wwPDB/EMDataBank EM Map/Model Validation Summary Report (i)

Feb 19, 2018 – 05:10 am GMT

PDB ID	:	5WSN
EMDB ID:	:	EMD-6685
Title	:	Structure of Japanese encephalitis virus
Authors	:	Wang, X.; Zhu, L.; Li, S.; Yuan, S.; Qin, C.; Fry, E.E.; Stuart, I.D.; Rao, Z.
Deposited on	:	2016-12-07
Resolution	:	4.30 Å(reported)
Based on PDB ID	:	3J57

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
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

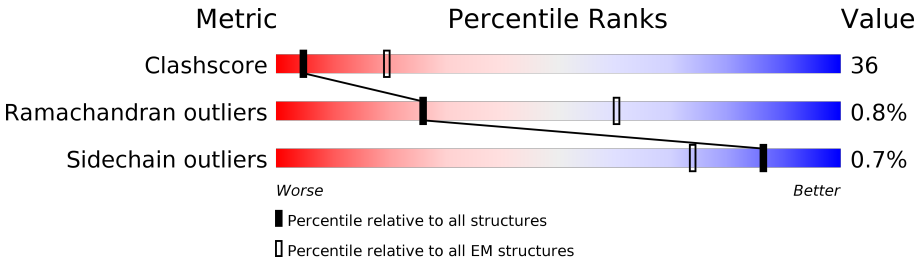
MolProbity	:	4.02b-467
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30686

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 4.30 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	136279	1886
Ramachandran outliers	132675	1663
Sidechain outliers	132484	1531

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	500	44% 53% .
1	C	500	49% 50% .
1	E	500	48% 51% .
2	B	74	51% 47% .
2	D	74	47% 51% .
2	F	74	38% 61% .

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 12981 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 E protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	500	Total 3753	C 2369	N 650	O 709	S 25	0	0
1	C	500	Total 3753	C 2369	N 650	O 709	S 25	0	0
1	E	500	Total 3753	C 2369	N 650	O 709	S 25	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	492	LEU	VAL	See Sequence details	UNP A1E4C6
C	492	LEU	VAL	See Sequence details	UNP A1E4C6
E	492	LEU	VAL	See Sequence details	UNP A1E4C6

- Molecule 2 is a protein called M protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	74	Total 574	C 370	N 98	O 104	S 2	0	0
2	D	74	Total 574	C 370	N 98	O 104	S 2	0	0
2	F	74	Total 574	C 370	N 98	O 104	S 2	0	0

There are 6 discrepancies between the modelled and reference sequences:

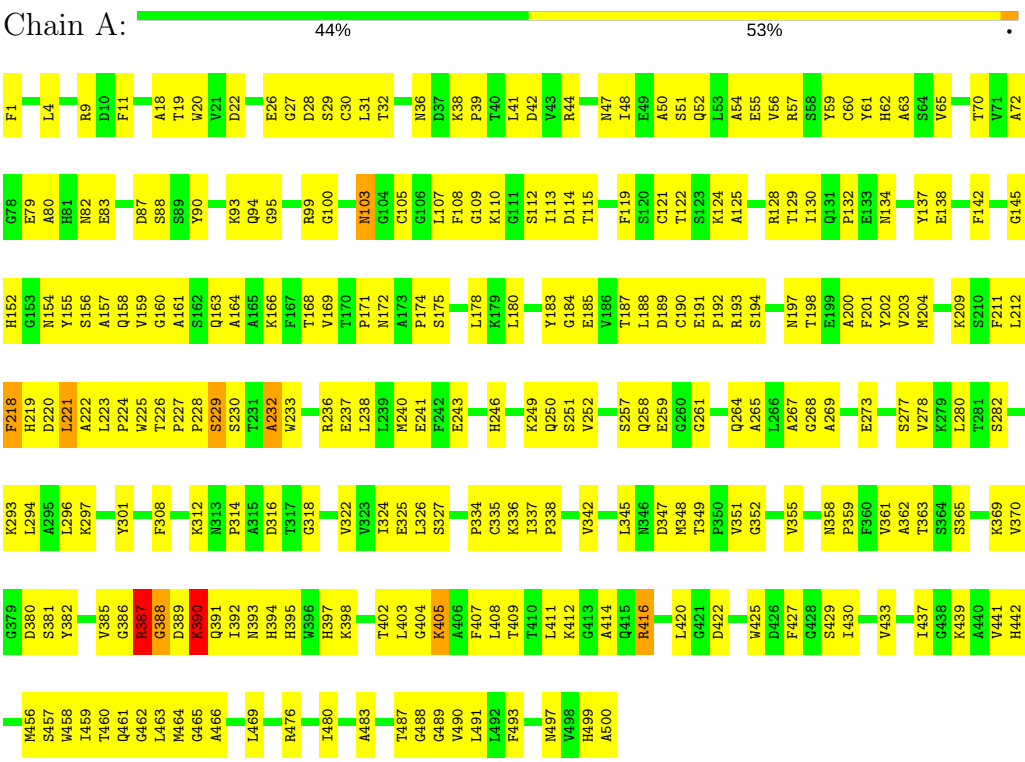
Chain	Residue	Modelled	Actual	Comment	Reference
B	16	THR	LYS	See Sequence details	UNP Q82863
B	18	THR	ALA	See Sequence details	UNP Q82863
D	16	THR	LYS	See Sequence details	UNP Q82863
D	18	THR	ALA	See Sequence details	UNP Q82863
F	16	THR	LYS	See Sequence details	UNP Q82863
F	18	THR	ALA	See Sequence details	UNP Q82863

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. 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.

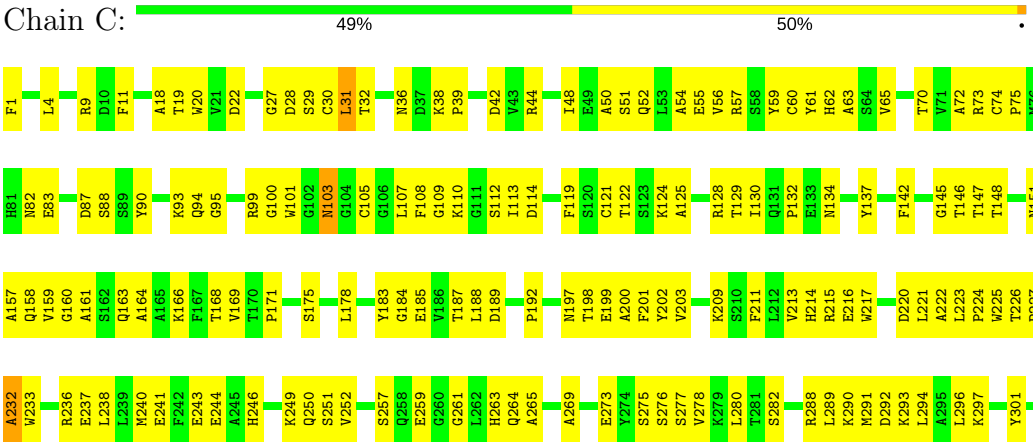
• Molecule 1: E protein

Chain A:



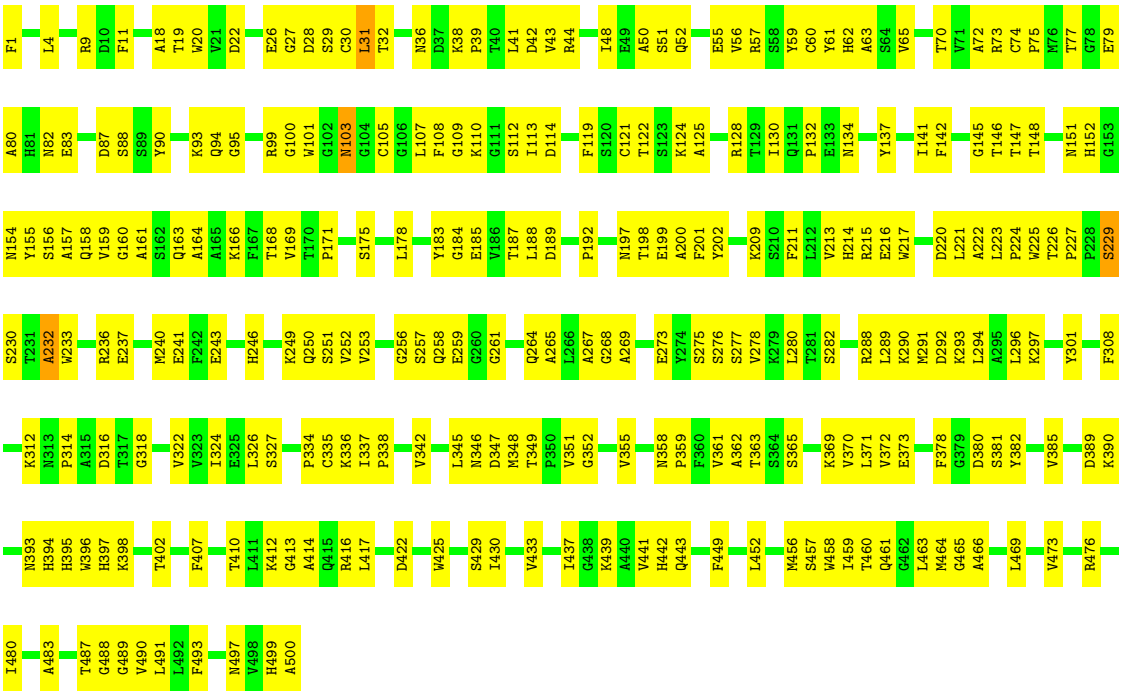
• Molecule 1: E protein

Chain C:



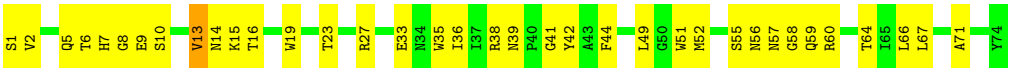
• Molecule 1: E protein

Chain E:



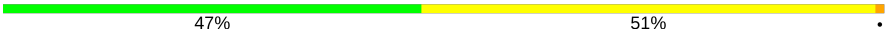
• Molecule 2: M protein

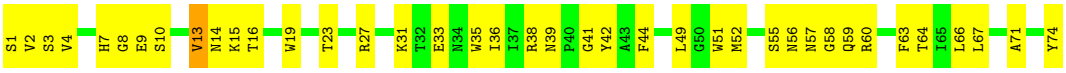
Chain B:



• Molecule 2: M protein

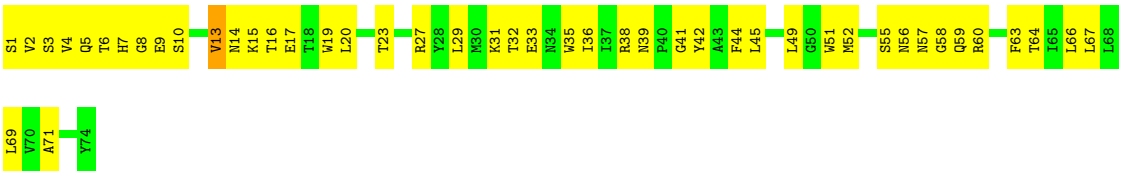
Chain D:





• Molecule 2: M protein

Chain F:


S1
V2
S3
V4
Q5
T6
H7
G8
E9
S10
V13
N14
K15
T16
E17
T18
W19
L20
T23
R27
Y28
L29
M30
K31
T32
E33
N34
W35
I36
I37
R38
N39
P40
G41
Y42
A43
F44
L45
L49
G50
W51
M52
S55
N56
N57
G58
Q59
R60
F63
T64
I65
L66
L67
L68
L69
V70
A71
Y74

4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of particles used	15260	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI POLARA 300	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	1.2	Depositor
Minimum defocus (nm)	1200	Depositor
Maximum defocus (nm)	3000	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (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	A	0.56	0/3835	0.62	1/5198 (0.0%)
1	C	0.58	0/3835	0.63	1/5198 (0.0%)
1	E	0.58	0/3835	0.63	1/5198 (0.0%)
2	B	0.44	0/586	0.59	0/799
2	D	0.44	0/586	0.59	0/799
2	F	0.44	0/586	0.59	0/799
All	All	0.56	0/13263	0.62	3/17991 (0.0%)

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
1	A	0	3
1	C	0	3
1	E	0	3
All	All	0	9

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	31	LEU	CA-CB-CG	-5.18	103.38	115.30
1	E	31	LEU	CA-CB-CG	-5.17	103.40	115.30
1	C	31	LEU	CA-CB-CG	-5.17	103.41	115.30

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	103	ASN	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	229	SER	Peptide
1	A	232	ALA	Peptide
1	C	103	ASN	Peptide
1	C	229	SER	Peptide

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	3753	0	3692	360	0
1	C	3753	0	3692	239	0
1	E	3753	0	3692	285	0
2	B	574	0	585	48	0
2	D	574	0	585	46	0
2	F	574	0	585	63	0
All	All	12981	0	12831	924	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 36.

The worst 5 of 924 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:348:MET:CG	1:A:387:ARG:HH21	1.15	1.56
1:A:312:LYS:NZ	1:A:325:GLU:HB2	1.37	1.36
1:A:348:MET:CB	1:A:387:ARG:HH21	1.40	1.34
1:A:348:MET:CB	1:A:387:ARG:NH2	1.93	1.32
1:A:348:MET:HB3	1:A:387:ARG:NH2	1.45	1.27

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 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	A	498/500 (100%)	448 (90%)	44 (9%)	6 (1%)	14 55
1	C	498/500 (100%)	451 (91%)	45 (9%)	2 (0%)	36 76
1	E	498/500 (100%)	451 (91%)	45 (9%)	2 (0%)	36 76
2	B	72/74 (97%)	61 (85%)	10 (14%)	1 (1%)	12 51
2	D	72/74 (97%)	61 (85%)	10 (14%)	1 (1%)	12 51
2	F	72/74 (97%)	61 (85%)	10 (14%)	1 (1%)	12 51
All	All	1710/1722 (99%)	1533 (90%)	164 (10%)	13 (1%)	26 64

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	490	VAL
1	C	490	VAL
1	E	490	VAL
1	A	387	ARG
1	A	221	LEU

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 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	A	399/399 (100%)	389 (98%)	10 (2%)	50 74
1	C	399/399 (100%)	399 (100%)	0	100 100
1	E	399/399 (100%)	399 (100%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	B	62/63 (98%)	62 (100%)	0	100 100
2	D	62/63 (98%)	62 (100%)	0	100 100
2	F	62/63 (98%)	62 (100%)	0	100 100
All	All	1383/1386 (100%)	1373 (99%)	10 (1%)	86 92

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	387	ARG
1	A	390	LYS
1	A	405	LYS
1	A	218	PHE
1	A	403	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	258	GLN
1	C	395	HIS
2	F	5	GLN
1	C	263	HIS
1	C	264	GLN

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)

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.