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

Feb 20, 2018 – 08:07 pm GMT

PDB ID	:	5KHC
EMDB ID:	:	EMD-8248
Title	:	Structure of rubella virus E1 glycoprotein ectodomain fitted into sub-tomogram averaged surface spike density of rubella virus
Authors	:	Mangala Prasad, V.; Klose, T.; Rossmann, M.G.
Deposited on	:	2016-06-14
Resolution	:	11.10 Å(reported)
Based on PDB ID	:	4ADG

This is a Full wwPDB/EMDataBank EM Map/Model Validation 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.

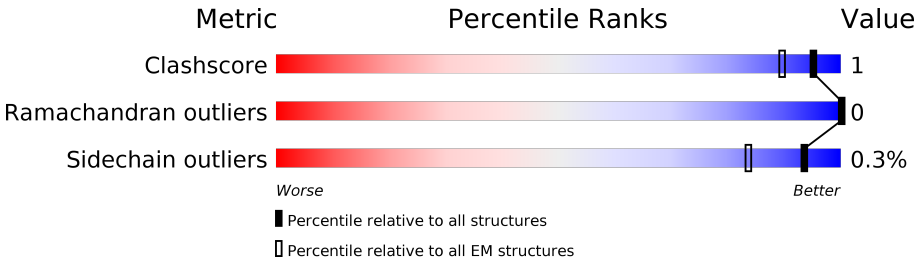
MolProbity	:	4.02b-467
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30686

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 11.10 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	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	473	85% · 12%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3179 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 E1 glycoprotein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	418	3179	2012	558	586	23	6	1

There are 37 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	437	PHE	-	expression tag	UNP P08563
A	438	GLU	-	expression tag	UNP P08563
A	439	ASP	-	expression tag	UNP P08563
A	440	ASP	-	expression tag	UNP P08563
A	441	ASP	-	expression tag	UNP P08563
A	442	ASP	-	expression tag	UNP P08563
A	443	LYS	-	expression tag	UNP P08563
A	444	ALA	-	expression tag	UNP P08563
A	445	GLY	-	expression tag	UNP P08563
A	446	TRP	-	expression tag	UNP P08563
A	447	SER	-	expression tag	UNP P08563
A	448	HIS	-	expression tag	UNP P08563
A	449	PRO	-	expression tag	UNP P08563
A	450	GLN	-	expression tag	UNP P08563
A	451	PHE	-	expression tag	UNP P08563
A	452	GLU	-	expression tag	UNP P08563
A	453	LYS	-	expression tag	UNP P08563
A	454	GLY	-	expression tag	UNP P08563
A	455	GLY	-	expression tag	UNP P08563
A	456	GLY	-	expression tag	UNP P08563
A	457	SER	-	expression tag	UNP P08563
A	458	GLY	-	expression tag	UNP P08563
A	459	GLY	-	expression tag	UNP P08563
A	460	GLY	-	expression tag	UNP P08563
A	461	SER	-	expression tag	UNP P08563
A	462	GLY	-	expression tag	UNP P08563
A	463	GLY	-	expression tag	UNP P08563
A	464	GLY	-	expression tag	UNP P08563

Continued on next page...

Continued from previous page...

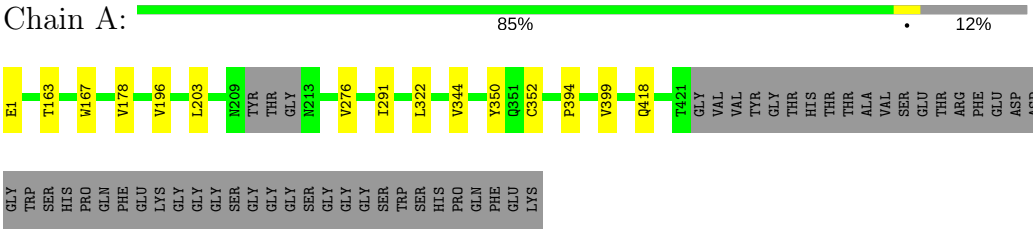
Chain	Residue	Modelled	Actual	Comment	Reference
A	465	SER	-	expression tag	UNP P08563
A	466	TRP	-	expression tag	UNP P08563
A	467	SER	-	expression tag	UNP P08563
A	468	HIS	-	expression tag	UNP P08563
A	469	PRO	-	expression tag	UNP P08563
A	470	GLN	-	expression tag	UNP P08563
A	471	PHE	-	expression tag	UNP P08563
A	472	GLU	-	expression tag	UNP P08563
A	473	LYS	-	expression tag	UNP P08563

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: E1 glycoprotein

Chain A:



4 Experimental information (i)

Property	Value	Source
Reconstruction method	SUBTOMOGRAM AVERAGING	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of subtomograms used	7290	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	90	Depositor
Minimum defocus (nm)	400	Depositor
Maximum defocus (nm)	500	Depositor
Magnification	11000	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	0.36	0/3295	0.59	0/4525

There are no bond length outliers.

There are no bond angle outliers.

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	3179	0	3019	7	0
All	All	3179	0	3019	7	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 1.

All (7) 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:276:VAL:HG12	1:A:291:ILE:HG12	1.94	0.49
1:A:163[B]:THR:HG21	1:A:203:LEU:HG	1.96	0.47
1:A:344:VAL:HG11	1:A:399:VAL:HG11	1.98	0.46
1:A:350:TYR:CZ	1:A:394:PRO:HB3	2.53	0.43
1:A:352:CYS:HB2	1:A:418:GLN:HG2	2.00	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:196:VAL:HG12	1:A:322:LEU:HD22	2.00	0.42
1:A:167:TRP:HB2	1:A:178:VAL:HG23	2.02	0.40

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	420/473 (89%)	410 (98%)	10 (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	339/376 (90%)	338 (100%)	1 (0%)	93 96

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	398	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.