F	0.32	0/2315	0.47	0/3128
2	G	0.31	0/2315	0.48	0/3128
All	All	0.31	0/23653	0.48	0/32032

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	D	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	D	177	ASP	Peptide

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	K	1053	0	1057	13	0
1	L	1053	0	1057	12	0
1	M	1053	0	1057	11	0
1	S	1053	0	1057	11	0
1	X	1053	0	1057	11	0
1	Y	1053	0	1057	14	0
1	Z	1053	0	1057	12	0
2	A	2276	0	2237	18	0
2	B	2276	0	2237	20	0
2	C	2276	0	2237	19	0
2	D	2276	0	2237	22	0
2	E	2276	0	2237	19	0
2	F	2276	0	2237	17	0
2	G	2276	0	2237	23	0
All	All	23303	0	23058	193	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 4.

The worst 5 of 193 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)
2:A:104:GLU:HG2	2:A:135:LYS:HB3	1.76	0.68
2:E:241:GLY:HA3	2:E:302:ARG:HH22	1.61	0.65
1:Y:6:ASN:ND2	2:E:128:LYS:O	2.30	0.64
2:C:327:GLN:NE2	2:D:126:MET:SD	2.72	0.63
2:E:74:THR:HG21	2:F:107:ARG:HG3	1.81	0.63

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

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	K	146/148 (99%)	136 (93%)	10 (7%)	0	100	100
1	L	146/148 (99%)	136 (93%)	10 (7%)	0	100	100
1	M	146/148 (99%)	136 (93%)	10 (7%)	0	100	100
1	S	146/148 (99%)	136 (93%)	10 (7%)	0	100	100
1	X	146/148 (99%)	136 (93%)	10 (7%)	0	100	100
1	Y	146/148 (99%)	136 (93%)	10 (7%)	0	100	100
1	Z	146/148 (99%)	136 (93%)	10 (7%)	0	100	100
2	A	292/352 (83%)	267 (91%)	25 (9%)	0	100	100
2	B	292/352 (83%)	270 (92%)	22 (8%)	0	100	100
2	C	292/352 (83%)	269 (92%)	22 (8%)	1 (0%)	44	80
2	D	292/352 (83%)	270 (92%)	22 (8%)	0	100	100
2	E	292/352 (83%)	270 (92%)	22 (8%)	0	100	100
2	F	292/352 (83%)	271 (93%)	21 (7%)	0	100	100
2	G	292/352 (83%)	266 (91%)	25 (9%)	1 (0%)	44	80
All	All	3066/3500 (88%)	2835 (92%)	229 (8%)	2 (0%)	58	88

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	G	62	VAL
2	C	161	PRO

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

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	K	109/109 (100%)	104 (95%)	5 (5%)	31	69
1	L	109/109 (100%)	104 (95%)	5 (5%)	31	69
1	M	109/109 (100%)	104 (95%)	5 (5%)	31	69
1	S	109/109 (100%)	104 (95%)	5 (5%)	31	69
1	X	109/109 (100%)	104 (95%)	5 (5%)	31	69
1	Y	109/109 (100%)	104 (95%)	5 (5%)	31	69
1	Z	109/109 (100%)	104 (95%)	5 (5%)	31	69
2	A	244/292 (84%)	228 (93%)	16 (7%)	19	59
2	B	244/292 (84%)	227 (93%)	17 (7%)	18	56
2	C	244/292 (84%)	224 (92%)	20 (8%)	13	49
2	D	244/292 (84%)	231 (95%)	13 (5%)	26	65
2	E	244/292 (84%)	229 (94%)	15 (6%)	22	61
2	F	244/292 (84%)	231 (95%)	13 (5%)	26	65
2	G	244/292 (84%)	228 (93%)	16 (7%)	19	59
All	All	2471/2807 (88%)	2326 (94%)	145 (6%)	27	62

5 of 145 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	262	ASN
2	C	238	LYS
2	G	151	ASP
2	B	295	LEU
2	C	95	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	121	GLN
2	B	134	GLN
2	D	134	GLN
2	A	256	GLN
2	C	184	ASN

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 ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.