



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 6, 2025 – 04:21 PM JST

PDB ID : 8YDU  
Title : Crystal structure of the receptor binding domain of SARS-CoV-2 Omicron BA.2 variant spike protein in complex with CeSPIACE  
Authors : Nakamura, S.; Numoto, N.; Fujiyoshi, Y.  
Deposited on : 2024-02-21  
Resolution : 1.70 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](mailto: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	:	1.8.5 (274361), CSD as541be (2020)
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

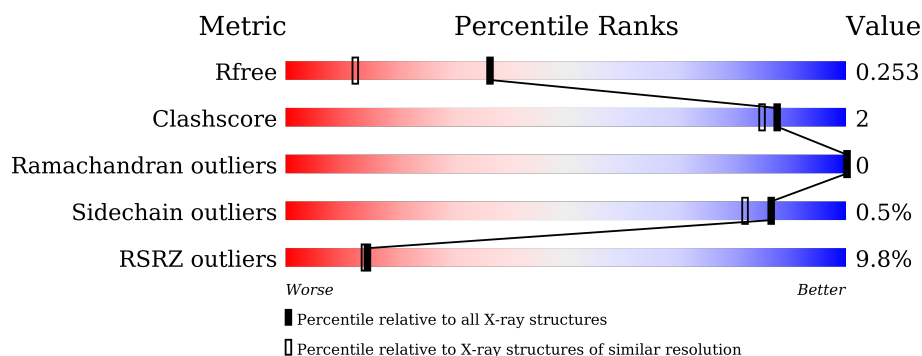
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.70 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	5161 (1.70-1.70)
Clashscore	180529	5671 (1.70-1.70)
Ramachandran outliers	177936	5594 (1.70-1.70)
Sidechain outliers	177891	5594 (1.70-1.70)
RSRZ outliers	164620	5159 (1.70-1.70)

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  $\geq 3$ , 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	39	<div> <div>3%</div> <div> <div></div> <div>95%</div> <div>5%</div> </div> </div>
2	B	230	<div> <div>10%</div> <div> <div></div> <div>80%</div> <div>5%</div> <div>15%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2025 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 SARS-CoV-2 inhibiting peptide CeSPIACE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	39	Total	C	N	O	S	0	0	0
			327	211	52	61	3			

- Molecule 2 is a protein called Spike protein S1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	196	Total	C	N	O	S	0	0	0
			1538	987	262	281	8			

There are 51 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	339	ASP	GLY	variant	UNP P0DTC2
B	371	PHE	SER	variant	UNP P0DTC2
B	373	PRO	SER	variant	UNP P0DTC2
B	376	ALA	THR	variant	UNP P0DTC2
B	405	ASN	ASP	variant	UNP P0DTC2
B	408	SER	ARG	variant	UNP P0DTC2
B	417	ASN	LYS	variant	UNP P0DTC2
B	440	LYS	ASN	variant	UNP P0DTC2
B	477	ASN	SER	variant	UNP P0DTC2
B	478	LYS	THR	variant	UNP P0DTC2
B	484	ALA	GLU	variant	UNP P0DTC2
B	493	ARG	GLN	variant	UNP P0DTC2
B	498	ARG	GLN	variant	UNP P0DTC2
B	501	TYR	ASN	variant	UNP P0DTC2
B	505	HIS	TYR	variant	UNP P0DTC2
B	527	SER	-	expression tag	UNP P0DTC2
B	528	ASN	-	expression tag	UNP P0DTC2
B	529	SER	-	expression tag	UNP P0DTC2
B	530	GLU	-	expression tag	UNP P0DTC2
B	531	ASN	-	expression tag	UNP P0DTC2

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	532	LEU	-	expression tag	UNP P0DTC2
B	533	TYR	-	expression tag	UNP P0DTC2
B	534	PHE	-	expression tag	UNP P0DTC2
B	535	GLN	-	expression tag	UNP P0DTC2
B	536	GLY	-	expression tag	UNP P0DTC2
B	537	SER	-	expression tag	UNP P0DTC2
B	538	HIS	-	expression tag	UNP P0DTC2
B	539	HIS	-	expression tag	UNP P0DTC2
B	540	HIS	-	expression tag	UNP P0DTC2
B	541	HIS	-	expression tag	UNP P0DTC2
B	542	HIS	-	expression tag	UNP P0DTC2
B	543	HIS	-	expression tag	UNP P0DTC2
B	544	HIS	-	expression tag	UNP P0DTC2
B	545	HIS	-	expression tag	UNP P0DTC2
B	546	HIS	-	expression tag	UNP P0DTC2
B	547	HIS	-	expression tag	UNP P0DTC2
B	548	GLY	-	expression tag	UNP P0DTC2
B	549	LEU	-	expression tag	UNP P0DTC2
B	550	ASN	-	expression tag	UNP P0DTC2
B	551	ASP	-	expression tag	UNP P0DTC2
B	552	ILE	-	expression tag	UNP P0DTC2
B	553	PHE	-	expression tag	UNP P0DTC2
B	554	GLU	-	expression tag	UNP P0DTC2
B	555	ALA	-	expression tag	UNP P0DTC2
B	556	GLN	-	expression tag	UNP P0DTC2
B	557	LYS	-	expression tag	UNP P0DTC2
B	558	ILE	-	expression tag	UNP P0DTC2
B	559	GLU	-	expression tag	UNP P0DTC2
B	560	TRP	-	expression tag	UNP P0DTC2
B	561	HIS	-	expression tag	UNP P0DTC2
B	562	GLU	-	expression tag	UNP P0DTC2

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	24	Total	O	0	0
			24	24		
4	B	130	Total	O	0	0
			130	130		

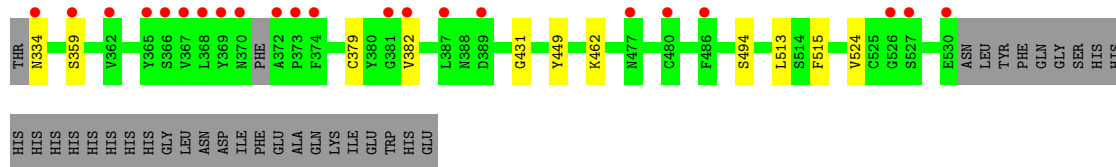
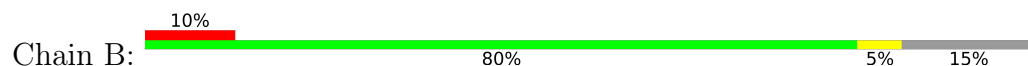
### 3 Residue-property plots [i](#)

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 inhibiting peptide CeSPIACE



- Molecule 2: Spike protein S1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	74.26Å 74.26Å 99.76Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	39.41 – 1.70 39.41 – 1.70	Depositor EDS
% Data completeness (in resolution range)	99.3 (39.41-1.70) 99.3 (39.41-1.70)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.41 (at 1.70Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, $R_{free}$	0.220 , 0.252 0.220 , 0.253	Depositor DCC
$R_{free}$ test set	1782 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	32.1	Xtriage
Anisotropy	0.601	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 38.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.032 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2025	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.70% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>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

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

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	0.32	0/332	0.49	0/442
2	B	0.36	0/1581	0.55	0/2150
All	All	0.35	0/1913	0.54	0/2592

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	327	0	328	1	0
2	B	1538	0	1452	6	0
3	B	6	0	8	0	0
4	A	24	0	0	0	0
4	B	130	0	0	0	0
All	All	2025	0	1788	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 2.

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)
2:B:379:CYS:HB3	2:B:382:VAL:HG23	1.95	0.48
2:B:431:GLY:HA2	2:B:515:PHE:HD2	1.81	0.46
2:B:449:TYR:HD2	2:B:494:SER:HB2	1.81	0.45
2:B:359:SER:HA	2:B:524:VAL:CG2	2.47	0.44
2:B:431:GLY:HA3	2:B:513:LEU:O	2.18	0.43
1:A:2:LYS:HG3	1:A:35:PHE:CZ	2.54	0.43
2:B:462:LYS:HB2	2:B:462:LYS:HE3	1.77	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 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	37/39 (95%)	36 (97%)	1 (3%)	0	100	100
2	B	192/230 (84%)	185 (96%)	7 (4%)	0	100	100
All	All	229/269 (85%)	221 (96%)	8 (4%)	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 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	35/36 (97%)	35 (100%)	0	100	100
2	B	163/199 (82%)	162 (99%)	1 (1%)	84	78

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	198/235 (84%)	197 (100%)	1 (0%)	86	82

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

Mol	Chain	Res	Type
2	B	334	ASN

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 ⓘ

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	GOL	B	601	-	5,5,5	0.94	0	5,5,5	1.01	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the

Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GOL	B	601	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	39/39 (100%)	0.82	1 (2%) 57 60	32, 41, 59, 84	0
2	B	196/230 (85%)	0.83	22 (11%) 11 10	28, 41, 67, 78	0
All	All	235/269 (87%)	0.83	23 (9%) 14 14	28, 41, 67, 84	0

All (23) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	372	ALA	6.3
2	B	382	VAL	3.7
2	B	369	TYR	3.6
2	B	480	CYS	3.6
2	B	366	SER	3.2
2	B	373	PRO	3.2
2	B	365	TYR	3.0
2	B	334	ASN	3.0
2	B	381	GLY	2.9
2	B	370	ASN	2.6
2	B	362	VAL	2.6
2	B	477	ASN	2.5
1	A	3	LEU	2.5
2	B	486	PHE	2.5
2	B	526	GLY	2.4
2	B	527	SER	2.4
2	B	367	VAL	2.3
2	B	389	ASP	2.3
2	B	368	LEU	2.3
2	B	387	LEU	2.3
2	B	374	PHE	2.2
2	B	359	SER	2.2
2	B	530	GLU	2.2

## 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 monosaccharides in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	GOL	B	601	6/6	0.90	0.12	44,48,50,54	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.