



wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Aug 9, 2017 – 12:46 PM EDT

PDB ID : 5XLR
EMDB ID: : EMD-6732
Title : Structure of SARS-CoV spike glycoprotein
Authors : Gui, M.; Song, W.; Xiang, Y.; Wang, X.
Deposited on : unknown
Resolution : 3.80 Å(reported)

This is a wwPDB/EMDatabank EM Map/Model Validation Summary Report
for a publicly released PDB/EMDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029824

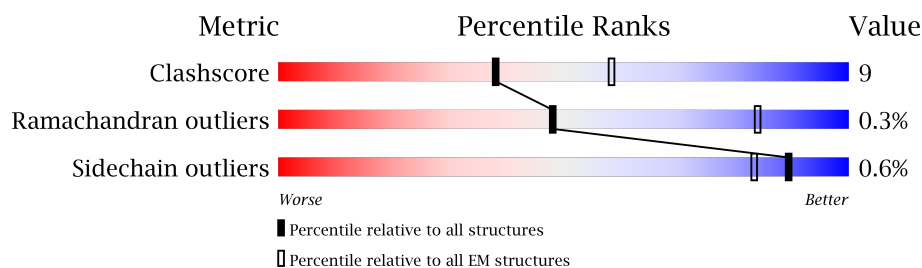
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.80 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	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

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	1203	66% 18% • 15%
1	B	1203	67% 18% • 15%
1	C	1203	67% 18% • 15%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 23901 atoms, of which 0 are hydrogens and 0 are deuteriums.

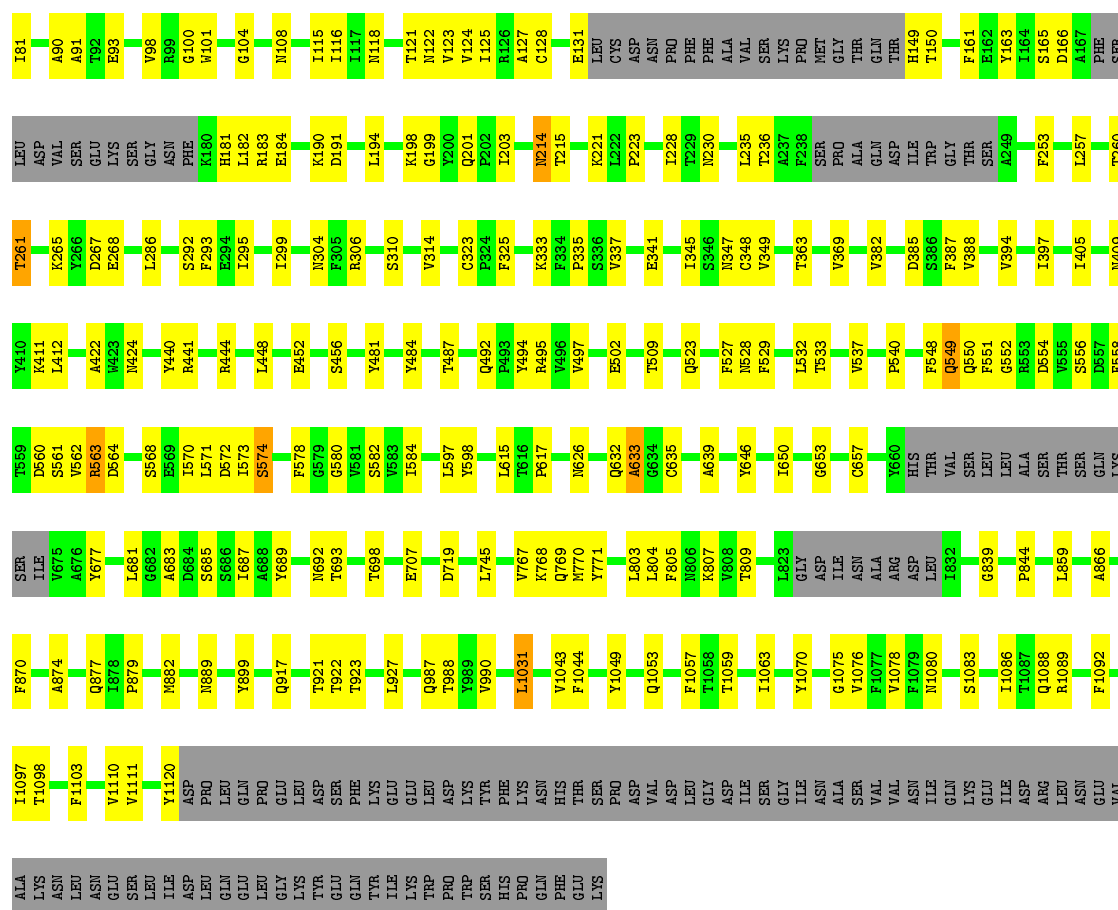
In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Spike glycoprotein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1022	Total	C	N	O	S	0	0
			7967	5093	1318	1512	44		
1	B	1022	Total	C	N	O	S	0	0
			7967	5093	1318	1512	44		
1	C	1022	Total	C	N	O	S	0	0
			7967	5093	1318	1512	44		

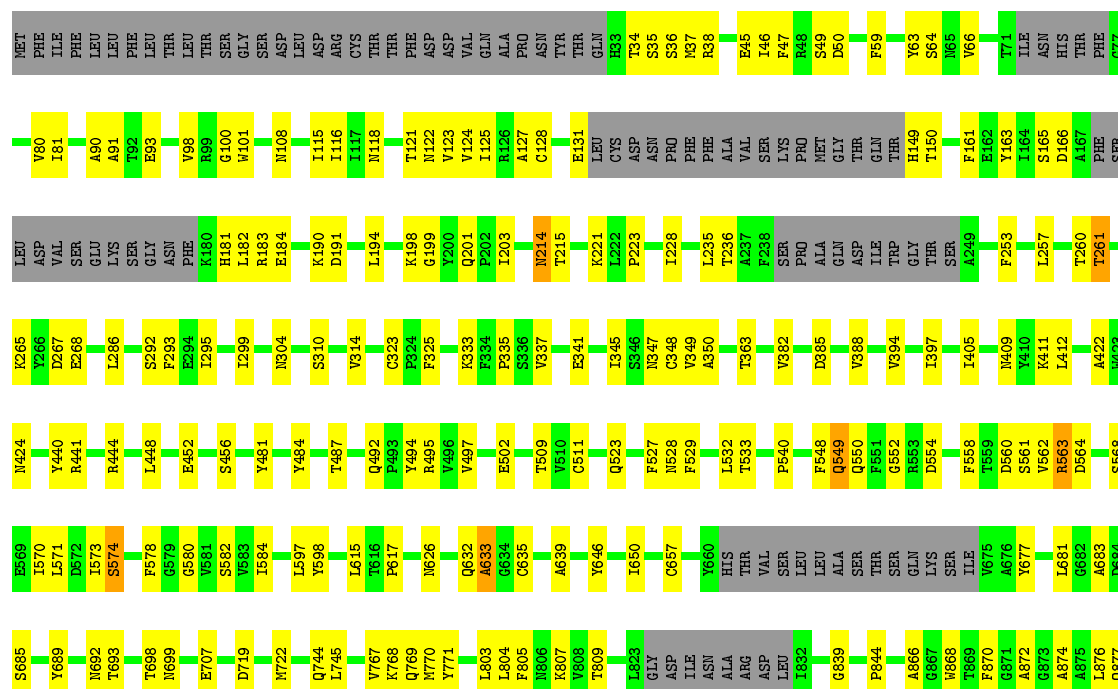
There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	667	ALA	ARG	engineered mutation	UNP P59594
A	1197	SER	-	expression tag	UNP P59594
A	1198	HIS	-	expression tag	UNP P59594
A	1199	PRO	-	expression tag	UNP P59594
A	1200	GLN	-	expression tag	UNP P59594
A	1201	PHE	-	expression tag	UNP P59594
A	1202	GLU	-	expression tag	UNP P59594
A	1203	LYS	-	expression tag	UNP P59594
B	667	ALA	ARG	engineered mutation	UNP P59594
B	1197	SER	-	expression tag	UNP P59594
B	1198	HIS	-	expression tag	UNP P59594
B	1199	PRO	-	expression tag	UNP P59594
B	1200	GLN	-	expression tag	UNP P59594
B	1201	PHE	-	expression tag	UNP P59594
B	1202	GLU	-	expression tag	UNP P59594
B	1203	LYS	-	expression tag	UNP P59594
C	667	ALA	ARG	engineered mutation	UNP P59594
C	1197	SER	-	expression tag	UNP P59594
C	1198	HIS	-	expression tag	UNP P59594
C	1199	PRO	-	expression tag	UNP P59594
C	1200	GLN	-	expression tag	UNP P59594
C	1201	PHE	-	expression tag	UNP P59594
C	1202	GLU	-	expression tag	UNP P59594
C	1203	LYS	-	expression tag	UNP P59594



• Molecule 1: Spike glycoprotein

Chain C: 67% 18% 15%



I1097	I1098	F1103	V1110	V1111	Y1120	ASP	PRO	LEU	GLN	PRO	GLY	LEU	LYS	TYR	GLU	SER	PHE	LYS	GLU	LEU	ASP	LYS	TYR	PHE	LYS	ASN	HIS	THR	SER	PRO	ASP	VAL	ASP	LEU	GLY	ASP	ILE	SER	GLY	ILE	ASN	ALA	SER	VAL	F1077	V1076	G1075	Y1070	I1063	F1057	T1059	T1058	Q1053	Y1049	F1044	V1043	L1031	V990	T988	Q987	L959	L927	T923	T922	T921	Q917	Q904	Y899	N896	N889	M882	P879	I878
ALA	LYS	ASN	ASN	GLU	SER	LEU	ILE	ASP	LEU	GLN	GLY	LEU	LYS	TYR	GLU	SER	PHE	LYS	GLU	LEU	ASP	LYS	TYR	PHE	LYS	ASN	HIS	THR	SER	PRO	ASP	VAL	ASP	LEU	GLY	ASP	ILE	SER	GLY	ILE	ASN	ALA	SER	VAL	F1077	V1076	G1075	Y1070	I1063	F1057	T1059	T1058	Q1053	Y1049	F1044	V1043	L1031	V990	T988	Q987	L959	L927	T923	T922	T921	Q917	Q904	Y899	N896	N889	M882	P879	I878

4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C3	Depositor
Number of particles used	42135	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 ($e^-/\text{\AA}^2$)	50	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	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	0.43	0/8151	0.64	1/11085 (0.0%)
1	B	0.43	0/8151	0.64	1/11085 (0.0%)
1	C	0.43	0/8151	0.64	1/11085 (0.0%)
All	All	0.43	0/24453	0.64	3/33255 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	8
1	B	0	8
1	C	0	8
All	All	0	24

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	1031	LEU	CA-CB-CG	5.73	128.48	115.30
1	C	1031	LEU	CA-CB-CG	5.69	128.39	115.30
1	A	1031	LEU	CA-CB-CG	5.68	128.36	115.30

There are no chirality outliers.

5 of 24 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	261	THR	Peptide
1	A	502	GLU	Mainchain,Peptide
1	A	548	PHE	Peptide
1	A	573	ILE	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	574	SER	Peptide

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	7967	0	7762	185	0
1	B	7967	0	7762	166	0
1	C	7967	0	7762	170	0
All	All	23901	0	23286	432	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 9.

The worst 5 of 432 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:194:LEU:O	1:A:223:PRO:HA	1.72	0.88
1:B:194:LEU:O	1:B:223:PRO:HA	1.72	0.88
1:C:194:LEU:O	1:C:223:PRO:HA	1.72	0.87
1:A:122:ASN:HA	1:A:166:ASP:O	1.74	0.87
1:B:34:THR:O	1:B:66:VAL:HB	1.75	0.87

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	1008/1203 (84%)	859 (85%)	146 (14%)	3 (0%)	44	81
1	B	1008/1203 (84%)	858 (85%)	147 (15%)	3 (0%)	44	81
1	C	1008/1203 (84%)	858 (85%)	147 (15%)	3 (0%)	44	81
All	All	3024/3609 (84%)	2575 (85%)	440 (15%)	9 (0%)	48	81

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	549	GLN
1	B	549	GLN
1	C	549	GLN
1	A	1092	PHE
1	B	1092	PHE

5.3.2 Protein sidechains ⓘ

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	881/1047 (84%)	876 (99%)	5 (1%)	89	95
1	B	881/1047 (84%)	876 (99%)	5 (1%)	89	95
1	C	881/1047 (84%)	876 (99%)	5 (1%)	89	95
All	All	2643/3141 (84%)	2628 (99%)	15 (1%)	89	95

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	323	CYS
1	B	333	LYS
1	C	333	LYS
1	B	214	ASN
1	C	323	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 45 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	409	ASN
1	B	759	ASN
1	C	889	ASN
1	B	437	ASN
1	B	769	GLN

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.