wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 01:50 PM BST

PDB ID	:	3J7G
EMDB ID:	:	EMD-5994
Title	:	Electron cryo-microscopy of human papillomavirus 16 and H16.V5 Fab fragments
Authors	:	Lee, H.; Hafenstein, S.
Deposited on	:	2014-06-24
Resolution	:	13.60 Å(reported)
Based on PDB ID	:	3OAE

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

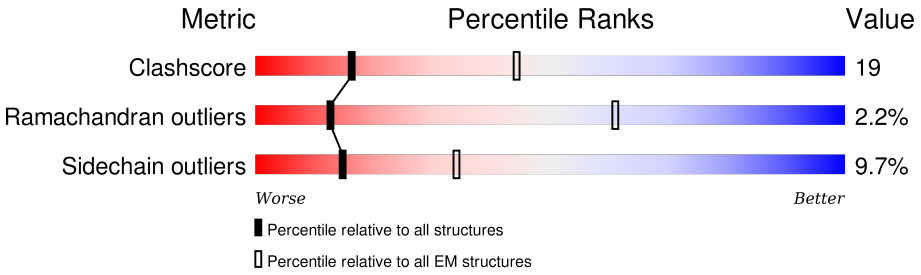
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

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 13.60 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	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	455	55% 31% 6% 7%
1	B	455	56% 30% 6% 7%
1	C	455	56% 30% 6% 7%
1	D	455	55% 31% 6% 7%
1	E	455	57% 30% 6% 7%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 16610 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 L1.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	421	Total	C	N	O	S	0	0
			3322	2116	556	630	20		
1	B	421	Total	C	N	O	S	0	0
			3322	2116	556	630	20		
1	C	421	Total	C	N	O	S	0	0
			3322	2116	556	630	20		
1	D	421	Total	C	N	O	S	0	0
			3322	2116	556	630	20		
1	E	421	Total	C	N	O	S	0	0
			3322	2116	556	630	20		

There are 20 discrepancies between the modelled and reference sequences:

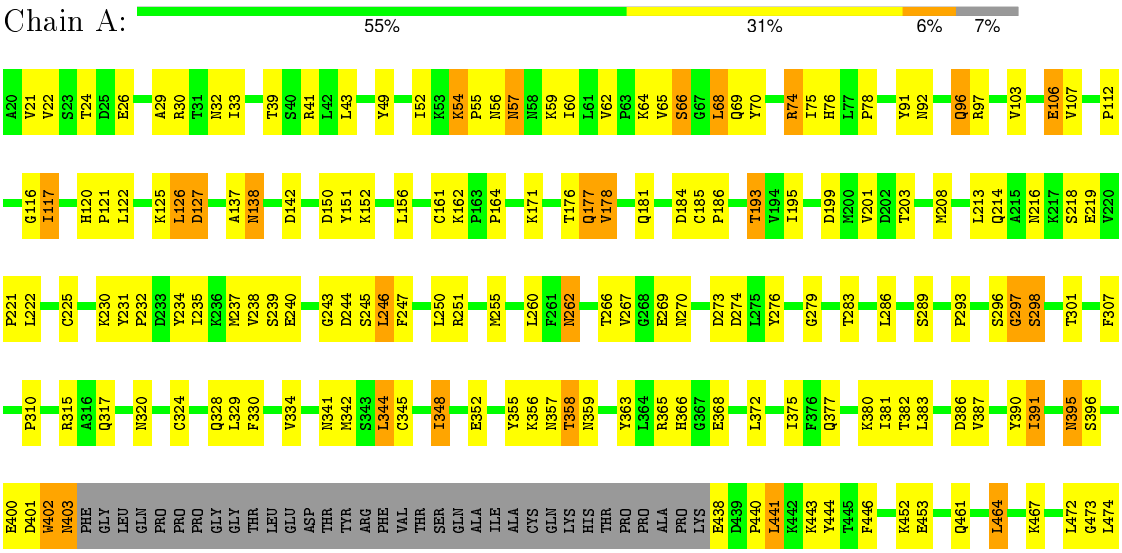
Chain	Residue	Modelled	Actual	Comment	Reference
A	20	ALA	-	EXPRESSION TAG	UNP Q4VRM0
A	177	GLN	ASN	CONFLICT	UNP Q4VRM0
A	181	GLN	ASN	CONFLICT	UNP Q4VRM0
A	472	LEU	ALA	CONFLICT	UNP Q4VRM0
B	20	ALA	-	EXPRESSION TAG	UNP Q4VRM0
B	177	GLN	ASN	CONFLICT	UNP Q4VRM0
B	181	GLN	ASN	CONFLICT	UNP Q4VRM0
B	472	LEU	ALA	CONFLICT	UNP Q4VRM0
C	20	ALA	-	EXPRESSION TAG	UNP Q4VRM0
C	177	GLN	ASN	CONFLICT	UNP Q4VRM0
C	181	GLN	ASN	CONFLICT	UNP Q4VRM0
C	472	LEU	ALA	CONFLICT	UNP Q4VRM0
D	20	ALA	-	EXPRESSION TAG	UNP Q4VRM0
D	177	GLN	ASN	CONFLICT	UNP Q4VRM0
D	181	GLN	ASN	CONFLICT	UNP Q4VRM0
D	472	LEU	ALA	CONFLICT	UNP Q4VRM0
E	20	ALA	-	EXPRESSION TAG	UNP Q4VRM0
E	177	GLN	ASN	CONFLICT	UNP Q4VRM0
E	181	GLN	ASN	CONFLICT	UNP Q4VRM0
E	472	LEU	ALA	CONFLICT	UNP Q4VRM0

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.

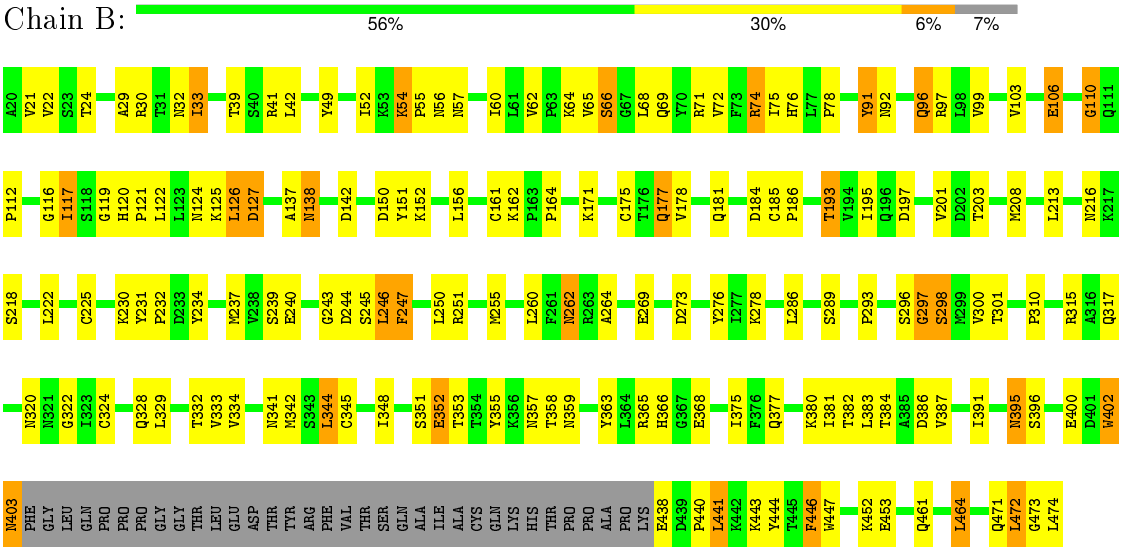
- Molecule 1: L1

Chain A:



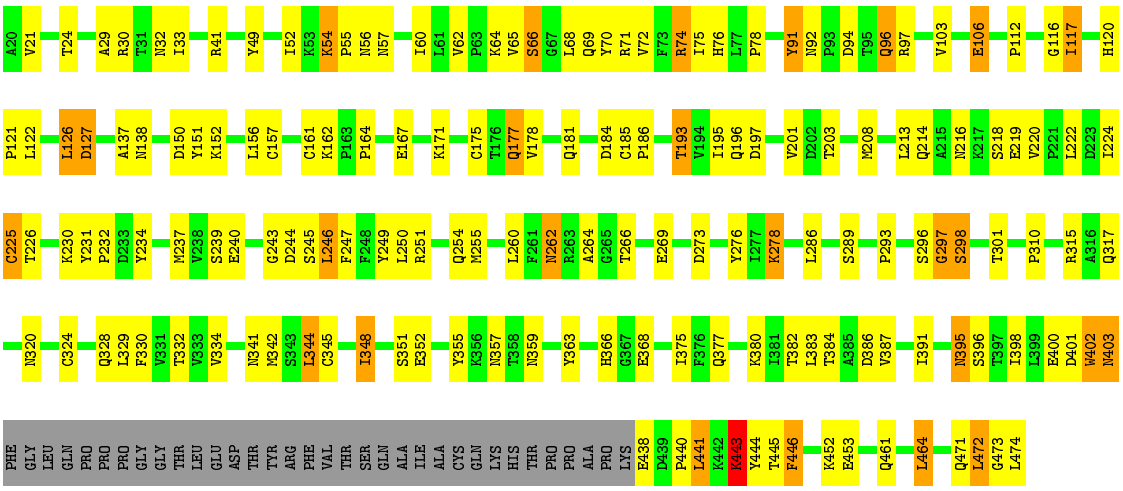
- Molecule 1: L1

Chain B:



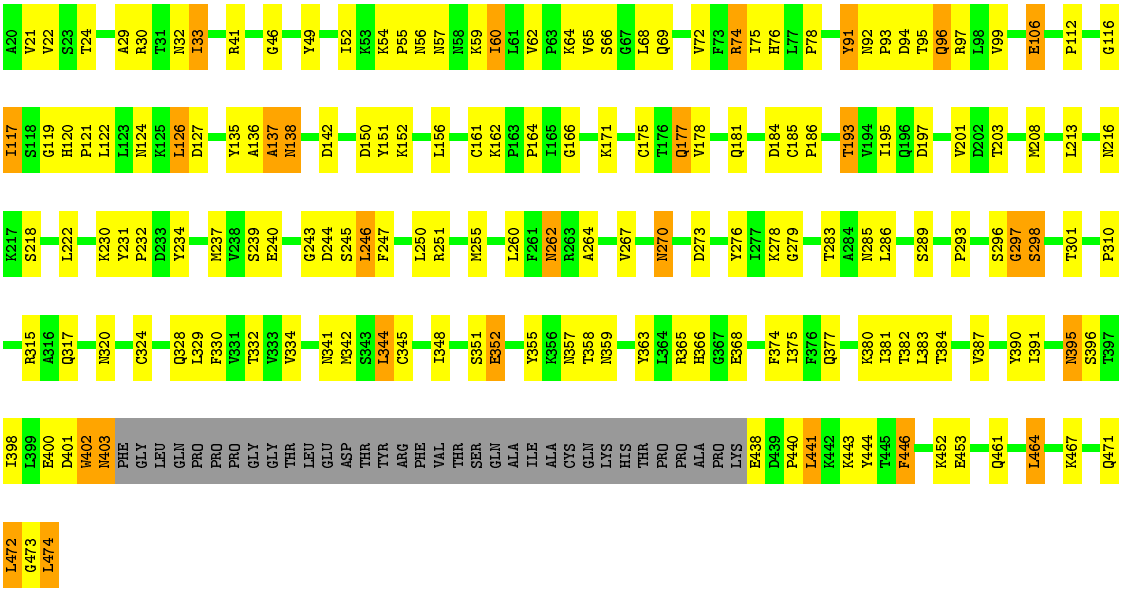
- Molecule 1: L1

Chain C:



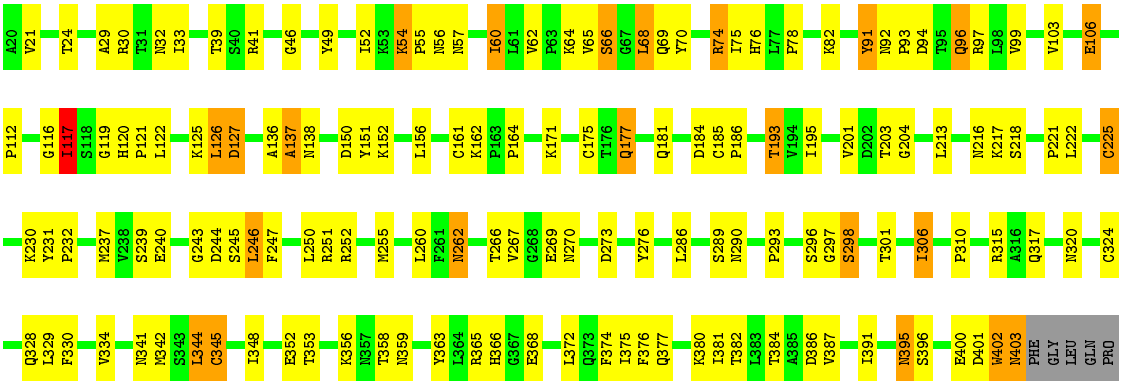
- Molecule 1: L1

Chain D:



- Molecule 1: L1

Chain E:




PRO
PRO
GLY
GLY
THR
LEU
GLU
ASP
THR
TYR
ARG
PHE
VAL
THR
SER
GLN
ALA
ILE
ALA
CYS
GLN
LYS
HIS
THR
PRO
PRO
ALA
PRO
LYS
E438
D439
P440
L441
K442
K443
Y444
T445
F446
K452
E453
Q461
L464
Q471
L472
G473
L474

4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of images	2075	Depositor
Resolution determination method	FSC 0.5	Depositor
CTF correction method	Each particle	Depositor
Microscope	JEOL 2100	Depositor
Voltage (kV)	200	Depositor
Electron dose ($e^-/Å^2$)	15	Depositor
Minimum defocus (nm)	690	Depositor
Maximum defocus (nm)	3990	Depositor
Magnification	80000	Depositor
Image detector	GATAN ULTRASCAN 4000 (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	0.58	6/3407 (0.2%)	0.72	3/4632 (0.1%)
1	B	0.49	0/3407	0.70	2/4632 (0.0%)
1	C	0.53	2/3407 (0.1%)	0.71	1/4632 (0.0%)
1	D	0.55	3/3407 (0.1%)	0.71	3/4632 (0.1%)
1	E	0.53	1/3407 (0.0%)	0.72	0/4632
All	All	0.54	12/17035 (0.1%)	0.71	9/23160 (0.0%)

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	59	LYS	CB-CG	-8.88	1.28	1.52
1	A	443	LYS	CD-CE	-6.57	1.34	1.51
1	A	57	ASN	CG-OD1	-6.33	1.10	1.24
1	C	443	LYS	CD-CE	-6.02	1.36	1.51
1	D	285	ASN	CG-OD1	-6.01	1.10	1.24

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	443	LYS	CD-CE-NZ	-6.44	96.89	111.70
1	C	297	GLY	N-CA-C	-5.74	98.75	113.10
1	D	297	GLY	N-CA-C	-5.69	98.88	113.10
1	D	59	LYS	CA-CB-CG	5.62	125.76	113.40
1	A	391	ILE	CG1-CB-CG2	-5.55	99.18	111.40

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	3322	0	3227	132	0
1	B	3322	0	3227	141	0
1	C	3322	0	3227	135	0
1	D	3322	0	3227	138	0
1	E	3322	0	3227	135	0
All	All	16610	0	16135	609	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 19.

The worst 5 of 609 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:54:LYS:HB3	1:A:57:ASN:HB3	1.23	1.09
1:C:97:ARG:HH21	1:C:403:ASN:HB2	1.13	1.07
1:B:461:GLN:HE22	1:C:21:VAL:H	1.04	1.01
1:C:54:LYS:HB3	1:C:57:ASN:HB3	1.40	1.01
1:D:461:GLN:HE22	1:E:21:VAL:H	1.05	1.01

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	417/455 (92%)	378 (91%)	29 (7%)	10 (2%)	7 47
1	B	417/455 (92%)	377 (90%)	30 (7%)	10 (2%)	7 47
1	C	417/455 (92%)	376 (90%)	33 (8%)	8 (2%)	10 52
1	D	417/455 (92%)	377 (90%)	30 (7%)	10 (2%)	7 47

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	E	417/455 (92%)	378 (91%)	31 (7%)	8 (2%)	10 52
All	All	2085/2275 (92%)	1886 (90%)	153 (7%)	46 (2%)	13 49

5 of 46 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	137	ALA
1	A	298	SER
1	A	402	TRP
1	B	137	ALA
1	B	298	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	368/396 (93%)	334 (91%)	34 (9%)	11 43
1	B	368/396 (93%)	335 (91%)	33 (9%)	12 44
1	C	368/396 (93%)	329 (89%)	39 (11%)	8 36
1	D	368/396 (93%)	334 (91%)	34 (9%)	11 43
1	E	368/396 (93%)	330 (90%)	38 (10%)	9 37
All	All	1840/1980 (93%)	1662 (90%)	178 (10%)	15 40

5 of 178 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	213	LEU
1	C	441	LEU
1	E	317	GLN
1	C	246	LEU
1	C	317	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 62 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	138	ASN
1	C	395	ASN
1	E	341	ASN
1	C	290	ASN
1	C	461	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.