



wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 01:01 am BST

PDB ID : 5ZHY
Title : Structural characterization of the HCoV-229E fusion core
Authors : Zhang, W.; Zheng, Q.; Yan, M.; Chen, X.; Yang, H.; Zhou, W.; Rao, Z.
Deposited on : 2018-03-13
Resolution : 2.44 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

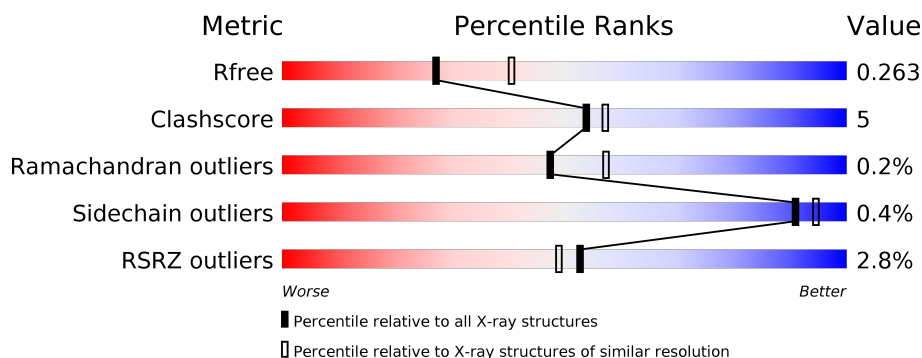
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.44 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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1564 (2.46-2.42)
Clashscore	141614	1631 (2.46-2.42)
Ramachandran outliers	138981	1617 (2.46-2.42)
Sidechain outliers	138945	1617 (2.46-2.42)
RSRZ outliers	127900	1547 (2.46-2.42)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. 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 $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	148	<div> <div>%</div> <div> <div></div> <div>47%</div> <div>14%</div> <div>39%</div> </div> </div>
1	B	148	<div> <div>3%</div> <div> <div></div> <div>54%</div> <div>6%</div> <div>40%</div> </div> </div>
1	C	148	<div> <div>3%</div> <div> <div></div> <div>55%</div> <div>5%</div> <div>40%</div> </div> </div>
1	D	148	<div> <div>%</div> <div> <div></div> <div>48%</div> <div>13%</div> <div>39%</div> </div> </div>
1	E	148	<div> <div></div> <div> <div>47%</div> <div>12%</div> <div>41%</div> </div> </div>
1	F	148	<div> <div>2%</div> <div> <div></div> <div>55%</div> <div>5%</div> <div>40%</div> </div> </div>

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 4039 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, 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, Spike glycoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	90	Total	C	N	O	S	0	0	0
			679	419	115	144	1			
1	B	89	Total	C	N	O	S	0	0	0
			671	416	113	141	1			
1	C	89	Total	C	N	O	S	0	0	0
			673	418	113	141	1			
1	D	90	Total	C	N	O	S	0	0	0
			681	421	116	143	1			
1	E	88	Total	C	N	O	S	0	0	0
			666	414	112	139	1			
1	F	89	Total	C	N	O	S	0	0	0
			669	414	112	142	1			

There are 162 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	69	LEU	-	linker	UNP P15423
A	70	VAL	-	linker	UNP P15423
A	71	PRO	-	linker	UNP P15423
A	72	ARG	-	linker	UNP P15423
A	73	GLY	-	linker	UNP P15423
A	74	SER	-	linker	UNP P15423
A	75	GLY	-	linker	UNP P15423
A	76	GLY	-	linker	UNP P15423
A	77	SER	-	linker	UNP P15423
A	78	GLY	-	linker	UNP P15423
A	79	GLY	-	linker	UNP P15423
A	80	SER	-	linker	UNP P15423
A	81	GLY	-	linker	UNP P15423
A	82	GLY	-	linker	UNP P15423
A	83	LEU	-	linker	UNP P15423
A	84	GLU	-	linker	UNP P15423
A	85	VAL	-	linker	UNP P15423

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	86	LEU	-	linker	UNP P15423
A	87	PHE	-	linker	UNP P15423
A	88	GLN	-	linker	UNP P15423
A	89	GLY	-	linker	UNP P15423
A	143	HIS	-	expression tag	UNP P15423
A	144	HIS	-	expression tag	UNP P15423
A	145	HIS	-	expression tag	UNP P15423
A	146	HIS	-	expression tag	UNP P15423
A	147	HIS	-	expression tag	UNP P15423
A	148	HIS	-	expression tag	UNP P15423
B	69	LEU	-	linker	UNP P15423
B	70	VAL	-	linker	UNP P15423
B	71	PRO	-	linker	UNP P15423
B	72	ARG	-	linker	UNP P15423
B	73	GLY	-	linker	UNP P15423
B	74	SER	-	linker	UNP P15423
B	75	GLY	-	linker	UNP P15423
B	76	GLY	-	linker	UNP P15423
B	77	SER	-	linker	UNP P15423
B	78	GLY	-	linker	UNP P15423
B	79	GLY	-	linker	UNP P15423
B	80	SER	-	linker	UNP P15423
B	81	GLY	-	linker	UNP P15423
B	82	GLY	-	linker	UNP P15423
B	83	LEU	-	linker	UNP P15423
B	84	GLU	-	linker	UNP P15423
B	85	VAL	-	linker	UNP P15423
B	86	LEU	-	linker	UNP P15423
B	87	PHE	-	linker	UNP P15423
B	88	GLN	-	linker	UNP P15423
B	89	GLY	-	linker	UNP P15423
B	143	HIS	-	expression tag	UNP P15423
B	144	HIS	-	expression tag	UNP P15423
B	145	HIS	-	expression tag	UNP P15423
B	146	HIS	-	expression tag	UNP P15423
B	147	HIS	-	expression tag	UNP P15423
B	148	HIS	-	expression tag	UNP P15423
C	69	LEU	-	linker	UNP P15423
C	70	VAL	-	linker	UNP P15423
C	71	PRO	-	linker	UNP P15423
C	72	ARG	-	linker	UNP P15423
C	73	GLY	-	linker	UNP P15423

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	74	SER	-	linker	UNP P15423
C	75	GLY	-	linker	UNP P15423
C	76	GLY	-	linker	UNP P15423
C	77	SER	-	linker	UNP P15423
C	78	GLY	-	linker	UNP P15423
C	79	GLY	-	linker	UNP P15423
C	80	SER	-	linker	UNP P15423
C	81	GLY	-	linker	UNP P15423
C	82	GLY	-	linker	UNP P15423
C	83	LEU	-	linker	UNP P15423
C	84	GLU	-	linker	UNP P15423
C	85	VAL	-	linker	UNP P15423
C	86	LEU	-	linker	UNP P15423
C	87	PHE	-	linker	UNP P15423
C	88	GLN	-	linker	UNP P15423
C	89	GLY	-	linker	UNP P15423
C	143	HIS	-	expression tag	UNP P15423
C	144	HIS	-	expression tag	UNP P15423
C	145	HIS	-	expression tag	UNP P15423
C	146	HIS	-	expression tag	UNP P15423
C	147	HIS	-	expression tag	UNP P15423
C	148	HIS	-	expression tag	UNP P15423
D	69	LEU	-	linker	UNP P15423
D	70	VAL	-	linker	UNP P15423
D	71	PRO	-	linker	UNP P15423
D	72	ARG	-	linker	UNP P15423
D	73	GLY	-	linker	UNP P15423
D	74	SER	-	linker	UNP P15423
D	75	GLY	-	linker	UNP P15423
D	76	GLY	-	linker	UNP P15423
D	77	SER	-	linker	UNP P15423
D	78	GLY	-	linker	UNP P15423
D	79	GLY	-	linker	UNP P15423
D	80	SER	-	linker	UNP P15423
D	81	GLY	-	linker	UNP P15423
D	82	GLY	-	linker	UNP P15423
D	83	LEU	-	linker	UNP P15423
D	84	GLU	-	linker	UNP P15423
D	85	VAL	-	linker	UNP P15423
D	86	LEU	-	linker	UNP P15423
D	87	PHE	-	linker	UNP P15423
D	88	GLN	-	linker	UNP P15423

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	89	GLY	-	linker	UNP P15423
D	143	HIS	-	expression tag	UNP P15423
D	144	HIS	-	expression tag	UNP P15423
D	145	HIS	-	expression tag	UNP P15423
D	146	HIS	-	expression tag	UNP P15423
D	147	HIS	-	expression tag	UNP P15423
D	148	HIS	-	expression tag	UNP P15423
E	69	LEU	-	linker	UNP P15423
E	70	VAL	-	linker	UNP P15423
E	71	PRO	-	linker	UNP P15423
E	72	ARG	-	linker	UNP P15423
E	73	GLY	-	linker	UNP P15423
E	74	SER	-	linker	UNP P15423
E	75	GLY	-	linker	UNP P15423
E	76	GLY	-	linker	UNP P15423
E	77	SER	-	linker	UNP P15423
E	78	GLY	-	linker	UNP P15423
E	79	GLY	-	linker	UNP P15423
E	80	SER	-	linker	UNP P15423
E	81	GLY	-	linker	UNP P15423
E	82	GLY	-	linker	UNP P15423
E	83	LEU	-	linker	UNP P15423
E	84	GLU	-	linker	UNP P15423
E	85	VAL	-	linker	UNP P15423
E	86	LEU	-	linker	UNP P15423
E	87	PHE	-	linker	UNP P15423
E	88	GLN	-	linker	UNP P15423
E	89	GLY	-	linker	UNP P15423
E	143	HIS	-	expression tag	UNP P15423
E	144	HIS	-	expression tag	UNP P15423
E	145	HIS	-	expression tag	UNP P15423
E	146	HIS	-	expression tag	UNP P15423
E	147	HIS	-	expression tag	UNP P15423
E	148	HIS	-	expression tag	UNP P15423
F	69	LEU	-	linker	UNP P15423
F	70	VAL	-	linker	UNP P15423
F	71	PRO	-	linker	UNP P15423
F	72	ARG	-	linker	UNP P15423
F	73	GLY	-	linker	UNP P15423
F	74	SER	-	linker	UNP P15423
F	75	GLY	-	linker	UNP P15423
F	76	GLY	-	linker	UNP P15423

Continued on next page...

Continued from previous page...

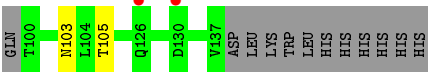
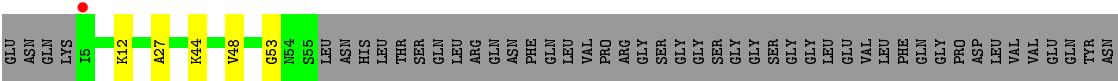
Chain	Residue	Modelled	Actual	Comment	Reference
F	77	SER	-	linker	UNP P15423
F	78	GLY	-	linker	UNP P15423
F	79	GLY	-	linker	UNP P15423
F	80	SER	-	linker	UNP P15423
F	81	GLY	-	linker	UNP P15423
F	82	GLY	-	linker	UNP P15423
F	83	LEU	-	linker	UNP P15423
F	84	GLU	-	linker	UNP P15423
F	85	VAL	-	linker	UNP P15423
F	86	LEU	-	linker	UNP P15423
F	87	PHE	-	linker	UNP P15423
F	88	GLN	-	linker	UNP P15423
F	89	GLY	-	linker	UNP P15423
F	143	HIS	-	expression tag	UNP P15423
F	144	HIS	-	expression tag	UNP P15423
F	145	HIS	-	expression tag	UNP P15423
F	146	HIS	-	expression tag	UNP P15423
F	147	HIS	-	expression tag	UNP P15423
F	148	HIS	-	expression tag	UNP P15423



● Molecule 1: Spike glycoprotein, Spike glycoprotein



● Molecule 1: Spike glycoprotein, Spike glycoprotein



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	30.11Å 77.35Å 77.39Å 76.89° 89.78° 90.20°	Depositor
Resolution (Å)	31.01 – 2.44 31.01 – 2.44	Depositor EDS
% Data completeness (in resolution range)	97.0 (31.01-2.44) 97.4 (31.01-2.44)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.67 (at 2.45Å)	Xtriage
Refinement program	PHENIX 1.8.4 _1496	Depositor
R, R_{free}	0.224 , 0.263 0.232 , 0.263	Depositor DCC
R_{free} test set	2018 reflections (8.20%)	wwPDB-VP
Wilson B-factor (Å ²)	48.9	Xtriage
Anisotropy	0.475	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 36.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.149 for h,-k,-l 0.147 for -h,-l,-k 0.447 for -h,l,k	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	4039	wwPDB-VP
Average B, all atoms (Å ²)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.43% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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	$\# Z > 5$	RMSZ	$\# Z > 5$
1	A	0.44	0/680	0.59	0/924
1	B	0.41	0/672	0.55	0/913
1	C	0.40	0/674	0.54	0/916
1	D	0.45	0/683	0.61	0/929
1	E	0.43	0/667	0.55	0/906
1	F	0.41	0/670	0.53	0/912
All	All	0.42	0/4046	0.57	0/5500

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	679	0	691	13	0
1	B	671	0	687	8	0
1	C	673	0	691	7	0
1	D	681	0	689	14	0
1	E	666	0	684	9	0
1	F	669	0	681	5	0
All	All	4039	0	4123	43	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 5.

The worst 5 of 43 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:101:ILE:HG21	1:D:51:GLN:HG2	1.68	0.76
1:C:47:ASP:O	1:C:51:GLN:HG3	1.85	0.76
1:A:47:ASP:O	1:A:51:GLN:HG3	1.93	0.68
1:B:51:GLN:HE22	1:C:100:THR:N	1.99	0.61
1:E:122:VAL:O	1:E:126:GLN:HG3	2.03	0.59

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 X-ray entries followed by that with respect to entries of similar resolution.

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	86/148 (58%)	85 (99%)	1 (1%)	0	100	100
1	B	85/148 (57%)	85 (100%)	0	0	100	100
1	C	85/148 (57%)	85 (100%)	0	0	100	100
1	D	86/148 (58%)	84 (98%)	2 (2%)	0	100	100
1	E	84/148 (57%)	84 (100%)	0	0	100	100
1	F	85/148 (57%)	84 (99%)	0	1 (1%)	13	13
All	All	511/888 (58%)	507 (99%)	3 (1%)	1 (0%)	47	57

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	53	GLY

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 X-ray entries followed by that with respect to entries of similar resolution.

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	79/129 (61%)	78 (99%)	1 (1%)	69	80
1	B	78/129 (60%)	78 (100%)	0	100	100
1	C	78/129 (60%)	78 (100%)	0	100	100
1	D	79/129 (61%)	78 (99%)	1 (1%)	69	80
1	E	77/129 (60%)	77 (100%)	0	100	100
1	F	78/129 (60%)	78 (100%)	0	100	100
All	All	469/774 (61%)	467 (100%)	2 (0%)	91	94

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	109	SER
1	D	109	SER

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

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.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	90/148 (60%)	0.54	2 (2%) 62 58	46, 58, 85, 102	0
1	B	89/148 (60%)	0.47	5 (5%) 24 20	43, 61, 77, 86	0
1	C	89/148 (60%)	0.47	4 (4%) 33 30	41, 59, 73, 103	0
1	D	90/148 (60%)	0.23	1 (1%) 80 79	46, 57, 83, 111	0
1	E	88/148 (59%)	0.21	0 100 100	42, 61, 75, 81	0
1	F	89/148 (60%)	0.41	3 (3%) 45 42	41, 58, 75, 84	0
All	All	535/888 (60%)	0.39	15 (2%) 53 49	41, 59, 79, 111	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	132	ILE	3.7
1	B	54	ASN	3.6
1	F	5	ILE	3.6
1	F	130	ASP	3.4
1	B	5	ILE	2.9

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.