wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 01:46 PM BST

PDB ID	:	3IZX
EMDB ID:	:	EMD-5256
Title	:	3.1 Angstrom cryoEM structure of cytoplasmic polyhedrosis virus
Authors	:	Yu, X.; Ge, P.; Jiang, J.; Atanasov, I.; Zhou, Z.H.
Deposited on	:	2011-01-15
Resolution	:	3.10 Å(reported)

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

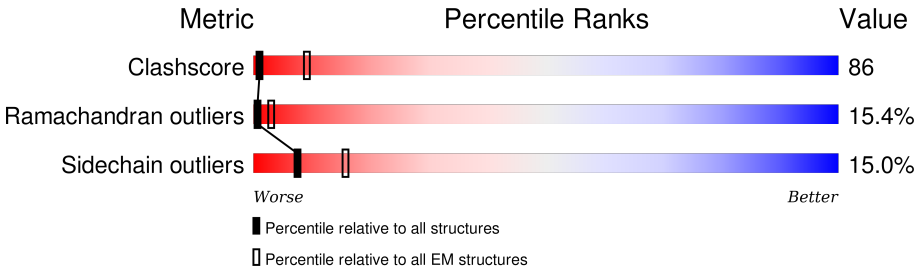
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

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 3.10 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	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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	1058	20% 54% 23% .
2	B	1333	17% 49% 21% . 11%
2	C	1333	21% 50% 20% . 6%
3	D	448	15% 34% 14% . 35%
3	E	448	15% 35% 14% . 35%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 32209 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 Structural protein VP3.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1057	Total	C	N	O	S	0	0
			8434	5345	1457	1587	45		

- Molecule 2 is a protein called Capsid protein VP1.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	1191	Total	C	N	O	S	0	0
			9397	5937	1634	1789	37		
2	C	1249	Total	C	N	O	S	0	0
			9844	6213	1712	1882	37		

- Molecule 3 is a protein called Viral structural protein 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	290	Total	C	N	O	S	0	0
			2267	1440	398	422	7		
3	E	290	Total	C	N	O	S	0	0
			2267	1440	398	422	7		

There are 2 discrepancies between the modelled and reference sequences:

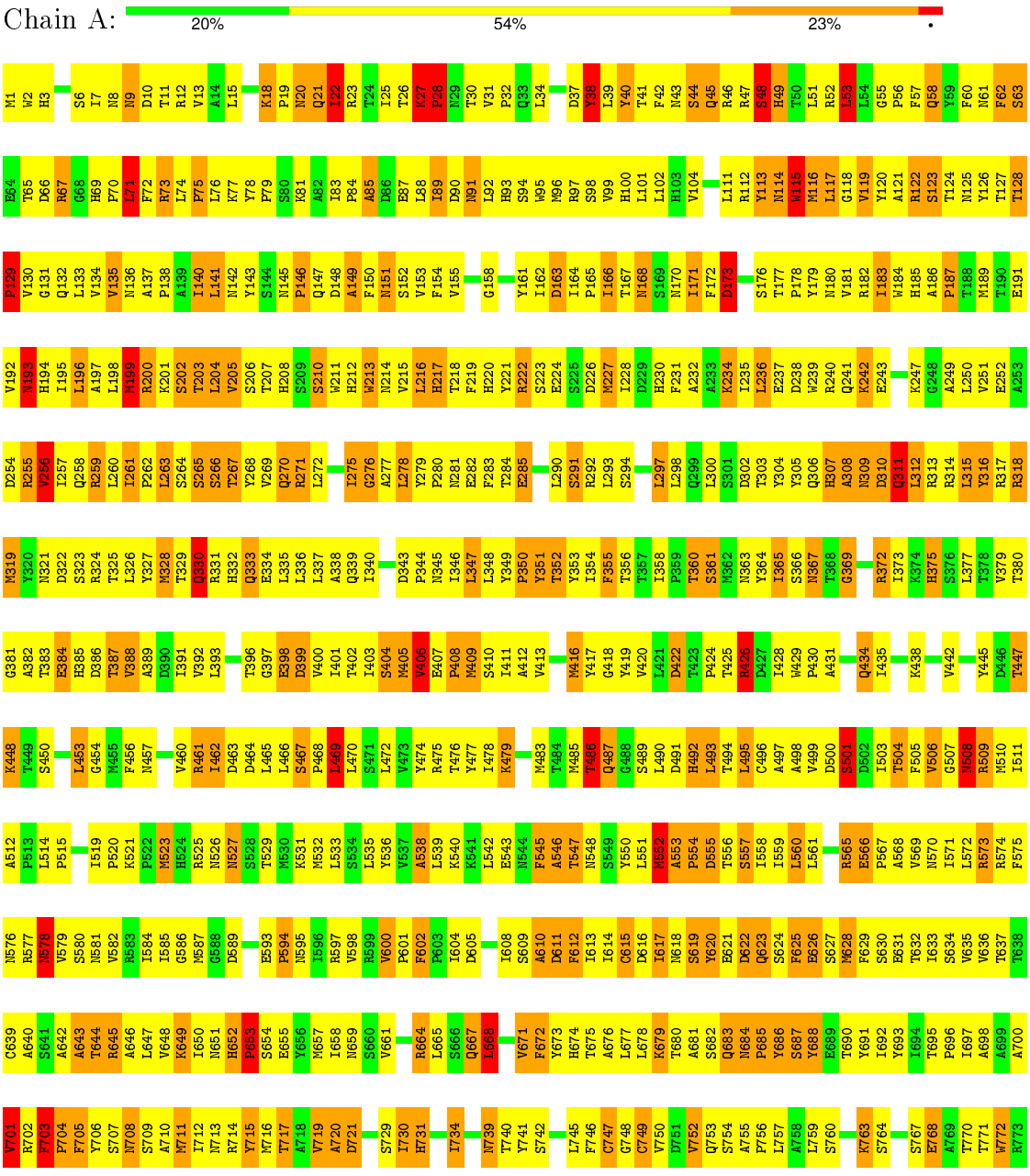
Chain	Residue	Modelled	Actual	Comment	Reference
D	37	TRP	TYR	ENGINEERED MUTATION	UNP C6K2M8
E	37	TRP	TYR	ENGINEERED MUTATION	UNP C6K2M8

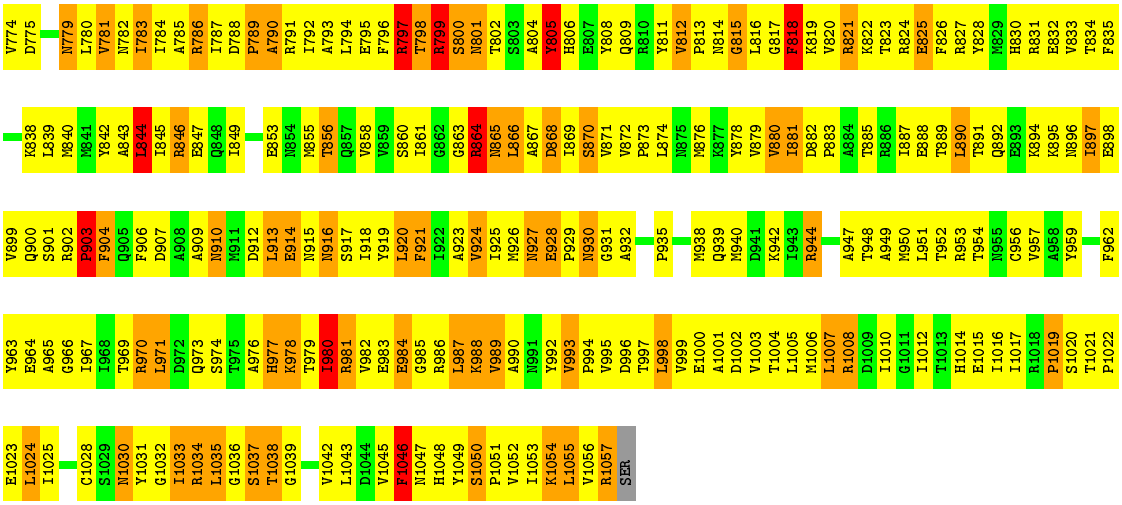
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Structural protein VP3

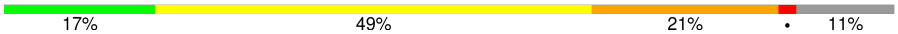
Chain A:

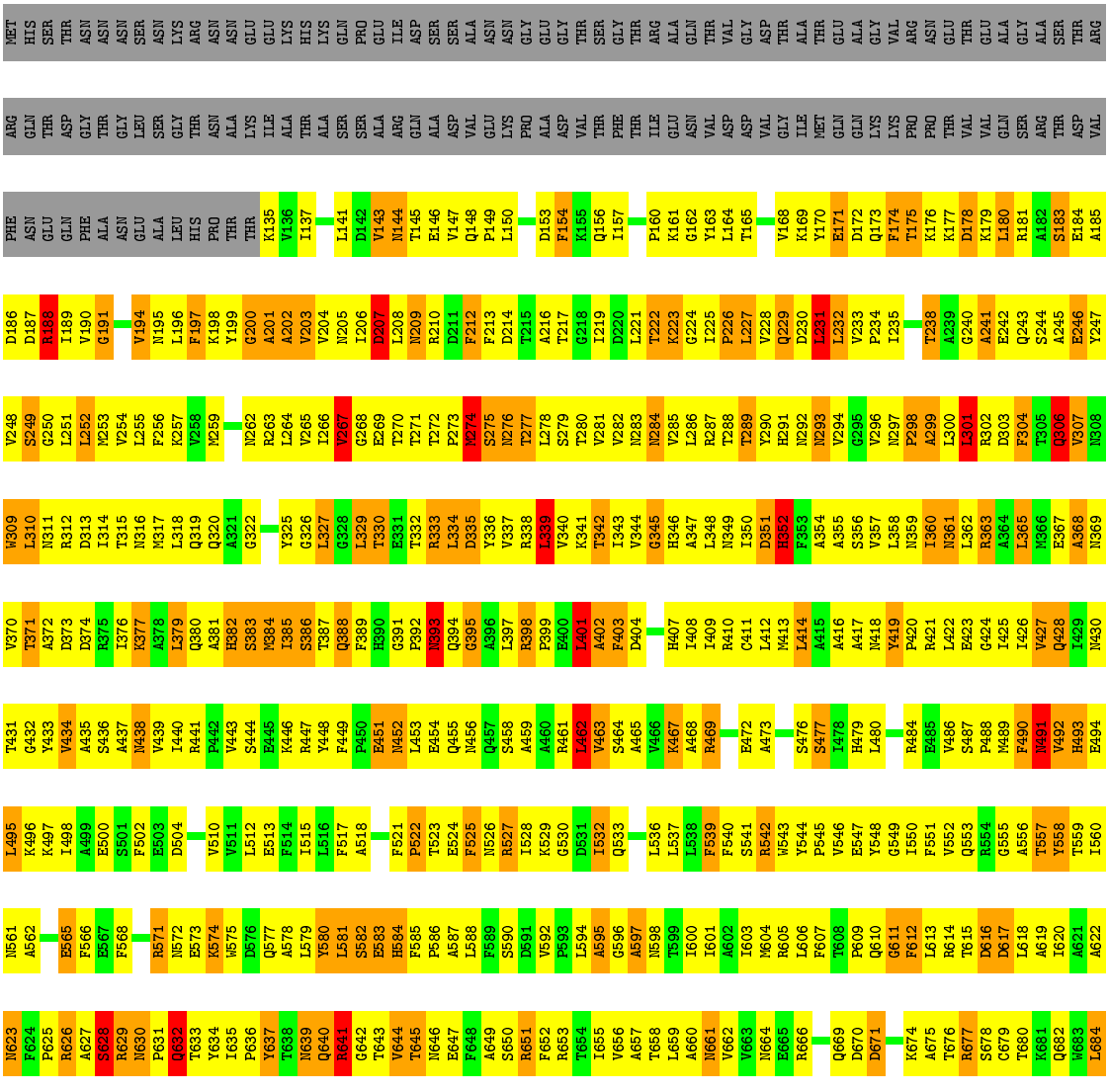


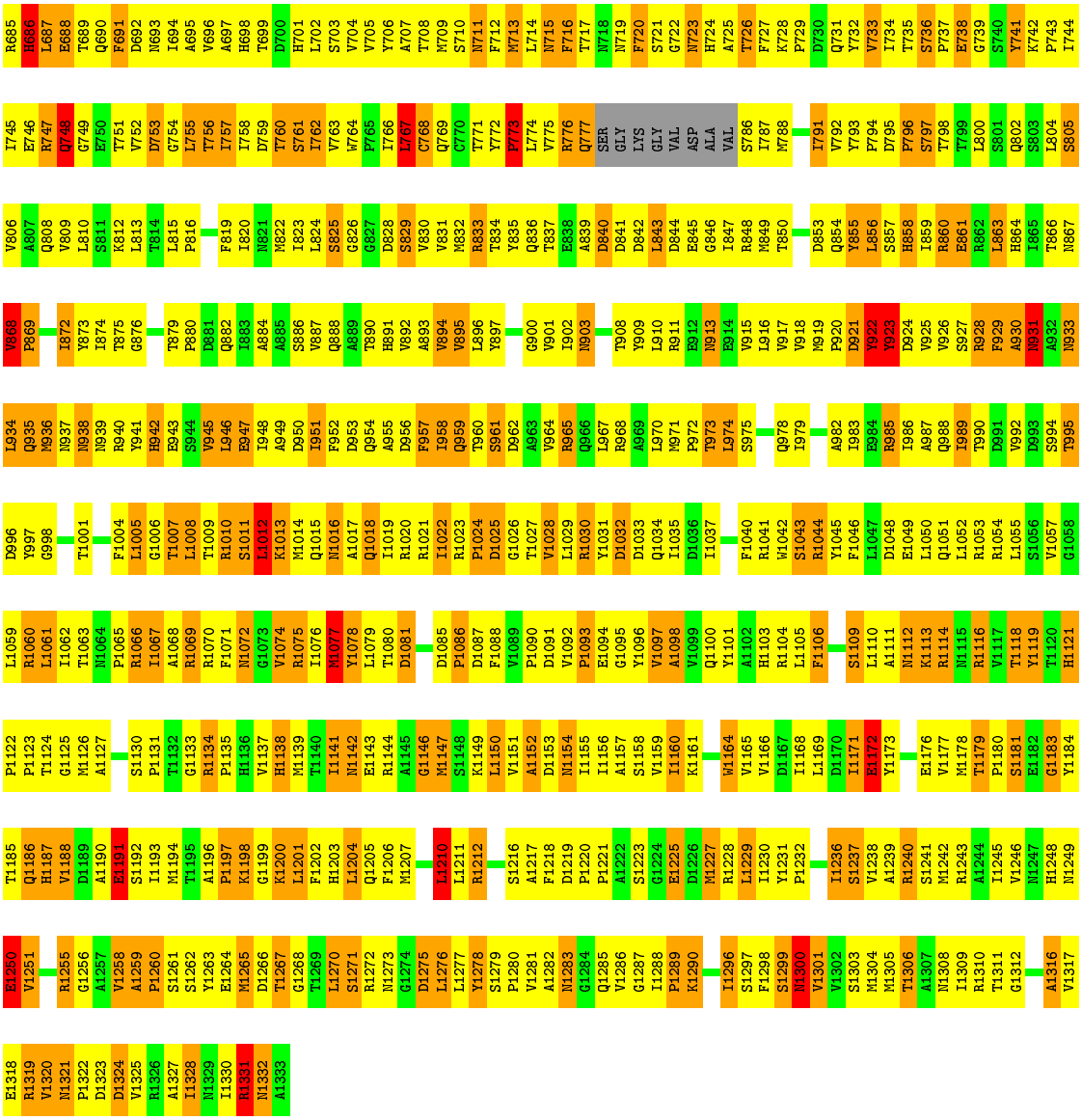


• Molecule 2: Capsid protein VP1

Chain B:

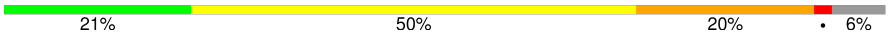


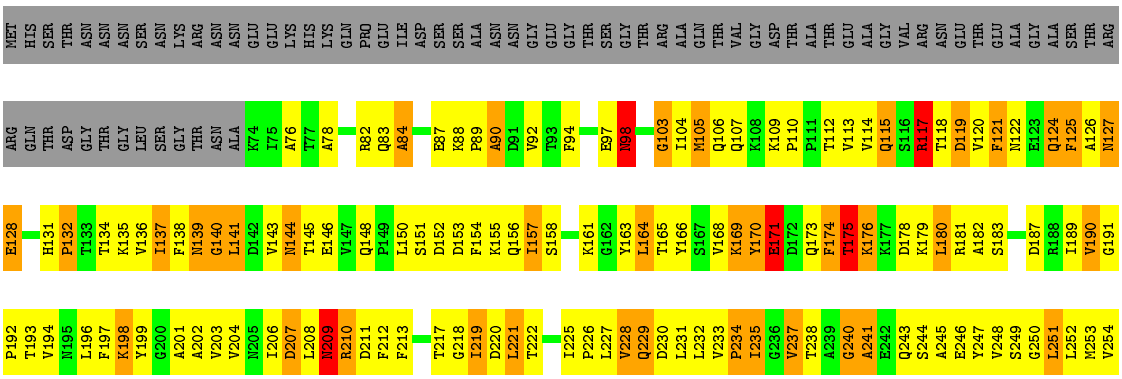


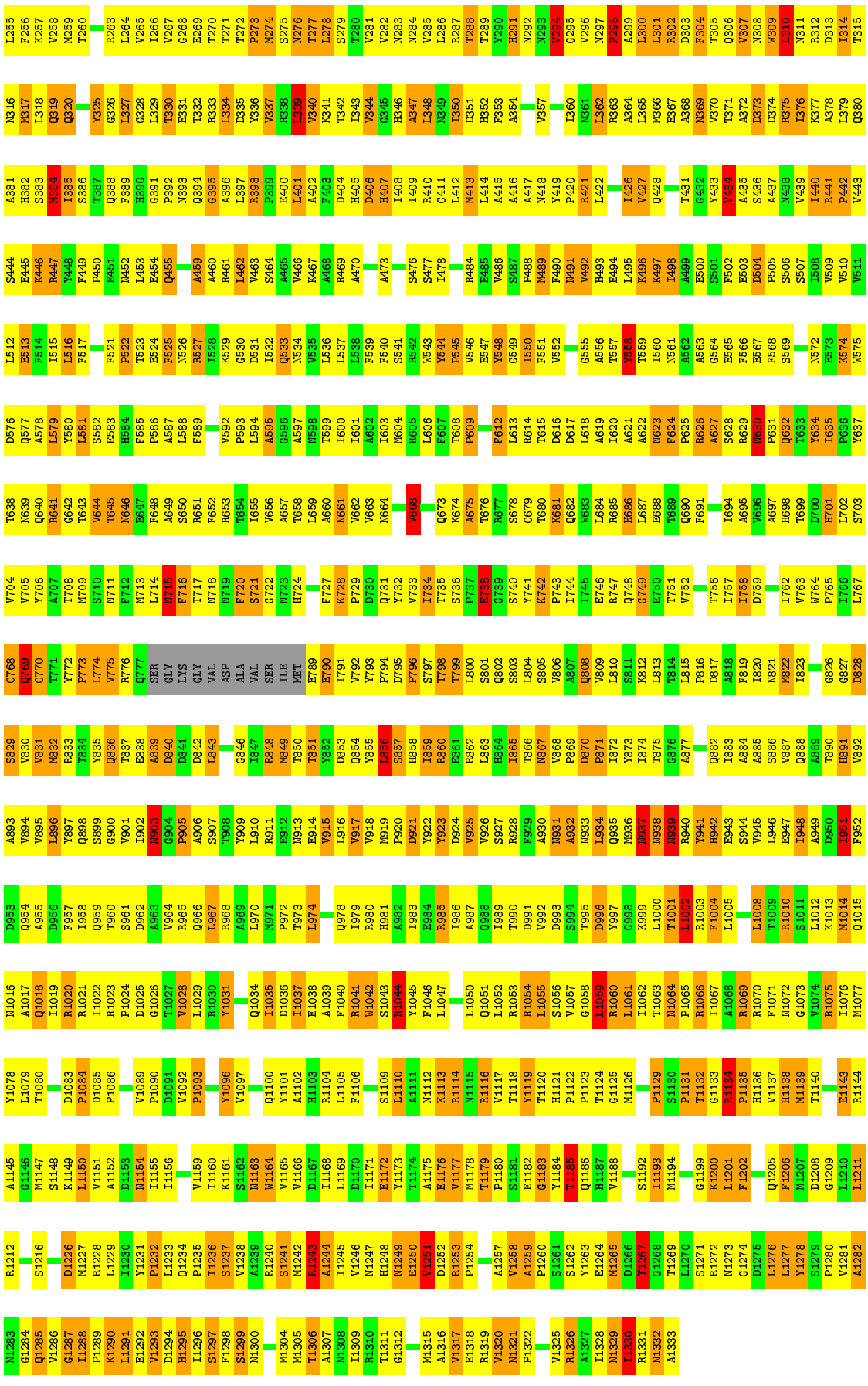


• Molecule 2: Capsid protein VP1

Chain C:

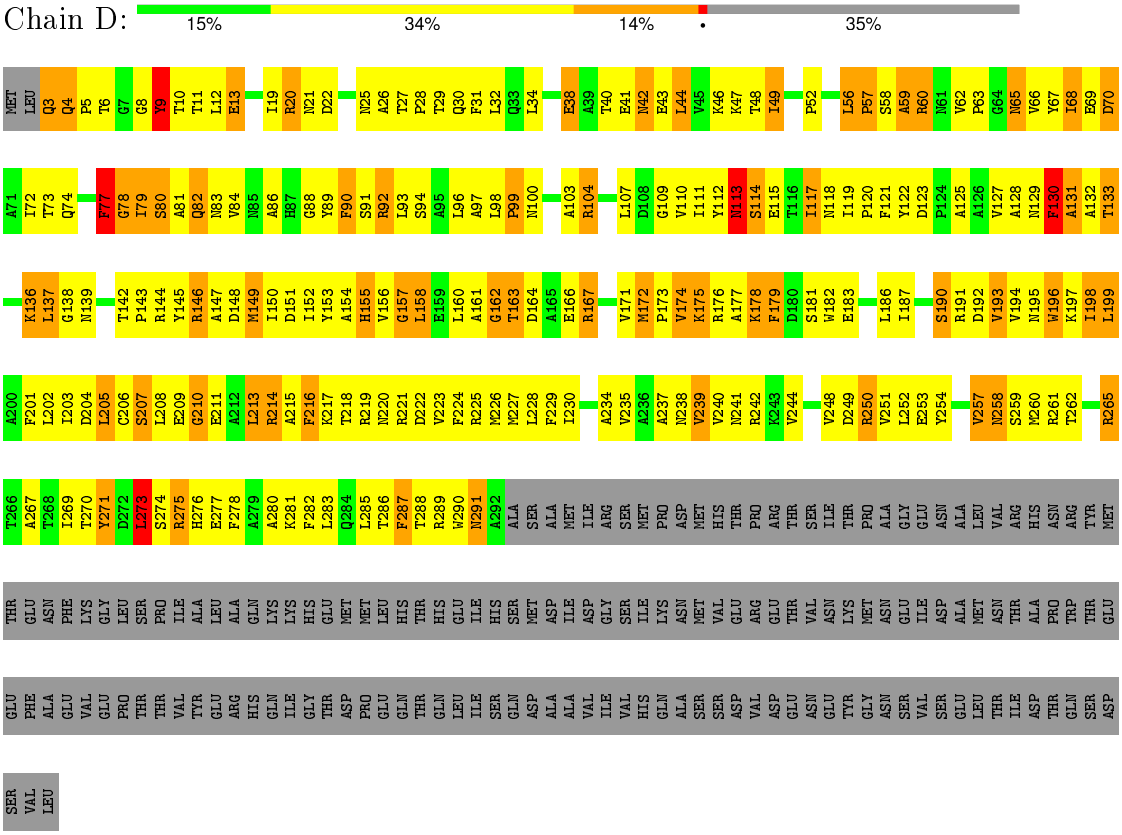






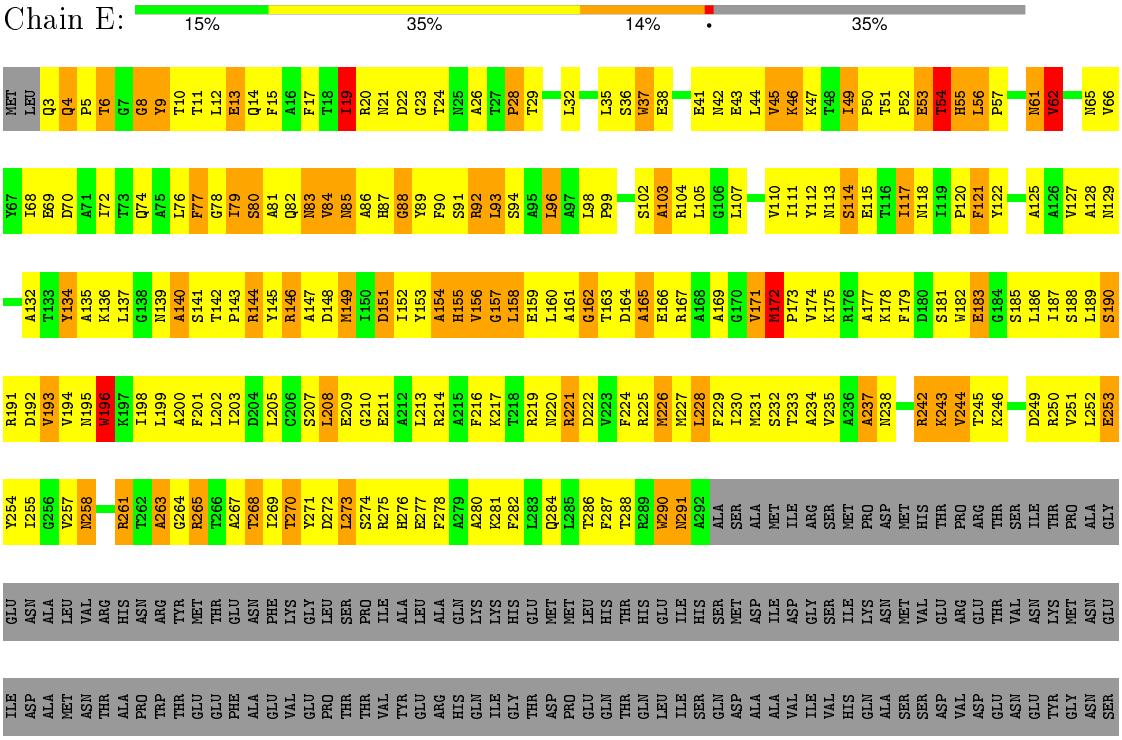
- Molecule 3: Viral structural protein 5

Chain D:



- Molecule 3: Viral structural protein 5

Chain E:



VAL
SER
GLU
LEU
THR
ILE
ASP
THR
GLN
SER
ASP
SER
VAL
LEU

4 Experimental information (i)

Property	Value	Source
Reconstruction method	Not provided	Depositor
Imposed symmetry	POINT, I	Depositor
Number of images	28993	Depositor
Resolution determination method	FSC at 0.143 cut-off	Depositor
CTF correction method	Each particle	Depositor
Microscope	OTHER	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	25	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2500	Depositor
Magnification	59000	Depositor
Image detector	Kodak SO 163 film	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	0.68	9/8619 (0.1%)	1.04	42/11737 (0.4%)
2	B	0.77	16/9590 (0.2%)	1.08	49/13056 (0.4%)
2	C	0.69	7/10045 (0.1%)	1.07	42/13678 (0.3%)
3	D	0.63	0/2314	1.03	17/3147 (0.5%)
3	E	0.65	0/2314	0.98	12/3147 (0.4%)
All	All	0.71	32/32882 (0.1%)	1.06	162/44765 (0.4%)

The worst 5 of 32 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	202	ALA	CA-CB	11.75	1.77	1.52
2	B	1098	ALA	CA-CB	8.29	1.69	1.52
2	C	114	VAL	CA-CB	8.29	1.72	1.54
2	C	1320	VAL	CA-CB	7.62	1.70	1.54
1	A	232	ALA	CA-CB	-7.29	1.37	1.52

The worst 5 of 162 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	550	ILE	CB-CA-C	-9.49	92.62	111.60
2	C	1059	LEU	CA-CB-CG	9.28	136.65	115.30
2	C	1316	ALA	CB-CA-C	8.39	122.68	110.10
1	A	971	LEU	CA-CB-CG	-8.22	96.39	115.30
2	B	863	LEU	CA-CB-CG	-8.15	96.55	115.30

There are no chirality outliers.

There are no planarity outliers.

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	A	8434	0	8395	1501	0
2	B	9397	0	9313	1584	0
2	C	9844	0	9749	1701	0
3	D	2267	0	2260	343	0
3	E	2267	0	2260	348	0
All	All	32209	0	31977	5388	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 86.

The worst 5 of 5388 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:C:1242:MET:HE1	2:C:1260:PRO:CD	1.25	1.60
2:C:615:THR:CG2	2:C:632:GLN:HB3	1.26	1.60
3:D:26:ALA:CB	3:D:30:GLN:HE21	1.13	1.60
2:C:832:MET:CE	2:C:946:LEU:HD12	1.34	1.56
2:B:202:ALA:CB	2:B:202:ALA:CA	1.77	1.56

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	1055/1058 (100%)	659 (62%)	233 (22%)	163 (16%)	0 0
2	B	1187/1333 (89%)	800 (67%)	183 (15%)	204 (17%)	0 0
2	C	1245/1333 (93%)	861 (69%)	207 (17%)	177 (14%)	0 1
3	D	288/448 (64%)	203 (70%)	46 (16%)	39 (14%)	0 1

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
3	E	288/448 (64%)	191 (66%)	55 (19%)	42 (15%)	0 1
All	All	4063/4620 (88%)	2714 (67%)	724 (18%)	625 (15%)	1 0

5 of 625 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	48	SER
1	A	63	SER
1	A	123	SER
1	A	135	VAL
1	A	202	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 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	942/943 (100%)	786 (83%)	156 (17%)	3 12
2	B	1038/1153 (90%)	891 (86%)	147 (14%)	4 18
2	C	1088/1153 (94%)	928 (85%)	160 (15%)	4 16
3	D	238/379 (63%)	204 (86%)	34 (14%)	4 17
3	E	238/379 (63%)	203 (85%)	35 (15%)	4 16
All	All	3544/4007 (88%)	3012 (85%)	532 (15%)	7 15

5 of 532 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	773	PRO
2	B	1290	LYS
3	D	271	TYR
2	B	922	TYR
2	B	1112	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 136 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	711	ASN
2	B	978	GLN
3	D	25	ASN
2	B	719	ASN
2	B	854	GLN

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.