wwPDB EM Map/Model Validation Report (i)

Jun 2, 2016 – 01:55 PM EDT

PDB ID	:	5FYN
EMDB ID:	:	EMD-3364
Title	:	Sub-tomogram averaging of Tula virus glycoprotein spike
Authors	:	Li, S.; Rissanen, I.; Zeltina, A.; Hepojoki, J.; Raghwani, J.; Harlos, K.; Pybus, O.G.; Huiskonen, J.T.; Bowden, T.A.
Deposited on	:	2016-03-08
Resolution	:	15.60 Å(reported)
Based on PDB ID	:	5FXU

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

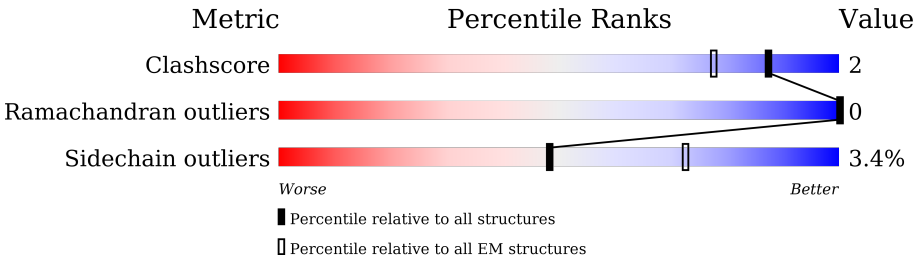
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027674

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 15.60 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	358	85% 7% 8%
1	B	358	86% 5% • 7%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5213 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 PUUMALA VIRUS GN GLYCOPROTEIN.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	331	Total	C	N	O	S	0	0
			2517	1599	409	488	21		
1	B	332	Total	C	N	O	S	4	0
			2551	1623	416	491	21		

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	26	GLU	-	EXPRESSION TAG	UNP Q9WJ31
A	27	THR	-	EXPRESSION TAG	UNP Q9WJ31
A	28	GLY	-	EXPRESSION TAG	UNP Q9WJ31
B	26	GLU	-	EXPRESSION TAG	UNP Q9WJ31
B	27	THR	-	EXPRESSION TAG	UNP Q9WJ31
B	28	GLY	-	EXPRESSION TAG	UNP Q9WJ31

- Molecule 2 is a polymer of unknown type called SUGAR (4-MER).

Mol	Chain	Residues	Atoms				AltConf
2	A	4	Total	C	N	O	0
			50	28	2	20	

- Molecule 3 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				AltConf
3	A	2	Total	C	N	O	0
			28	16	2	10	
3	B	2	Total	C	N	O	0
			28	16	2	10	

- Molecule 4 is a polymer of unknown type called SUGAR (3-MER).

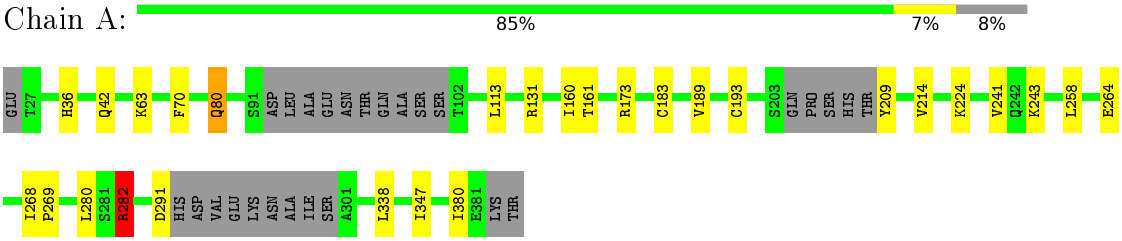
Mol	Chain	Residues	Atoms				AltConf
			Total	C	N	O	
4	B	3	39	22	2	15	0

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 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.

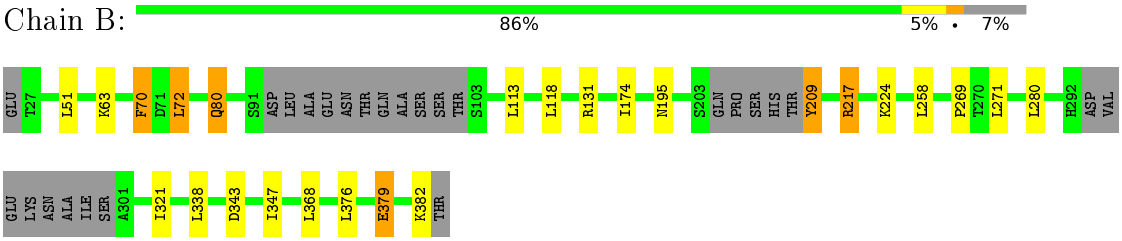
- Molecule 1: PUUMALA VIRUS GN GLYCOPROTEIN

Chain A:



- Molecule 1: PUUMALA VIRUS GN GLYCOPROTEIN

Chain B:



4 Experimental information (i)

Property	Value	Source
Reconstruction method	TOMOGRAPHY	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of tilted images used	Not provided	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	EACH TILTED IMAGE	Depositor
Microscope	OTHER	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	60	Depositor
Minimum defocus (nm)	2000	Depositor
Maximum defocus (nm)	3800	Depositor
Magnification	160000	Depositor
Image detector	GATAN K2 SUMMIT	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: BMA, NAG, MAN

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	0.60	0/2567	0.79	3/3488 (0.1%)
1	B	0.62	0/2611	0.78	4/3548 (0.1%)
All	All	0.61	0/5178	0.79	7/7036 (0.1%)

There are no bond length outliers.

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	282	ARG	NE-CZ-NH1	9.98	125.29	120.30
1	B	217	ARG	NE-CZ-NH2	-6.46	117.07	120.30
1	B	131	ARG	NE-CZ-NH1	5.71	123.15	120.30
1	A	282	ARG	NE-CZ-NH2	-5.66	117.47	120.30
1	A	131	ARG	NE-CZ-NH1	5.64	123.12	120.30
1	B	258	LEU	CB-CG-CD2	5.11	119.69	111.00
1	B	217	ARG	NE-CZ-NH1	5.05	122.83	120.30

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	2517	0	2512	16	0
1	B	2551	0	2556	13	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	50	0	43	0	0
3	A	28	0	25	0	0
3	B	28	0	25	0	0
4	B	39	0	34	0	0
All	All	5213	0	5195	26	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 2.

All (26) 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:160:ILE:O	1:B:209:TYR:HD2	1.18	1.27
1:A:160:ILE:O	1:B:209:TYR:CD2	2.01	1.14
1:B:224:LYS:NZ	1:B:343:ASP:OD1	2.27	0.67
1:A:241:VAL:HG12	1:A:241:VAL:O	2.02	0.59
1:A:36:HIS:NE2	1:A:183:CYS:O	2.37	0.57
1:A:269:PRO:HB3	1:A:280:LEU:HD11	1.86	0.57
1:B:195[B]:ASN:HD22	1:B:379:GLU:HG2	1.72	0.55
1:B:269:PRO:HB3	1:B:280:LEU:HD11	1.87	0.55
1:A:282:ARG:HH11	1:A:282:ARG:HG2	1.72	0.54
1:A:193:CYS:O	1:A:380:ILE:O	2.25	0.54
1:A:338:LEU:HG	1:A:347:ILE:HD11	1.92	0.51
1:A:241:VAL:CG1	1:A:241:VAL:O	2.59	0.50
1:B:338:LEU:HG	1:B:347:ILE:HD11	1.92	0.50
1:A:80:GLN:HG2	1:A:113:LEU:O	2.14	0.48
1:B:80:GLN:HG2	1:B:113:LEU:O	2.13	0.48
1:A:189:VAL:HG11	1:A:214:VAL:HG12	1.97	0.45
1:A:161:THR:HA	1:B:209:TYR:CE2	2.52	0.45
1:B:70:PHE:CE1	1:B:72:LEU:HD13	2.52	0.44
1:A:36:HIS:CD2	1:A:183:CYS:O	2.70	0.44
1:A:42:GLN:HB2	1:A:264:GLU:OE2	2.18	0.43
1:A:268:ILE:HA	1:A:269:PRO:HA	1.86	0.43
1:B:70:PHE:HE1	1:B:72:LEU:HD13	1.83	0.43
1:A:282:ARG:HH11	1:A:282:ARG:CG	2.32	0.43
1:B:321:ILE:HD12	1:B:347:ILE:HD13	2.01	0.42
1:B:118:LEU:HD12	1:B:368[A]:LEU:HD21	2.02	0.41
1:B:51:LEU:HD11	1:B:174:ILE:HD11	2.01	0.41

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	323/358 (90%)	316 (98%)	7 (2%)	0	100	100
1	B	328/358 (92%)	319 (97%)	9 (3%)	0	100	100
All	All	651/716 (91%)	635 (98%)	16 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	291/315 (92%)	281 (97%)	10 (3%)	44	75
1	B	296/315 (94%)	286 (97%)	10 (3%)	44	75
All	All	587/630 (93%)	567 (97%)	20 (3%)	48	75

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	63	LYS
1	A	70	PHE
1	A	80	GLN
1	A	173	ARG
1	A	209	TYR
1	A	224	LYS
1	A	243	LYS
1	A	258	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	282	ARG
1	A	291	ASP
1	B	63	LYS
1	B	70	PHE
1	B	72	LEU
1	B	80	GLN
1	B	209	TYR
1	B	217	ARG
1	B	271	LEU
1	B	376	LEU
1	B	379	GLU
1	B	382	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

11 carbohydrates are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	NAG	A	1382	1,2	14,14,15	0.40	0	15,19,21	1.26	2 (13%)
2	NAG	A	1383	2	14,14,15	0.54	0	15,19,21	0.84	0
2	BMA	A	1384	2	11,11,12	0.68	0	15,15,17	1.50	5 (33%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	MAN	A	1385	2	11,11,12	0.82	0	15,15,17	2.29	4 (26%)
3	NAG	A	1386	1,3	14,14,15	0.37	0	15,19,21	1.40	3 (20%)
3	NAG	A	1387	3	14,14,15	0.43	0	15,19,21	1.06	1 (6%)
3	NAG	B	1383	1,3	14,14,15	0.39	0	15,19,21	1.25	2 (13%)
3	NAG	B	1384	3	14,14,15	0.32	0	15,19,21	1.07	1 (6%)
4	NAG	B	1385	1,4	14,14,15	0.50	0	15,19,21	0.87	0
4	NAG	B	1386	4	14,14,15	0.50	0	15,19,21	1.01	0
4	BMA	B	1387	4	11,11,12	0.67	0	15,15,17	0.94	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	NAG	A	1382	1,2	-	0/6/23/26	0/1/1/1
2	NAG	A	1383	2	-	0/6/23/26	0/1/1/1
2	BMA	A	1384	2	-	0/2/19/22	0/1/1/1
2	MAN	A	1385	2	-	0/2/19/22	0/1/1/1
3	NAG	A	1386	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	1387	3	-	0/6/23/26	0/1/1/1
3	NAG	B	1383	1,3	-	0/6/23/26	0/1/1/1
3	NAG	B	1384	3	-	0/6/23/26	0/1/1/1
4	NAG	B	1385	1,4	-	0/6/23/26	0/1/1/1
4	NAG	B	1386	4	-	0/6/23/26	0/1/1/1
4	BMA	B	1387	4	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

All (18) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1382	NAG	O4-C4-C3	-3.25	103.03	110.36
2	A	1384	BMA	O5-C5-C4	-2.23	106.43	110.13
2	A	1384	BMA	C2-C3-C4	2.05	114.62	111.05
3	A	1386	NAG	C2-N2-C7	2.13	125.88	123.11
2	A	1385	MAN	O4-C4-C5	2.13	114.85	109.23
2	A	1382	NAG	C4-C3-C2	2.16	114.69	111.34
2	A	1384	BMA	C1-O5-C5	2.31	115.53	112.14
2	A	1384	BMA	O5-C1-C2	2.31	114.58	110.89
3	B	1383	NAG	C8-C7-N2	2.33	120.57	116.10
2	A	1385	MAN	O5-C5-C4	2.60	114.43	110.13

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	1383	NAG	C2-N2-C7	2.62	126.51	123.11
3	A	1386	NAG	C8-C7-N2	2.64	121.16	116.10
2	A	1384	BMA	C1-C2-C3	2.72	112.85	109.55
2	A	1385	MAN	O3-C3-C4	2.75	116.56	110.36
3	B	1384	NAG	C1-O5-C5	3.26	116.93	112.14
3	A	1386	NAG	C1-O5-C5	3.29	116.97	112.14
3	A	1387	NAG	C1-O5-C5	3.40	117.14	112.14
2	A	1385	MAN	C1-O5-C5	7.09	122.57	112.14

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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.