



wwPDB EM Map/Model Validation Report ⓘ

Apr 10, 2016 – 02:13 PM BST

PDB ID : 5I08
EMDB ID: : EMD-8069
Title : Prefusion structure of a human coronavirus spike protein
Authors : Kirchdoerfer, R.N.; Cottrell, C.A.; Wang, N.; Pallesen, J.; Yassine, H.M.;
Turner, H.L.; Corbett, K.S.; Graham, B.S.; McLellan, J.S.; Ward, A.B.
Deposited on : 2016-02-03
Resolution : 4.04 Å(reported)
Based on PDB ID : ?

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

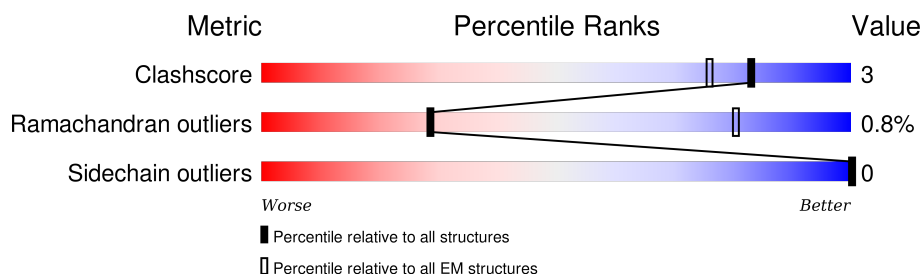
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 4.04 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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	1299	
1	B	1299	
1	C	1299	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 22683 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 Spike glycoprotein, envelope glycoprotein chimera.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	958	Total 7561	C 4833	N 1241	O 1447	S 40	0	0
1	B	958	Total 7561	C 4833	N 1241	O 1447	S 40	0	0
1	C	958	Total 7561	C 4833	N 1241	O 1447	S 40	0	0

There are 39 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	752	GLY	ARG	engineered mutation	UNP Q0ZME7
A	753	GLY	ARG	engineered mutation	UNP Q0ZME7
A	754	SER	LYS	engineered mutation	UNP Q0ZME7
A	755	GLY	ARG	engineered mutation	UNP Q0ZME7
A	756	SER	ARG	engineered mutation	UNP Q0ZME7
A	1277	GLY	-	linker	UNP Q0ZME7
A	1278	SER	-	linker	UNP Q0ZME7
A	1307	LEU	-	expression tag	UNP M1E1E4
A	1308	GLU	-	expression tag	UNP M1E1E4
A	1309	VAL	-	expression tag	UNP M1E1E4
A	1310	LEU	-	expression tag	UNP M1E1E4
A	1311	PHE	-	expression tag	UNP M1E1E4
A	1312	GLN	-	expression tag	UNP M1E1E4
B	752	GLY	ARG	engineered mutation	UNP Q0ZME7
B	753	GLY	ARG	engineered mutation	UNP Q0ZME7
B	754	SER	LYS	engineered mutation	UNP Q0ZME7
B	755	GLY	ARG	engineered mutation	UNP Q0ZME7
B	756	SER	ARG	engineered mutation	UNP Q0ZME7
B	1277	GLY	-	linker	UNP Q0ZME7
B	1278	SER	-	linker	UNP Q0ZME7
B	1307	LEU	-	expression tag	UNP M1E1E4
B	1308	GLU	-	expression tag	UNP M1E1E4
B	1309	VAL	-	expression tag	UNP M1E1E4
B	1310	LEU	-	expression tag	UNP M1E1E4

Continued on next page...

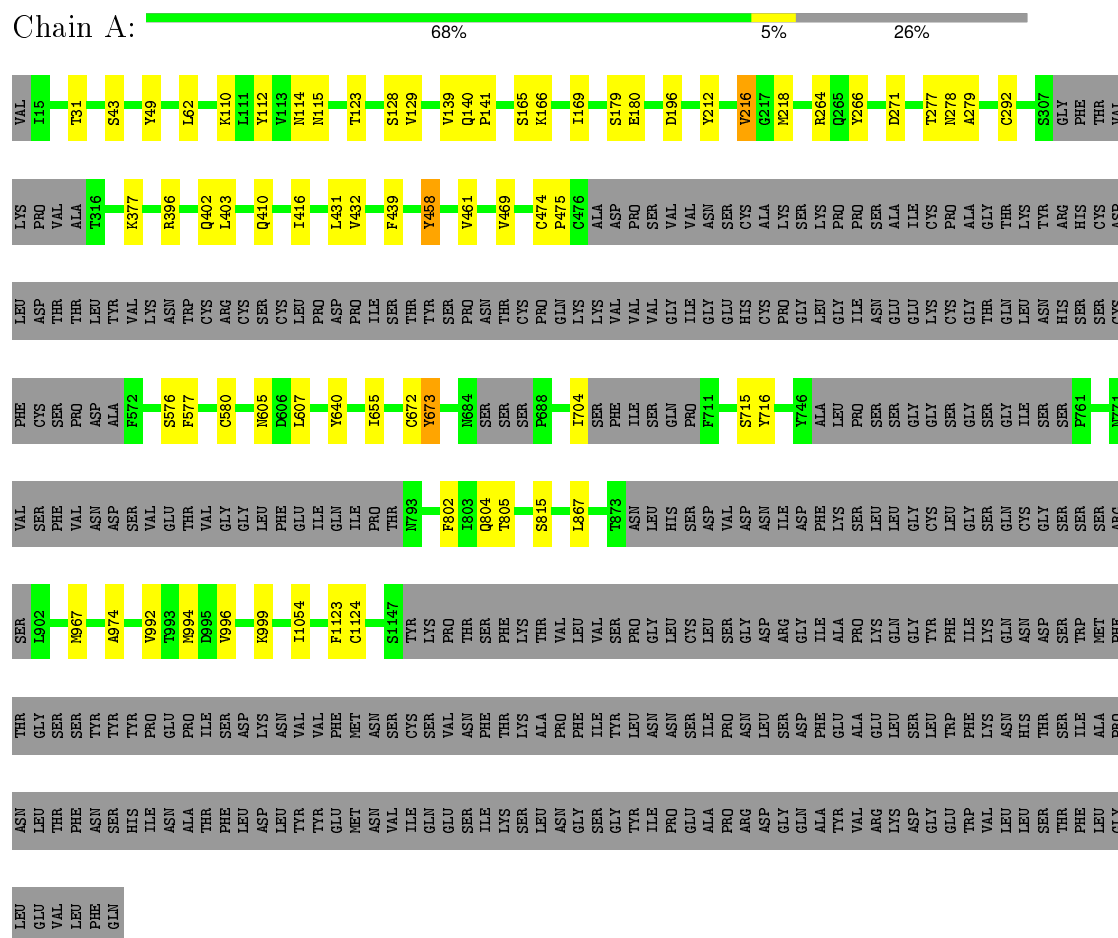
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1311	PHE	-	expression tag	UNP M1E1E4
B	1312	GLN	-	expression tag	UNP M1E1E4
C	752	GLY	ARG	engineered mutation	UNP Q0ZME7
C	753	GLY	ARG	engineered mutation	UNP Q0ZME7
C	754	SER	LYS	engineered mutation	UNP Q0ZME7
C	755	GLY	ARG	engineered mutation	UNP Q0ZME7
C	756	SER	ARG	engineered mutation	UNP Q0ZME7
C	1277	GLY	-	linker	UNP Q0ZME7
C	1278	SER	-	linker	UNP Q0ZME7
C	1307	LEU	-	expression tag	UNP M1E1E4
C	1308	GLU	-	expression tag	UNP M1E1E4
C	1309	VAL	-	expression tag	UNP M1E1E4
C	1310	LEU	-	expression tag	UNP M1E1E4
C	1311	PHE	-	expression tag	UNP M1E1E4
C	1312	GLN	-	expression tag	UNP M1E1E4

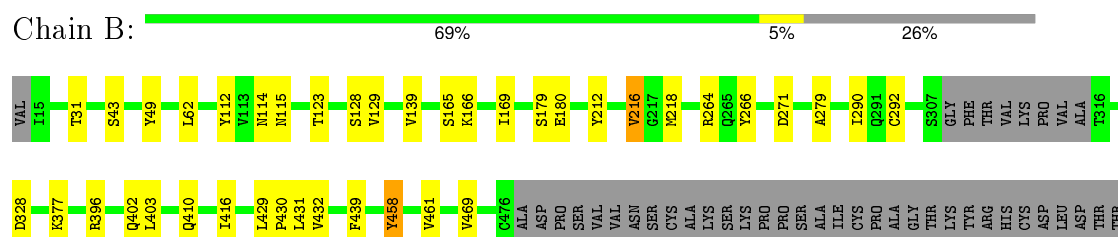
3 Residue-property plots

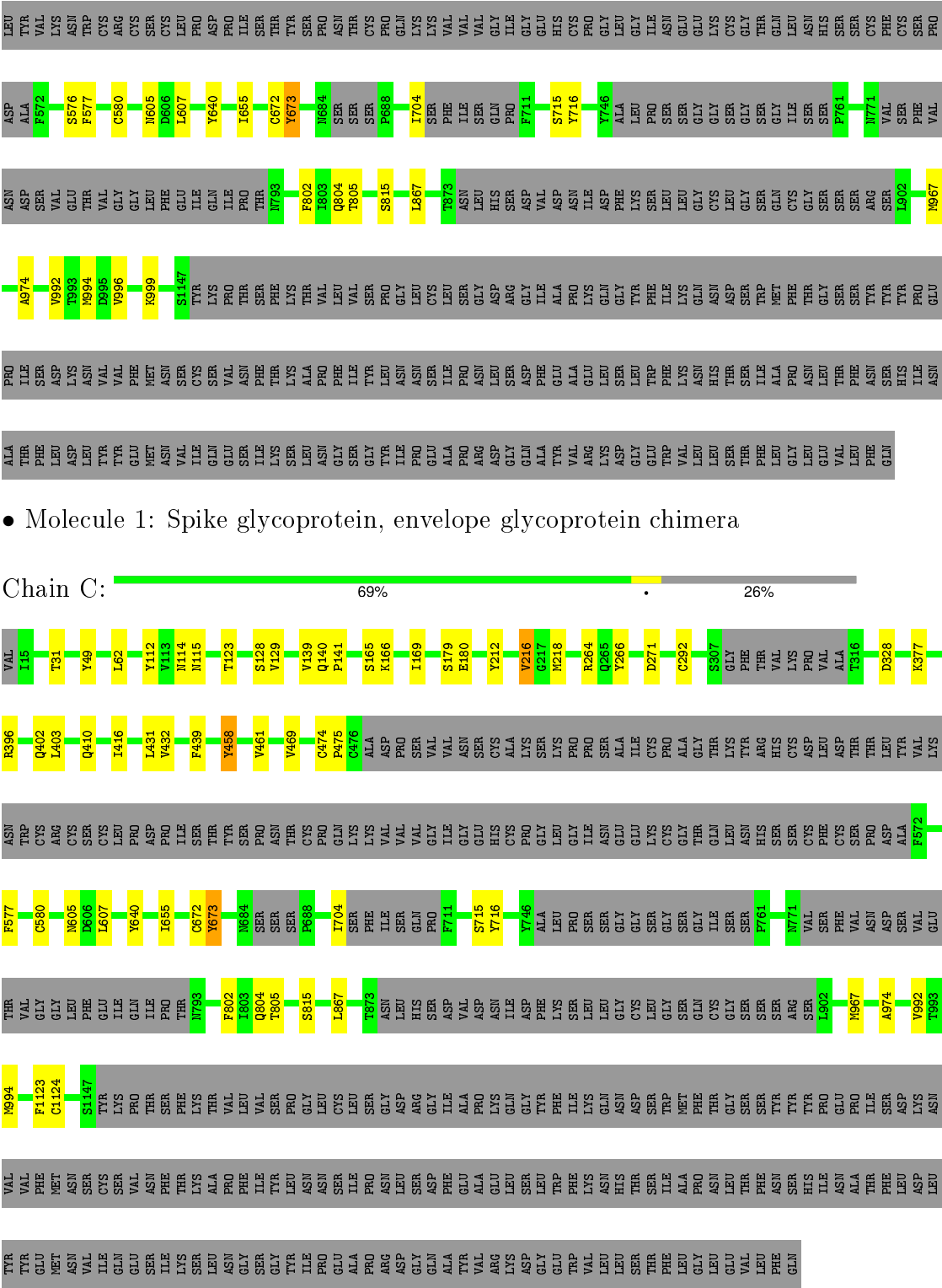
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric 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: Spike glycoprotein, envelope glycoprotein chimera



- Molecule 1: Spike glycoprotein, envelope glycoprotein chimera





4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	31435	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	Not provided	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	Not provided	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	22500	Depositor
Image detector	Not provided	Depositor

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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	0.76	0/7733	0.86	7/10518 (0.1%)
1	B	0.76	0/7733	0.86	7/10518 (0.1%)
1	C	0.76	0/7733	0.86	7/10518 (0.1%)
All	All	0.76	0/23199	0.86	21/31554 (0.1%)

There are no bond length outliers.

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	B	458	TYR	CB-CG-CD1	-7.68	116.39	121.00
1	C	458	TYR	CB-CG-CD1	-7.68	116.39	121.00
1	A	458	TYR	CB-CG-CD1	-7.67	116.39	121.00
1	B	266	TYR	CB-CG-CD2	-6.96	116.82	121.00
1	C	266	TYR	CB-CG-CD2	-6.95	116.83	121.00

There are no chirality outliers.

There are no planarity outliers.

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	7561	0	7292	48	0
1	B	7561	0	7292	44	0
1	C	7561	0	7292	43	0
All	All	22683	0	21876	135	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 3.

The worst 5 of 135 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:B:802:PHE:CD1	1:B:802:PHE:O	2.17	0.97
1:A:802:PHE:O	1:A:802:PHE:CD1	2.17	0.97
1:C:802:PHE:O	1:C:802:PHE:CD1	2.18	0.96
1:B:804:GLN:O	1:B:804:GLN:HG2	1.80	0.81
1:A:804:GLN:HG2	1:A:804:GLN:O	1.80	0.80

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	942/1299 (72%)	868 (92%)	66 (7%)	8 (1%)	24	69
1	B	942/1299 (72%)	868 (92%)	66 (7%)	8 (1%)	24	69
1	C	942/1299 (72%)	868 (92%)	66 (7%)	8 (1%)	24	69
All	All	2826/3897 (72%)	2604 (92%)	198 (7%)	24 (1%)	29	69

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	218	MET
1	A	673	TYR
1	B	218	MET
1	B	673	TYR
1	C	218	MET

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	860/1161 (74%)	860 (100%)	0	100	100
1	B	860/1161 (74%)	860 (100%)	0	100	100
1	C	860/1161 (74%)	860 (100%)	0	100	100
All	All	2580/3483 (74%)	2580 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	175	HIS
1	C	269	ASN
1	B	269	ASN
1	A	269	ASN
1	C	175	HIS

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.