



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 15, 2021 – 12:14 PM EST

PDB ID : 6X45  
Title : SARS-CoV2 spike glycoprotein N-terminal heptad repeat domain + SARS-CoV2(QEYKKEKE)  
Authors : Kreitler, D.F.; Outlaw, V.K.; Gellman, S.H.  
Deposited on : 2020-05-22  
Resolution : 2.20 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

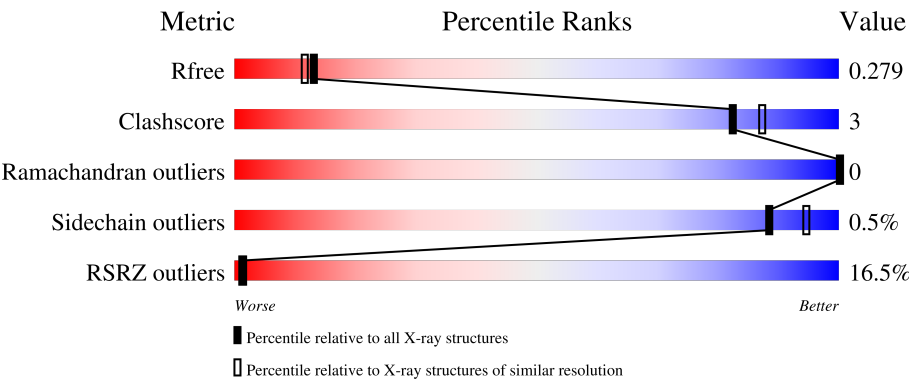
MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.23.2  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.23.2

# 1 Overall quality at a glance i

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

The reported resolution of this entry is 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	D	38	<div><div>18%</div><div>87%</div><div>11%</div><div>.</div></div>
1	E	38	<div><div>11%</div><div>84%</div><div>5%</div><div>11%</div></div>
1	F	38	<div><div>13%</div><div>84%</div><div>5%</div><div>11%</div></div>
2	A	57	<div><div>12%</div><div>86%</div><div>.</div><div>11%</div></div>
2	B	57	<div><div>18%</div><div>84%</div><div>9%</div><div>7%</div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	C	57	<div><div></div><div>18%</div><div>84%</div><div>7%</div><div>9%</div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3983 atoms, of which 1999 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 S2'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	34	Total	C	H	N	O	0	0	0
			537	170	272	41	54			
1	D	37	Total	C	H	N	O	0	0	1
			589	184	299	45	61			
1	F	34	Total	C	H	N	O	0	0	0
			525	167	263	41	54			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	1167	ACE	-	acetylation	UNP P0DTC2
E	1171	GLN	GLY	engineered mutation	UNP P0DTC2
E	1180	GLU	GLN	engineered mutation	UNP P0DTC2
E	1181	TYR	LYS	engineered mutation	UNP P0DTC2
E	1184	LYS	ASP	engineered mutation	UNP P0DTC2
E	1185	LYS	ARG	engineered mutation	UNP P0DTC2
E	1187	GLU	ASN	engineered mutation	UNP P0DTC2
E	1192	LYS	ASN	engineered mutation	UNP P0DTC2
E	1194	GLU	ASN	engineered mutation	UNP P0DTC2
E	1204	NH2	-	amidation	UNP P0DTC2
D	1167	ACE	-	acetylation	UNP P0DTC2
D	1171	GLN	GLY	engineered mutation	UNP P0DTC2
D	1180	GLU	GLN	engineered mutation	UNP P0DTC2
D	1181	TYR	LYS	engineered mutation	UNP P0DTC2
D	1184	LYS	ASP	engineered mutation	UNP P0DTC2
D	1185	LYS	ARG	engineered mutation	UNP P0DTC2
D	1187	GLU	ASN	engineered mutation	UNP P0DTC2
D	1192	LYS	ASN	engineered mutation	UNP P0DTC2
D	1194	GLU	ASN	engineered mutation	UNP P0DTC2
D	1204	NH2	-	amidation	UNP P0DTC2
F	1167	ACE	-	acetylation	UNP P0DTC2
F	1171	GLN	GLY	engineered mutation	UNP P0DTC2
F	1180	GLU	GLN	engineered mutation	UNP P0DTC2

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	1181	TYR	LYS	engineered mutation	UNP P0DTC2
F	1184	LYS	ASP	engineered mutation	UNP P0DTC2
F	1185	LYS	ARG	engineered mutation	UNP P0DTC2
F	1187	GLU	ASN	engineered mutation	UNP P0DTC2
F	1192	LYS	ASN	engineered mutation	UNP P0DTC2
F	1194	GLU	ASN	engineered mutation	UNP P0DTC2
F	1204	NH2	-	amidation	UNP P0DTC2

- Molecule 2 is a protein called Spike protein S2'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	52	Total	C	H	N	O	0	0	0
			769	238	385	68	78			
2	B	53	Total	C	H	N	O	0	0	0
			795	245	400	70	80			
2	A	51	Total	C	H	N	O	0	0	0
			760	234	380	68	78			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	911	ACE	-	acetylation	UNP P0DTC2
C	967	NH2	-	amidation	UNP P0DTC2
B	911	ACE	-	acetylation	UNP P0DTC2
B	967	NH2	-	amidation	UNP P0DTC2
A	911	ACE	-	acetylation	UNP P0DTC2
A	967	NH2	-	amidation	UNP P0DTC2

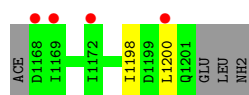
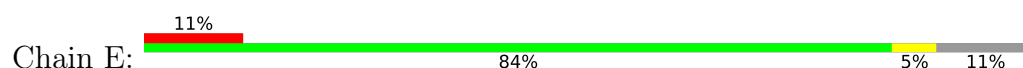
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	D	2	Total	O	0	0
			2	2		
3	C	2	Total	O	0	0
			2	2		
3	B	4	Total	O	0	0
			4	4		

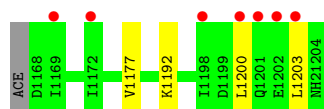
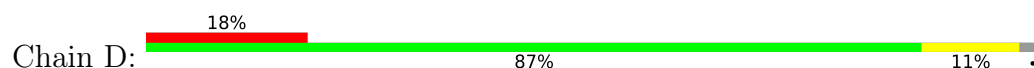
### 3 Residue-property plots [i](#)

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.

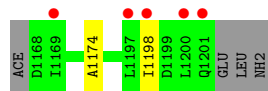
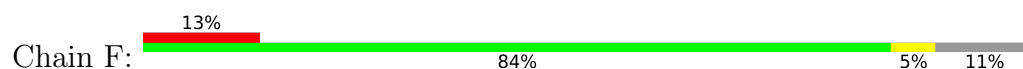
- Molecule 1: Spike protein S2'



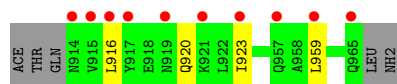
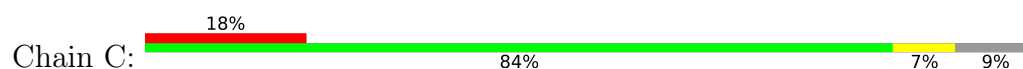
- Molecule 1: Spike protein S2'



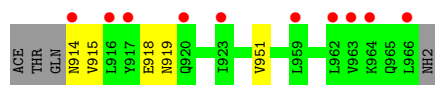
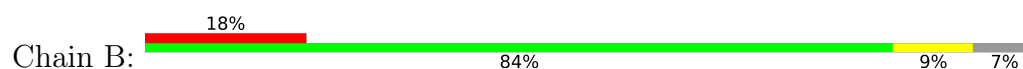
- Molecule 1: Spike protein S2'



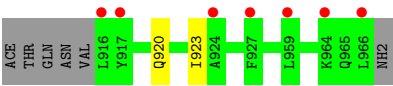
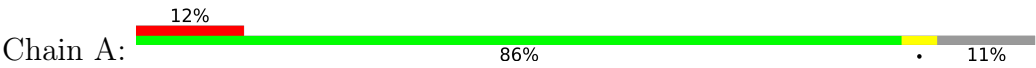
- Molecule 2: Spike protein S2'



- Molecule 2: Spike protein S2'



● Molecule 2: Spike protein S2'



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	53.54Å 54.46Å 80.12Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.04 – 2.20 80.12 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.1 (45.04-2.20) 91.3 (80.12-2.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.09 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, $R_{free}$	0.243 , 0.278 0.243 , 0.279	Depositor DCC
$R_{free}$ test set	1217 reflections (9.87%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	51.7	Xtriage
Anisotropy	0.269	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 76.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.017 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3983	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.42% 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 [i](#)

### 5.1 Standard geometry [i](#)

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

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	D	0.29	0/289	0.41	0/388
1	E	0.24	0/265	0.40	0/356
1	F	0.23	0/262	0.41	0/352
2	A	0.23	0/381	0.34	0/514
2	B	0.23	0/396	0.33	0/535
2	C	0.23	0/385	0.34	0/520
All	All	0.24	0/1978	0.37	0/2665

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	D	290	299	297	3	0
1	E	265	272	272	2	0
1	F	262	263	263	2	0
2	A	380	380	380	2	0
2	B	395	400	400	3	0
2	C	384	385	385	3	0
3	B	4	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	2	0	0	0	0
3	D	2	0	0	1	0
All	All	1984	1999	1997	10	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 3.

All (10) 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:D:1192:LYS:NZ	3:D:1301:HOH:O	2.27	0.68
2:C:916:LEU:O	2:C:920:GLN:N	2.30	0.63
2:C:959:LEU:HD23	1:F:1174:ALA:HB2	1.83	0.60
1:F:1198:ILE:HD13	2:A:923:ILE:HG12	1.85	0.59
1:E:1200:LEU:O	1:E:1200:LEU:HD12	2.03	0.58
1:E:1198:ILE:HD13	2:C:923:ILE:HG22	1.90	0.54
2:B:914:ASN:O	2:B:918:GLU:N	2.32	0.43
2:B:915:VAL:O	2:B:919:ASN:HB2	2.20	0.42
1:D:1200:LEU:HD11	2:A:920:GLN:O	2.18	0.42
1:D:1177:VAL:HB	2:B:951:VAL:HG21	2.01	0.41

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	D	35/38 (92%)	35 (100%)	0	0	100	100
1	E	32/38 (84%)	31 (97%)	1 (3%)	0	100	100
1	F	32/38 (84%)	32 (100%)	0	0	100	100
2	A	49/57 (86%)	49 (100%)	0	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	51/57 (90%)	50 (98%)	1 (2%)	0	100	100
2	C	50/57 (88%)	50 (100%)	0	0	100	100
All	All	249/285 (87%)	247 (99%)	2 (1%)	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	D	33/34 (97%)	32 (97%)	1 (3%)	41	53
1	E	29/34 (85%)	29 (100%)	0	100	100
1	F	28/34 (82%)	28 (100%)	0	100	100
2	A	41/47 (87%)	41 (100%)	0	100	100
2	B	43/47 (92%)	43 (100%)	0	100	100
2	C	41/47 (87%)	41 (100%)	0	100	100
All	All	215/243 (88%)	214 (100%)	1 (0%)	88	94

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

Mol	Chain	Res	Type
1	D	1203	LEU

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

### 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 ⓘ

### 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	D	36/38 (94%)	1.05	7 (19%) 1 1	44, 69, 153, 202	0
1	E	34/38 (89%)	0.80	4 (11%) 4 4	50, 78, 156, 165	0
1	F	34/38 (89%)	1.00	5 (14%) 2 2	55, 81, 150, 203	0
2	A	51/57 (89%)	0.95	7 (13%) 3 2	42, 70, 150, 188	0
2	B	53/57 (92%)	0.88	10 (18%) 1 1	34, 66, 143, 176	0
2	C	52/57 (91%)	0.99	10 (19%) 1 1	46, 78, 162, 208	0
All	All	260/285 (91%)	0.94	43 (16%) 1 1	34, 78, 160, 208	0

All (43) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	916	LEU	9.7
2	C	915	VAL	8.3
2	B	966	LEU	6.7
1	D	1200	LEU	6.2
1	F	1198	ILE	6.0
2	A	917	TYR	5.1
1	E	1172	ILE	4.7
2	C	914	ASN	4.5
2	B	962	LEU	4.4
1	E	1169	ILE	4.3
1	F	1201	GLN	4.2
1	D	1203	LEU	4.2
2	C	916	LEU	3.8
1	D	1198	ILE	3.8
1	D	1169	ILE	3.7
1	E	1168	ASP	3.5
2	B	964	LYS	3.5
1	E	1200	LEU	3.3
1	D	1202	GLU	3.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	D	1172	ILE	3.2
2	C	923	ILE	3.2
2	C	921	LYS	3.1
2	B	914	ASN	3.1
2	B	923	ILE	3.0
2	B	963	VAL	3.0
2	C	959	LEU	3.0
2	C	957	GLN	2.9
2	B	917	TYR	2.9
2	C	919	ASN	2.9
2	B	916	LEU	2.8
2	A	924	ALA	2.8
1	F	1169	ILE	2.8
2	B	959	LEU	2.8
2	A	927	PHE	2.7
1	F	1197	LEU	2.5
2	B	920	GLN	2.5
1	F	1200	LEU	2.4
2	A	964	LYS	2.4
2	C	965	GLN	2.3
2	A	966	LEU	2.2
1	D	1201	GLN	2.1
2	A	959	LEU	2.1
2	C	917	TYR	2.1

## 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.