



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 21, 2023 – 01:01 AM JST

PDB ID : 7E7X  
Title : SARS-CoV-2 Spike Protein N terminal domain in Complex with N11 Fab  
Authors : Zhang, Z.; Shuo, D.; Xiao, J.  
Deposited on : 2021-02-28  
Resolution : 2.78 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](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  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
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) : 2.36

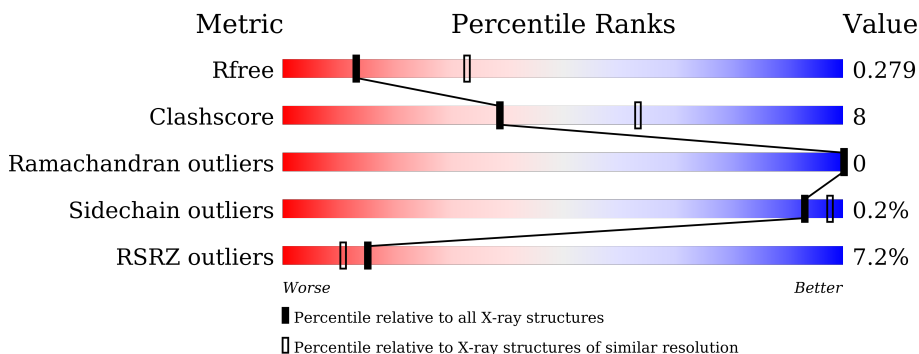
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.78 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	4107 (2.80-2.76)
Clashscore	141614	4575 (2.80-2.76)
Ramachandran outliers	138981	4487 (2.80-2.76)
Sidechain outliers	138945	4489 (2.80-2.76)
RSRZ outliers	127900	4027 (2.80-2.76)

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	297	
1	B	297	
2	H	234	
2	M	234	
3	L	216	
3	N	216	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9767 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 protein S1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	230	Total 1862	C 1212	N 301	O 344	S 5	0	0	0
1	B	206	Total 1675	C 1090	N 271	O 309	S 5	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	304	HIS	-	expression tag	UNP P0DTC2
A	305	HIS	-	expression tag	UNP P0DTC2
A	306	HIS	-	expression tag	UNP P0DTC2
A	307	HIS	-	expression tag	UNP P0DTC2
A	308	HIS	-	expression tag	UNP P0DTC2
A	309	HIS	-	expression tag	UNP P0DTC2
B	304	HIS	-	expression tag	UNP P0DTC2
B	305	HIS	-	expression tag	UNP P0DTC2
B	306	HIS	-	expression tag	UNP P0DTC2
B	307	HIS	-	expression tag	UNP P0DTC2
B	308	HIS	-	expression tag	UNP P0DTC2
B	309	HIS	-	expression tag	UNP P0DTC2

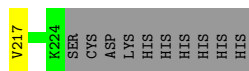
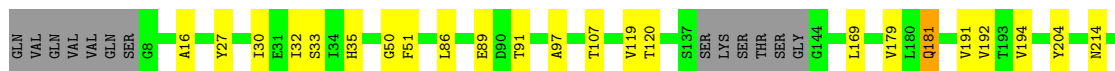
- Molecule 2 is a protein called N11 Fab heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	H	211	Total 1561	C 990	N 253	O 311	S 7	0	0	0
2	M	211	Total 1561	C 990	N 253	O 311	S 7	0	0	0

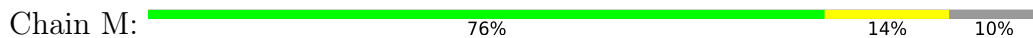
- Molecule 3 is a protein called N11 Fab Light chain.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>					<b>ZeroOcc</b>	<b>AltConf</b>	<b>Trace</b>
3	L	211	Total	C	N	O	S	0	0	0
			1554	975	254	320	5			
3	N	211	Total	C	N	O	S	0	0	0
			1554	975	254	320	5			

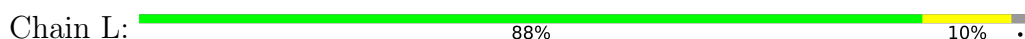




• Molecule 2: N11 Fab heavy chain



• Molecule 3: N11 Fab Light chain



• Molecule 3: N11 Fab Light chain



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	91.42Å 113.28Å 210.91Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.47 – 2.78 19.47 – 2.78	Depositor EDS
% Data completeness (in resolution range)	98.9 (19.47-2.78) 98.9 (19.47-2.78)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.24 (at 2.79Å)	Xtrriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, $R_{free}$	0.248 , 0.279 0.248 , 0.279	Depositor DCC
$R_{free}$ test set	1988 reflections (3.60%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	53.3	Xtrriage
Anisotropy	0.698	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 36.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	9767	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	59.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 48.34 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 8.7294e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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 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  > 5$	RMSZ	# $ Z  > 5$
1	A	0.53	3/1914 (0.2%)	0.63	3/2602 (0.1%)
1	B	0.35	0/1719	0.66	4/2334 (0.2%)
2	H	0.32	0/1599	0.59	2/2184 (0.1%)
2	M	0.46	2/1599 (0.1%)	0.64	3/2184 (0.1%)
3	L	0.32	0/1594	0.53	1/2180 (0.0%)
3	N	0.30	0/1594	0.51	1/2180 (0.0%)
All	All	0.40	5/10019 (0.0%)	0.60	14/13664 (0.1%)

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
2	H	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	M	23	LYS	CE-NZ	12.32	1.79	1.49
1	A	150	LYS	CD-CE	11.97	1.81	1.51
1	A	150	LYS	CE-NZ	10.23	1.74	1.49
1	A	156	GLU	CD-OE2	6.89	1.33	1.25
2	M	62	GLN	CG-CD	5.58	1.63	1.51

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	195	LYS	CD-CE-NZ	-10.07	88.54	111.70
2	M	23	LYS	CD-CE-NZ	-9.26	90.41	111.70
1	B	90	VAL	CG1-CB-CG2	7.96	123.63	110.90
2	H	181	GLN	CA-CB-CG	7.85	130.66	113.40
1	B	242	LEU	CB-CG-CD1	-7.69	97.92	111.00



There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	H	181	GLN	Sidechain

## 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	1862	0	1782	39	0
1	B	1675	0	1603	41	1
2	H	1561	0	1518	16	0
2	M	1561	0	1518	25	0
3	L	1554	0	1484	17	0
3	N	1554	0	1484	15	0
All	All	9767	0	9389	144	1

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 8.

The worst 5 of 144 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:150:LYS:CD	1:A:150:LYS:CE	1.81	1.57
1:A:150:LYS:CE	1:A:150:LYS:NZ	1.74	1.48
2:M:23:LYS:CE	2:M:23:LYS:NZ	1.79	1.41
1:A:150:LYS:CE	1:A:150:LYS:CG	2.44	0.95
2:M:23:LYS:NZ	2:M:23:LYS:CD	2.33	0.90

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:200:TYR:OH	1:B:228:ASP:OD1[2_655]	2.18	0.02

## 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	220/297 (74%)	204 (93%)	16 (7%)	0	100	100
1	B	192/297 (65%)	177 (92%)	15 (8%)	0	100	100
2	H	207/234 (88%)	198 (96%)	9 (4%)	0	100	100
2	M	207/234 (88%)	199 (96%)	8 (4%)	0	100	100
3	L	209/216 (97%)	202 (97%)	7 (3%)	0	100	100
3	N	209/216 (97%)	201 (96%)	8 (4%)	0	100	100
All	All	1244/1494 (83%)	1181 (95%)	63 (5%)	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	206/267 (77%)	206 (100%)	0	100	100
1	B	186/267 (70%)	185 (100%)	1 (0%)	88	95
2	H	172/196 (88%)	172 (100%)	0	100	100
2	M	172/196 (88%)	171 (99%)	1 (1%)	86	95
3	L	173/183 (94%)	173 (100%)	0	100	100
3	N	173/183 (94%)	173 (100%)	0	100	100
All	All	1082/1292 (84%)	1080 (100%)	2 (0%)	93	98

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

Mol	Chain	Res	Type
1	B	279	TYR
2	M	23	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	14	GLN
1	B	115	GLN
3	N	198	GLN
1	B	245	HIS
1	A	149	ASN

### 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 monosaccharides 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.

## 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, 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	230/297 (77%)	0.79	33 (14%) 2 1	46, 68, 96, 117	0
1	B	206/297 (69%)	1.40	56 (27%) 0 0	58, 84, 111, 126	0
2	H	211/234 (90%)	-0.10	0 100 100	35, 49, 65, 74	0
2	M	211/234 (90%)	-0.04	1 (0%) 91 90	37, 52, 70, 77	0
3	L	211/216 (97%)	-0.24	1 (0%) 91 90	31, 43, 58, 66	0
3	N	211/216 (97%)	-0.10	1 (0%) 91 90	37, 48, 60, 68	0
All	All	1280/1494 (85%)	0.29	92 (7%) 15 11	31, 54, 96, 126	0

The worst 5 of 92 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	28	TYR	9.5
1	B	18	LEU	6.2
1	B	280	ASN	6.1
1	B	268	GLY	5.7
1	A	98	SER	5.7

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.