



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 30, 2021 – 11:07 am BST

PDB ID : 7OAQ  
Title : Nanobody H3 AND C1 bound to RBD with Kent mutation  
Authors : Naismith, J.H.; Mikolajek, H.  
Deposited on : 2021-04-20  
Resolution : 1.55 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 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	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

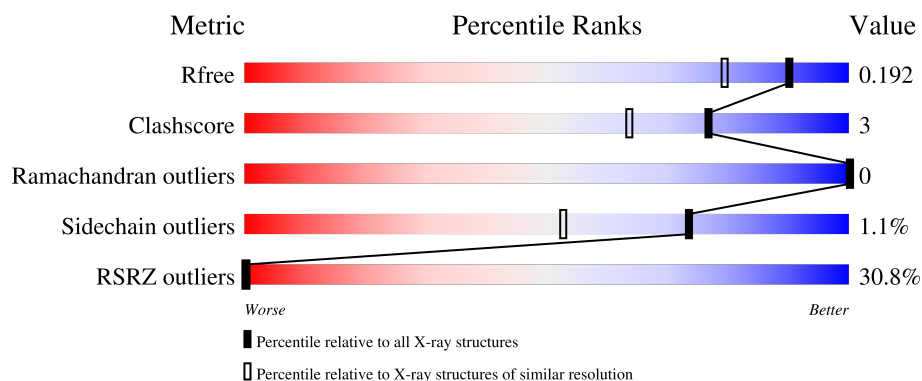
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.55 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	1483 (1.56-1.56)
Clashscore	141614	1529 (1.56-1.56)
Ramachandran outliers	138981	1498 (1.56-1.56)
Sidechain outliers	138945	1495 (1.56-1.56)
RSRZ outliers	127900	1465 (1.56-1.56)

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	131	<div> <div>66%</div> <div> <div></div> <div>91%</div> <div>5%</div> </div> </div>
2	EEE	210	<div> <div>13%</div> <div> <div></div> <div>89%</div> <div>5%</div> <div>6%</div> </div> </div>
3	FFF	136	<div> <div>18%</div> <div> <div></div> <div>89%</div> <div>7%</div> </div> </div>

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	CIT	EEE	602	-	-	X	-

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 3991 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 H3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	124	Total	C	N	O	S	0	3	0
			976	614	167	191	4			

- Molecule 2 is a protein called Spike protein S1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	EEE	197	Total	C	N	O	S	0	7	0
			1598	1035	260	295	8			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
EEE	501	TYR	ASN	engineered mutation	UNP P0DTC2
EEE	533	LYS	-	expression tag	UNP P0DTC2
EEE	534	HIS	-	expression tag	UNP P0DTC2
EEE	535	HIS	-	expression tag	UNP P0DTC2
EEE	536	HIS	-	expression tag	UNP P0DTC2
EEE	537	HIS	-	expression tag	UNP P0DTC2
EEE	538	HIS	-	expression tag	UNP P0DTC2
EEE	539	HIS	-	expression tag	UNP P0DTC2

- Molecule 3 is a protein called H3.

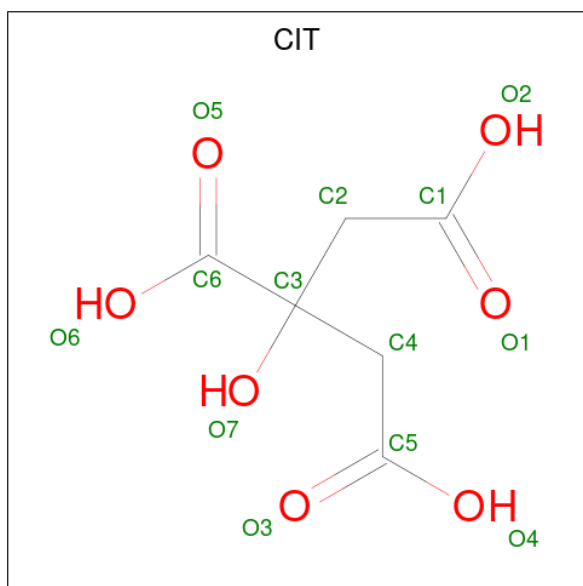
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	FFF	127	Total	C	N	O	S	0	8	0
			1014	638	169	202	5			

- Molecule 4 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
4	EEE	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is CITRIC ACID (three-letter code: CIT) (formula:  $C_6H_8O_7$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	EEE	1	Total	C	O	0	0
			13	6	7		

- Molecule 6 is CHLORIDE ION (three-letter code: CL) (formula:  $Cl$ ).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	FFF	1	Total	Cl	0	0
			1	1		

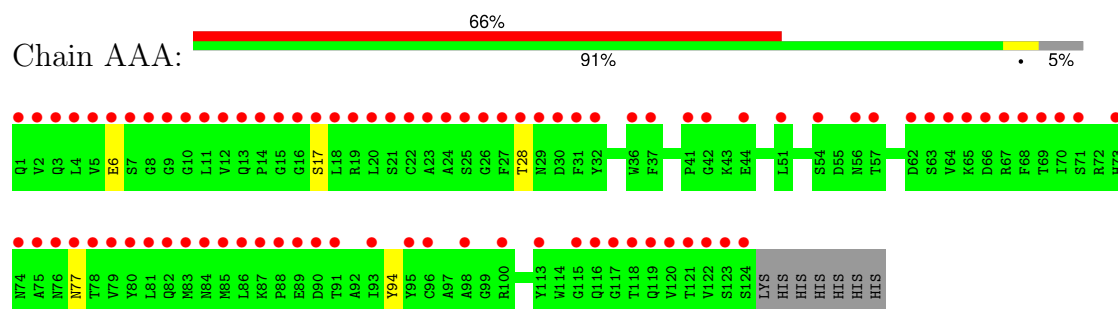
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	AAA	30	Total	O	0	0
			30	30		
7	EEE	206	Total	O	0	0
			206	206		
7	FFF	139	Total	O	0	0
			139	139		

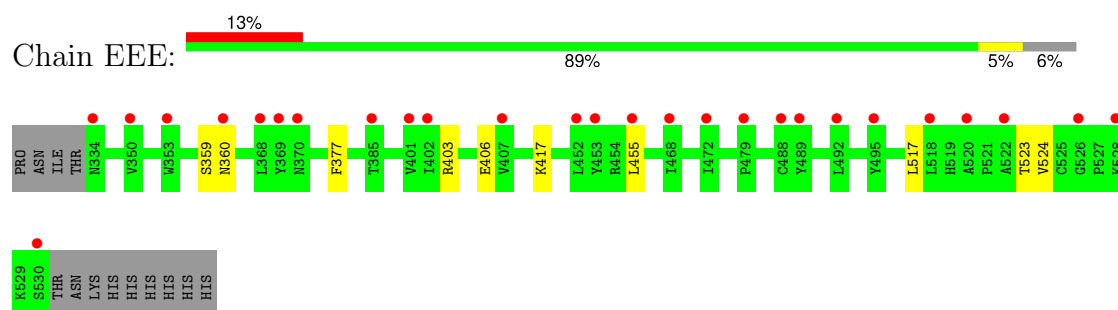
### 3 Residue-property plots [i](#)

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.

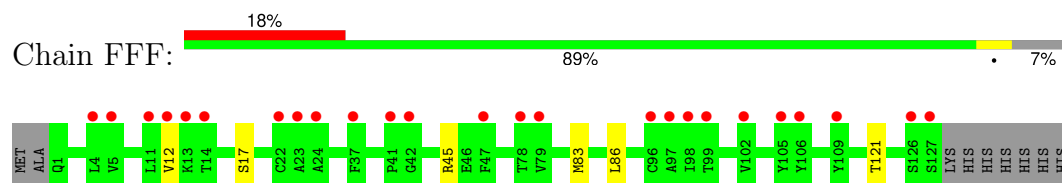
- Molecule 1: H3



- Molecule 2: Spike protein S1



- Molecule 3: H3



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	105.86Å 105.86Å 112.71Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	52.98 – 1.55 52.93 – 1.55	Depositor EDS
% Data completeness (in resolution range)	99.3 (52.98-1.55) 99.3 (52.93-1.55)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.08 (at 1.55Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.155 , 0.178 0.168 , 0.192	Depositor DCC
$R_{free}$ test set	4722 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.6	Xtriage
Anisotropy	0.818	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3991	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.41% 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: CL, CIT, 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	0.72	0/1010	0.77	0/1374
2	EEE	0.66	0/1665	0.76	0/2265
3	FFF	0.71	0/1060	0.78	0/1434
All	All	0.69	0/3735	0.77	0/5073

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	976	0	927	1	0
2	EEE	1598	0	1545	11	0
3	FFF	1014	0	986	5	0
4	EEE	14	0	13	0	0
5	EEE	13	0	5	8	0
6	FFF	1	0	0	0	0
7	AAA	30	0	0	0	0
7	EEE	206	0	0	5	0
7	FFF	139	0	0	3	0
All	All	3991	0	3476	21	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.

The worst 5 of 21 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)
5:EEE:602:CIT:O6	7:EEE:701:HOH:O	1.97	0.82
5:EEE:602:CIT:O4	7:EEE:702:HOH:O	2.01	0.78
5:EEE:602:CIT:O3	7:EEE:703:HOH:O	2.01	0.77
2:EEE:359[B]:SER:HA	2:EEE:524:VAL:CG2	2.23	0.68
2:EEE:406:GLU:OE2	5:EEE:602:CIT:O5	2.19	0.61

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	AAA	125/131 (95%)	120 (96%)	5 (4%)	0	100	100
2	EEE	202/210 (96%)	198 (98%)	4 (2%)	0	100	100
3	FFF	133/136 (98%)	132 (99%)	1 (1%)	0	100	100
All	All	460/477 (96%)	450 (98%)	10 (2%)	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	AAA	106/110 (96%)	102 (96%)	4 (4%)	33	6
2	EEE	177/183 (97%)	176 (99%)	1 (1%)	86	73
3	FFF	112/112 (100%)	112 (100%)	0	100	100
All	All	395/405 (98%)	390 (99%)	5 (1%)	73	44

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	17[A]	SER
1	AAA	17[B]	SER
1	AAA	28	THR
1	AAA	77	ASN
2	EEE	377	PHE

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](#)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 for Mogul analysis.

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	CIT	EEE	602	-	3,12,12	0.46	0	3,17,17	0.83	0
4	NAG	EEE	601	2	14,14,15	0.42	0	17,19,21	1.55	2 (11%)

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	CIT	EEE	602	-	-	4/6/16/16	-
4	NAG	EEE	601	2	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	EEE	601	NAG	C1-O5-C5	4.93	118.87	112.19
4	EEE	601	NAG	O5-C1-C2	-2.49	107.35	111.29

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	EEE	602	CIT	C2-C3-C4-C5
5	EEE	602	CIT	O7-C3-C4-C5
5	EEE	602	CIT	C6-C3-C4-C5
5	EEE	602	CIT	C1-C2-C3-O7

There are no ring outliers.

1 monomer is involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	EEE	602	CIT	8	0

## 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 [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	124/131 (94%)	3.81	86 (69%) 0 0	30, 63, 88, 108	0
2	EEE	197/210 (93%)	1.04	27 (13%) 3 2	19, 29, 48, 81	0
3	FFF	127/136 (93%)	1.32	25 (19%) 1 1	18, 24, 50, 72	0
All	All	448/477 (93%)	1.88	138 (30%) 0 0	18, 32, 79, 108	0

The worst 5 of 138 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	18	LEU	12.0
1	AAA	122	VAL	11.1
1	AAA	75	ALA	9.7
1	AAA	10	GLY	9.4
1	AAA	28	THR	9.3

### 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
4	NAG	EEE	601	14/15	0.78	0.31	40,48,56,61	0
5	CIT	EEE	602	13/13	0.86	0.33	32,40,64,87	0
6	CL	FFF	201	1/1	1.00	0.22	23,23,23,23	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.