Full wwPDB X-ray Structure Validation Report ⓘ

Feb 10, 2025 – 10:12 AM EST

PDB ID	:	8VYA
Title	:	SARS-CoV-2 Omicron Variant Spike Glycoprotein Fusion Core (Q954H)
Authors	:	Outlaw, V.K.; Apurba, R.; Vithanage, N.
Deposited on	:	2024-02-07
Resolution	:	2.12 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

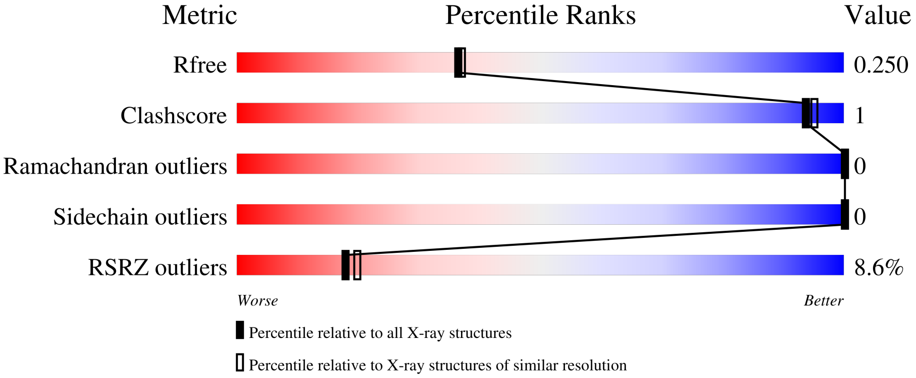
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

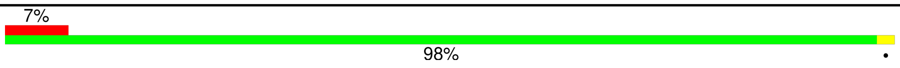
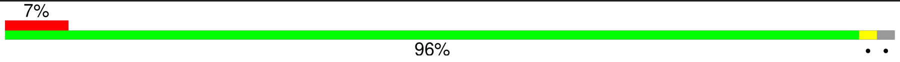
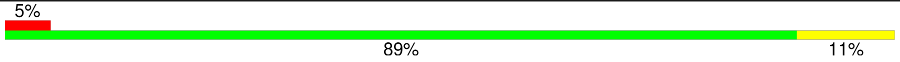
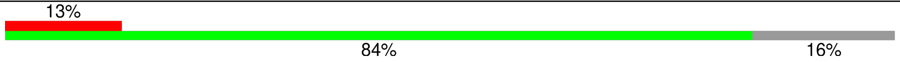
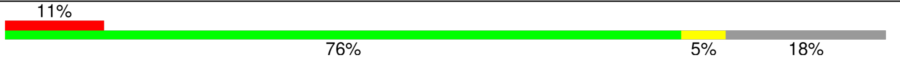
The reported resolution of this entry is 2.12 Å.

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.



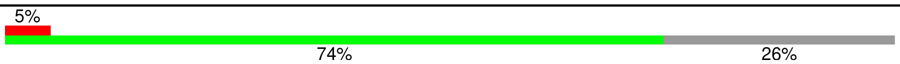
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	7689 (2.14-2.10)
Clashscore	180529	8431 (2.14-2.10)
Ramachandran outliers	177936	8366 (2.14-2.10)
Sidechain outliers	177891	8367 (2.14-2.10)
RSRZ outliers	164620	7689 (2.14-2.10)

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 of 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 <=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	57	
1	B	57	
1	C	57	
2	D	38	
2	E	38	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	38	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4022 atoms, of which 1998 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 SARS-CoV-2 Omicron variant spike glycoprotein N-terminal heptad repeat domain (Q954H).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	57	Total 840	C 258	H 422	N 75	O 85	0	0	1
1	B	56	Total 841	C 258	H 425	N 75	O 83	0	0	1
1	C	57	Total 840	C 258	H 422	N 75	O 85	0	0	1

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	911	ACE	-	acetylation	UNP P0DTC2
A	954	HIS	GLN	engineered mutation	UNP P0DTC2
A	967	NH2	-	amidation	UNP P0DTC2
B	911	ACE	-	acetylation	UNP P0DTC2
B	954	HIS	GLN	engineered mutation	UNP P0DTC2
B	967	NH2	-	amidation	UNP P0DTC2
C	911	ACE	-	acetylation	UNP P0DTC2
C	954	HIS	GLN	engineered mutation	UNP P0DTC2
C	967	NH2	-	amidation	UNP P0DTC2

- Molecule 2 is a protein called SARS-CoV-2 Omicron variant spike glycoprotein C-terminal heptad repeat domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	32	Total 517	C 157	H 263	N 44	O 53	0	0	0
2	E	31	Total 490	C 149	H 248	N 43	O 50	0	0	0
2	F	28	Total 435	C 132	H 218	N 39	O 46	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1167	ACE	-	acetylation	UNP P0DTC2
D	1204	NH2	-	amidation	UNP P0DTC2
E	1167	ACE	-	acetylation	UNP P0DTC2
E	1204	NH2	-	amidation	UNP P0DTC2
F	1167	ACE	-	acetylation	UNP P0DTC2
F	1204	NH2	-	amidation	UNP P0DTC2

- Molecule 3 is water.

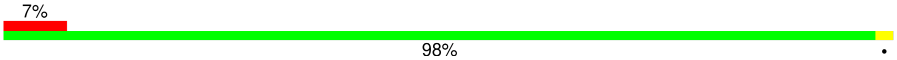
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	7	Total O 7 7	0	0
3	B	11	Total O 11 11	0	0
3	C	12	Total O 12 12	0	0
3	D	6	Total O 6 6	0	0
3	E	12	Total O 12 12	0	0
3	F	11	Total O 11 11	0	0

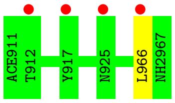
3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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 and electron density. 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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). 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: SARS-CoV-2 Omicron variant spike glycoprotein N-terminal heptad repeat domain (Q954H)

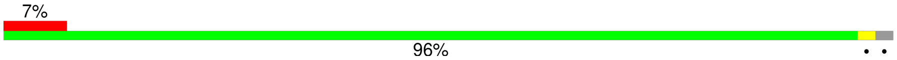
Chain A:

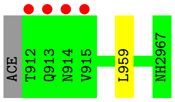




- Molecule 1: SARS-CoV-2 Omicron variant spike glycoprotein N-terminal heptad repeat domain (Q954H)

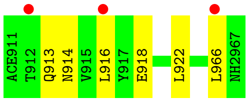
Chain B:





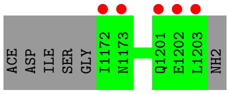
- Molecule 1: SARS-CoV-2 Omicron variant spike glycoprotein N-terminal heptad repeat domain (Q954H)

Chain C:



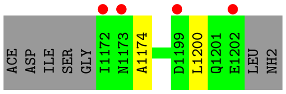
- Molecule 2: SARS-CoV-2 Omicron variant spike glycoprotein C-terminal heptad repeat domain

Chain D:



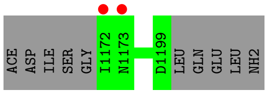
- Molecule 2: SARS-CoV-2 Omicron variant spike glycoprotein C-terminal heptad repeat domain

Chain E:



- Molecule 2: SARS-CoV-2 Omicron variant spike glycoprotein C-terminal heptad repeat domain

Chain F:



4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	51.38Å 56.69Å 84.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.88 – 2.12 43.88 – 2.12	Depositor EDS
% Data completeness (in resolution range)	85.4 (43.88-2.12) 78.5 (43.88-2.12)	Depositor EDS
R_{merge}	0.21	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.24 (at 2.12Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.216 , 0.250 0.215 , 0.250	Depositor DCC
R_{free} test set	13301 reflections (9.93%)	wwPDB-VP
Wilson B-factor (Å^2)	23.6	Xtriage
Anisotropy	0.078	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 53.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.36$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4022	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.10% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: NH2, ACE

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	0.28	0/417	0.39	0/566
1	B	0.30	0/417	0.41	0/564
1	C	0.30	0/417	0.40	0/566
2	D	0.24	0/253	0.42	0/341
2	E	0.27	0/241	0.46	0/325
2	F	0.29	0/216	0.43	0/291
All	All	0.28	0/1961	0.41	0/2653

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 (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	418	422	420	1	0
1	B	416	425	423	1	0
1	C	418	422	420	4	0
2	D	254	263	262	0	0
2	E	242	248	247	2	0
2	F	217	218	217	0	0
3	A	7	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	11	0	0	0	0
3	C	12	0	0	0	0
3	D	6	0	0	0	0
3	E	12	0	0	0	0
3	F	11	0	0	0	0
All	All	2024	1998	1989	5	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 (5) 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:C:913:GLN:HA	1:C:916:LEU:HD12	1.85	0.58
1:A:966:LEU:HD21	1:C:966:LEU:HD13	1.94	0.49
1:C:922:LEU:HD23	2:E:1200:LEU:HD23	1.97	0.46
1:B:959:LEU:HD23	2:E:1174:ALA:HB2	1.99	0.45
1:C:914:ASN:O	1:C:918:GLU:HG2	2.22	0.40

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 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	55/57 (96%)	55 (100%)	0	0	100	100
1	B	54/57 (95%)	54 (100%)	0	0	100	100
1	C	55/57 (96%)	55 (100%)	0	0	100	100
2	D	30/38 (79%)	30 (100%)	0	0	100	100
2	E	29/38 (76%)	29 (100%)	0	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	F	26/38 (68%)	26 (100%)	0	0	100	100
All	All	249/285 (87%)	249 (100%)	0	0	100	100

There are no Ramachandran outliers to report.

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	46/47 (98%)	46 (100%)	0	100	100
1	B	46/47 (98%)	46 (100%)	0	100	100
1	C	46/47 (98%)	46 (100%)	0	100	100
2	D	30/33 (91%)	30 (100%)	0	100	100
2	E	28/33 (85%)	28 (100%)	0	100	100
2	F	25/33 (76%)	25 (100%)	0	100	100
All	All	221/240 (92%)	221 (100%)	0	100	100

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

Sometimes 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 oligosaccharides 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

6.1 Protein, DNA and RNA chains

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	55/57 (96%)	0.58	4 (7%)	22	25	25, 40, 76, 95	0
1	B	55/57 (96%)	0.43	4 (7%)	22	25	24, 38, 77, 91	0
1	C	55/57 (96%)	0.39	3 (5%)	32	35	23, 37, 74, 80	0
2	D	32/38 (84%)	0.52	5 (15%)	6	7	28, 41, 97, 101	0
2	E	31/38 (81%)	0.29	4 (12%)	9	10	24, 36, 67, 87	0
2	F	28/38 (73%)	0.43	2 (7%)	23	26	28, 38, 62, 86	0
All	All	256/285 (89%)	0.45	22 (8%)	18	20	23, 39, 80, 101	0

All (22) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	1172	ILE	5.5
1	B	914	ASN	4.6
2	D	1172	ILE	4.4
1	A	917	TYR	3.8
2	E	1172	ILE	3.2
1	B	912	THR	3.0
1	B	915	VAL	2.9
1	C	912	THR	2.9
1	A	966	LEU	2.9
1	B	913	GLN	2.7
2	E	1202	GLU	2.6
2	D	1203	LEU	2.5
1	A	925	ASN	2.4
1	C	916	LEU	2.4
2	D	1173	ASN	2.3
2	E	1199	ASP	2.3
2	F	1173	ASN	2.2
1	A	912	THR	2.2
2	D	1201	GLN	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	D	1202	GLU	2.2
2	E	1173	ASN	2.1
1	C	966	LEU	2.1

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.