



# Full wwPDB X-ray Structure Validation Report ⓘ

May 4, 2021 – 03:03 pm BST

PDB ID : 7NKT  
Title : RBD domain of SARS-CoV2 in complex with neutralizing nanobody NM1226  
Authors : Ostertag, E.; Zocher, G.; Stehle, T.  
Deposited on : 2021-02-18  
Resolution : 2.30 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.18  
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) : 2.18

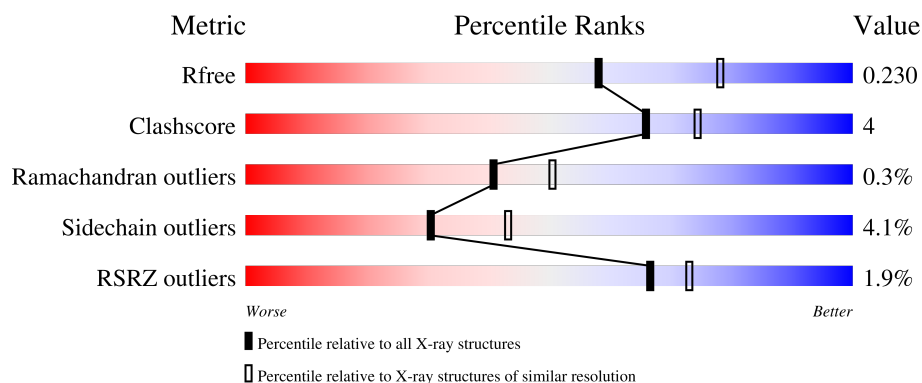
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*



The reported resolution of this entry is 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	AAA	229	 3% 77% 8% • 14%
2	BBB	144	 76% 10% • 13%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	PEG	BBB	203	-	-	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 2696 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
1	AAA	196	Total	C	N	O	S	0	2	0
			1566	1004	263	291	8			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	542	HIS	-	expression tag	UNP P0DTC2
AAA	543	HIS	-	expression tag	UNP P0DTC2
AAA	544	HIS	-	expression tag	UNP P0DTC2
AAA	545	HIS	-	expression tag	UNP P0DTC2
AAA	546	HIS	-	expression tag	UNP P0DTC2
AAA	547	HIS	-	expression tag	UNP P0DTC2

- Molecule 2 is a protein called neutralizing nanobody NM1226.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	BBB	125	Total	C	N	O	S	0	0	0
			953	599	159	190	5			

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



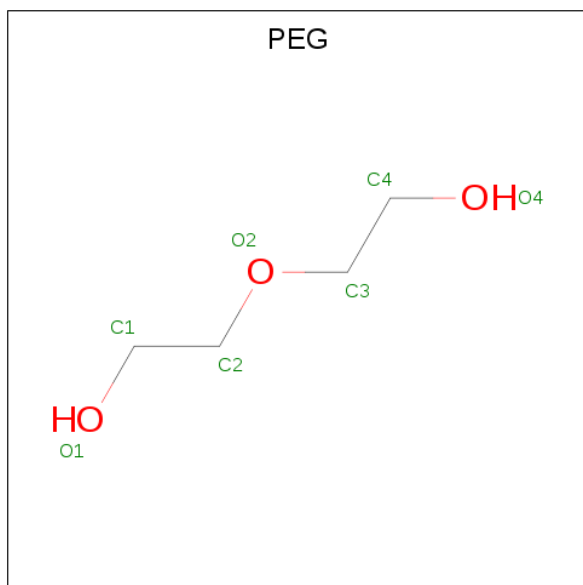
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	AAA	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 4 is PHOSPHATE ION (three-letter code: PO4) (formula:  $O_4P$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	AAA	1	Total	O	P	0	0
			5	4	1		

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula:  $C_4H_{10}O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	BBB	1	Total	C	O	0	0
			7	4	3		
5	BBB	1	Total	C	O	0	0
			7	4	3		
5	BBB	1	Total	C	O	0	0
			7	4	3		

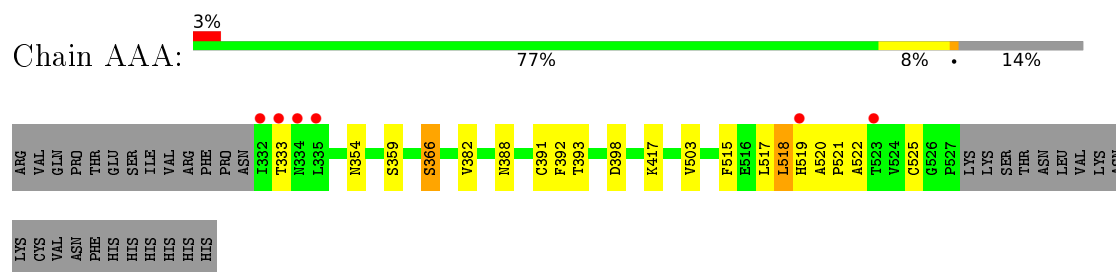
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	AAA	107	Total	O	0	0
			107	107		
6	BBB	30	Total	O	0	0
			30	30		

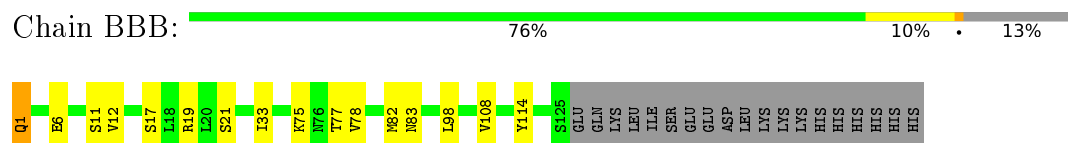
### 3 Residue-property plots [i](#)

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 S1



- Molecule 2: neutralizing nanobody NM1226



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 41	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	128.09 Å   128.09 Å   77.68 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	46.10 – 2.30 46.10 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.9 (46.10-2.30) 99.9 (46.10-2.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.99 (at 2.29 Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, $R_{free}$	0.184 , 0.224 0.189 , 0.230	Depositor DCC
$R_{free}$ test set	1687 reflections (6.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	54.7	Xtriage
Anisotropy	0.176	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 38.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.025 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2696	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.63% 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: PEG, PO4, NAG

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	AAA	0.67	0/1613	0.82	0/2195
2	BBB	0.67	0/975	0.79	0/1320
All	All	0.67	0/2588	0.81	0/3515

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	AAA	1566	0	1488	11	0
2	BBB	953	0	900	11	0
3	AAA	14	0	13	0	0
4	AAA	5	0	0	0	0
5	BBB	21	0	30	0	0
6	AAA	107	0	0	1	0
6	BBB	30	0	0	0	0
All	All	2696	0	2431	20	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 4.

All (20) 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:BBB:33:ILE:HG13	2:BBB:78:VAL:HG21	1.71	0.72
1:AAA:393:THR:HG21	1:AAA:518:LEU:HD11	1.78	0.66
1:AAA:393:THR:HG21	1:AAA:518:LEU:CD1	2.28	0.64
2:BBB:1:GLN:N	2:BBB:114:TYR:OH	2.34	0.60
1:AAA:503:VAL:HG11	2:BBB:108:VAL:HG11	1.82	0.60
2:BBB:6:GLU:HA	2:BBB:21:SER:O	2.02	0.60
2:BBB:1:GLN:N	2:BBB:114:TYR:CZ	2.73	0.56
1:AAA:417:LYS:NZ	6:AAA:701:HOH:O	2.43	0.51
2:BBB:98:LEU:C	2:BBB:98:LEU:HD12	2.31	0.50
1:AAA:503:VAL:HG11	2:BBB:108:VAL:CG1	2.42	0.49
2:BBB:17:SER:HB2	2:BBB:83:ASN:OD1	2.14	0.47
1:AAA:354:ASN:O	1:AAA:398:ASP:HA	2.15	0.47
1:AAA:392:PHE:CD1	1:AAA:515:PHE:HB3	2.50	0.46
1:AAA:393:THR:HA	1:AAA:522:ALA:HA	1.97	0.46
2:BBB:17:SER:HA	2:BBB:82:MET:O	2.15	0.46
2:BBB:75:LYS:O	2:BBB:77:THR:HG23	2.18	0.43
2:BBB:33:ILE:CG1	2:BBB:78:VAL:HG21	2.45	0.41
1:AAA:517:LEU:HD23	1:AAA:517:LEU:HA	1.95	0.41
1:AAA:520:ALA:HB1	1:AAA:521:PRO:HD2	2.02	0.41
1:AAA:366:SER:HB3	1:AAA:388:ASN:HD21	1.86	0.41

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	AAA	196/229 (86%)	190 (97%)	5 (3%)	1 (0%)	29	35
2	BBB	123/144 (85%)	118 (96%)	5 (4%)	0	100	100
All	All	319/373 (86%)	308 (97%)	10 (3%)	1 (0%)	41	50

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	519	HIS

### 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	AAA	171/202 (85%)	164 (96%)	7 (4%)	30	43
2	BBB	99/118 (84%)	95 (96%)	4 (4%)	31	44
All	All	270/320 (84%)	259 (96%)	11 (4%)	30	43

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	333	THR
1	AAA	359	SER
1	AAA	366	SER
1	AAA	382	VAL
1	AAA	391	CYS
1	AAA	518	LEU
1	AAA	525	CYS
2	BBB	1	GLN
2	BBB	11	SER
2	BBB	12	VAL
2	BBB	19	ARG

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 [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](#)

5 ligands are 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	$\# Z  > 2$
5	PEG	BBB	202	-	6,6,6	0.17	0	5,5,5	0.10	0
5	PEG	BBB	201	-	6,6,6	0.23	0	5,5,5	0.15	0
5	PEG	BBB	203	-	6,6,6	0.21	0	5,5,5	0.13	0
3	NAG	AAA	601	1	14,14,15	0.36	0	17,19,21	1.01	1 (5%)
4	PO4	AAA	602	-	4,4,4	0.68	0	6,6,6	0.44	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
5	PEG	BBB	202	-	-	0/4/4/4	-
5	PEG	BBB	201	-	-	2/4/4/4	-
5	PEG	BBB	203	-	-	2/4/4/4	-
3	NAG	AAA	601	1	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
3	AAA	601	NAG	C1-O5-C5	2.04	114.96	112.19

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	BBB	203	PEG	O2-C3-C4-O4
5	BBB	201	PEG	O1-C1-C2-O2
5	BBB	201	PEG	C1-C2-O2-C3
5	BBB	203	PEG	C4-C3-O2-C2

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 [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	AAA	196/229 (85%)	-0.13	6 (3%) 49 56	38, 54, 103, 143	0
2	BBB	125/144 (86%)	-0.20	0 100 100	45, 69, 97, 126	0
All	All	321/373 (86%)	-0.16	6 (1%) 66 73	38, 59, 103, 143	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	332	ILE	5.4
1	AAA	333	THR	4.8
1	AAA	519	HIS	2.4
1	AAA	335	LEU	2.3
1	AAA	334	ASN	2.3
1	AAA	523	THR	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
5	PEG	BBB	201	7/7	0.64	0.33	84,97,104,109	0
5	PEG	BBB	203	7/7	0.65	0.60	81,98,117,121	0
5	PEG	BBB	202	7/7	0.83	0.16	89,99,103,107	0
4	PO4	AAA	602	5/5	0.93	0.21	69,77,93,98	0
3	NAG	AAA	601	14/15	0.95	0.11	70,80,86,87	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.