



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 02:01 pm BST

PDB ID : 2GE8  
Title : Structure of the C-terminal dimerization domain of infectious bronchitis virus nucleocapsid protein  
Authors : Jayaram, H.; Fan, H.; Bowman, B.R.; Ooi, A.; Jayaram, J.; Collisson, E.W.; Lescar, J.; Prasad, B.V.  
Deposited on : 2006-03-18  
Resolution : 2.20 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
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

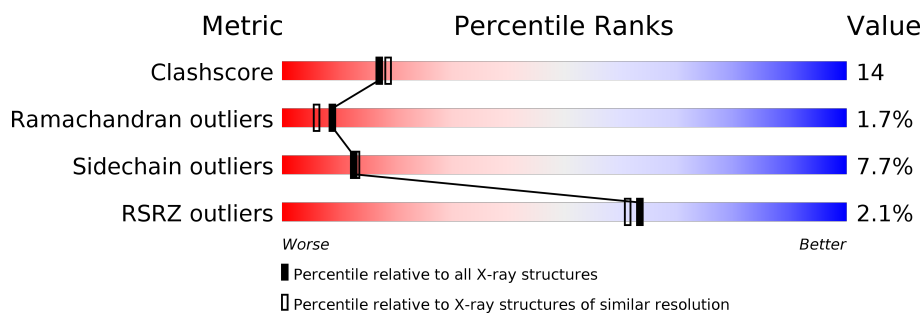
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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))
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 on 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	114	<div> <div>2%</div> <div>64%</div> <div>24%</div> <div>9%</div> <div>..</div> </div>
1	B	114	<div> <div>2%</div> <div>69%</div> <div>19%</div> <div>6%</div> <div>..</div> </div>
1	C	114	<div> <div>2%</div> <div>73%</div> <div>19%</div> <div>..</div> </div>
1	D	114	<div> <div>78%</div> <div>17%</div> <div>..</div> </div>
1	F	114	<div> <div>2%</div> <div>58%</div> <div>30%</div> <div>5%</div> <div>7%</div> </div>
1	G	114	<div> <div>%</div> <div>70%</div> <div>25%</div> <div>..</div> </div>
1	I	114	<div> <div>%</div> <div>67%</div> <div>25%</div> <div>6%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	J	114	<div><div></div><div></div><div></div><div></div><div></div></div> <div>7%57%32%5%5%</div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 6881 atoms, of which 0 are hydrogens and 0 are deuteriums.

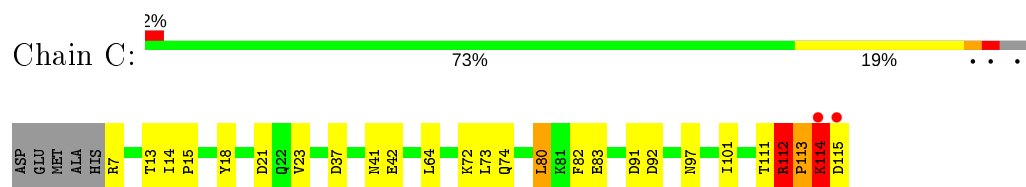
In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Nucleocapsid protein.

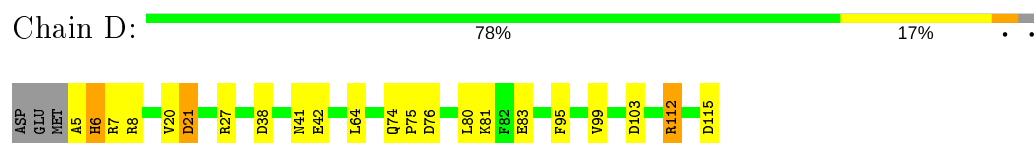
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	112	Total	C	N	O	S	0	0	0
			882	551	160	164	7			
1	B	109	Total	C	N	O	S	0	0	0
			859	537	155	161	6			
1	F	106	Total	C	N	O	S	0	0	0
			829	520	148	155	6			
1	G	113	Total	C	N	O	S	0	0	0
			890	556	161	166	7			
1	C	109	Total	C	N	O	S	0	0	0
			859	537	155	161	6			
1	D	111	Total	C	N	O	S	0	0	0
			874	546	159	163	6			
1	I	107	Total	C	N	O	S	0	0	0
			841	527	152	156	6			
1	J	108	Total	C	N	O	S	0	0	0
			847	531	151	159	6			



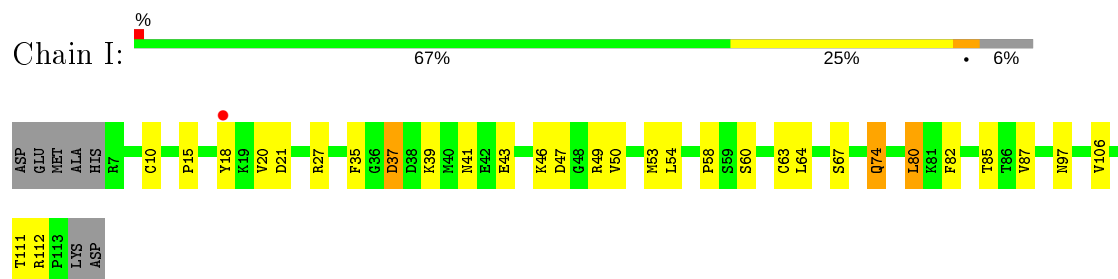
- Molecule 1: Nucleocapsid protein



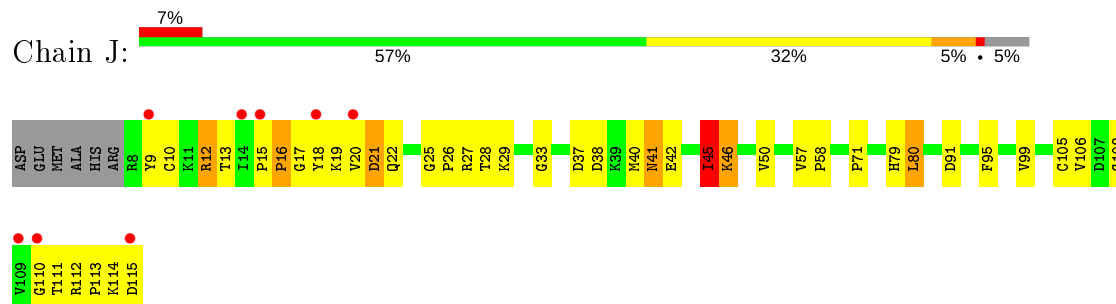
- Molecule 1: Nucleocapsid protein



- Molecule 1: Nucleocapsid protein



- Molecule 1: Nucleocapsid protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	108.99 Å   128.53 Å   71.44 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	47.78 – 2.20 47.78 – 2.20	Depositor EDS
% Data completeness (in resolution range)	100.0 (47.78-2.20) 98.9 (47.78-2.20)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	0.12	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.14 (at 2.20 Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.237 ,   0.291 0.252 ,   (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.2	Xtriage
Anisotropy	0.495	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 30.4	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	0.92	EDS
Total number of atoms	6881	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.23% 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

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	0.95	0/902	1.09	5/1214 (0.4%)
1	B	1.03	0/878	1.00	3/1182 (0.3%)
1	C	0.92	0/878	0.91	2/1182 (0.2%)
1	D	0.94	0/894	1.06	5/1204 (0.4%)
1	F	0.92	0/847	0.87	0/1142
1	G	0.98	2/910 (0.2%)	0.89	0/1226
1	I	0.88	0/860	0.81	0/1160
1	J	0.93	1/866 (0.1%)	0.95	0/1168
All	All	0.95	3/7035 (0.0%)	0.95	15/9478 (0.2%)

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
1	A	1	2
1	B	0	2
1	D	0	1
All	All	1	5

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	32	GLU	CD-OE2	7.38	1.33	1.25
1	J	105	CYS	CB-SG	-5.98	1.72	1.81
1	G	63	CYS	CB-SG	-5.56	1.72	1.81

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	6	HIS	N-CA-C	-8.24	88.74	111.00
1	A	21	ASP	C-N-CA	6.05	136.83	121.70

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	7	ARG	N-CA-C	5.96	127.10	111.00
1	B	115	ASP	CB-CG-OD2	-5.90	112.99	118.30
1	A	21	ASP	N-CA-C	5.88	126.88	111.00

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	103	ASP	CA

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	21	ASP	Peptide
1	A	4	MET	Peptide
1	B	114	LYS	Peptide
1	B	67	SER	Peptide
1	D	5	ALA	Peptide

## 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	882	0	870	33	0
1	B	859	0	849	24	0
1	C	859	0	849	16	0
1	D	874	0	861	18	0
1	F	829	0	814	34	0
1	G	890	0	876	26	0
1	I	841	0	832	31	0
1	J	847	0	836	40	0
All	All	6881	0	6787	190	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 14.

The worst 5 of 190 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:F:92:ASP:O	1:F:94:GLN:O	1.71	1.07
1:J:16:PRO:HA	1:J:45:ILE:HD11	1.45	0.97
1:D:81:LYS:HE2	1:D:83:GLU:HG3	1.44	0.95
1:A:27:ARG:HB2	1:A:109:VAL:HG13	1.50	0.92
1:J:16:PRO:O	1:J:18:TYR:N	2.06	0.88

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	110/114 (96%)	100 (91%)	7 (6%)	3 (3%)	5	2
1	B	107/114 (94%)	102 (95%)	3 (3%)	2 (2%)	8	5
1	C	107/114 (94%)	104 (97%)	1 (1%)	2 (2%)	8	5
1	D	109/114 (96%)	106 (97%)	3 (3%)	0	100	100
1	F	104/114 (91%)	96 (92%)	7 (7%)	1 (1%)	15	14
1	G	111/114 (97%)	106 (96%)	4 (4%)	1 (1%)	17	16
1	I	105/114 (92%)	98 (93%)	6 (6%)	1 (1%)	15	14
1	J	106/114 (93%)	94 (89%)	7 (7%)	5 (5%)	2	1
All	All	859/912 (94%)	806 (94%)	38 (4%)	15 (2%)	9	6

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	95	PHE
1	C	113	PRO
1	C	114	LYS
1	J	16	PRO
1	J	17	GLY

### 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	A	98/100 (98%)	87 (89%)	11 (11%)	6	5
1	B	96/100 (96%)	90 (94%)	6 (6%)	18	20
1	C	96/100 (96%)	89 (93%)	7 (7%)	14	15
1	D	97/100 (97%)	92 (95%)	5 (5%)	23	28
1	F	92/100 (92%)	81 (88%)	11 (12%)	5	4
1	G	99/100 (99%)	92 (93%)	7 (7%)	14	16
1	I	94/100 (94%)	89 (95%)	5 (5%)	22	27
1	J	95/100 (95%)	88 (93%)	7 (7%)	13	14
All	All	767/800 (96%)	708 (92%)	59 (8%)	13	13

5 of 59 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	100	LYS
1	G	91	ASP
1	J	41	ASN
1	F	111	THR
1	G	64	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	74	GLN
1	D	41	ASN
1	I	79	HIS
1	C	41	ASN
1	I	97	ASN

### 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 carbohydrates 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	A	112/114 (98%)	0.01	2 (1%) 68 66	15, 28, 45, 53	0
1	B	109/114 (95%)	-0.23	2 (1%) 68 66	15, 23, 32, 46	0
1	C	109/114 (95%)	-0.11	2 (1%) 68 66	17, 27, 45, 60	0
1	D	111/114 (97%)	-0.21	0 100 100	16, 25, 38, 40	0
1	F	106/114 (92%)	0.17	2 (1%) 66 65	16, 37, 49, 66	0
1	G	113/114 (99%)	-0.14	1 (0%) 84 83	17, 26, 41, 46	0
1	I	107/114 (93%)	0.46	1 (0%) 84 83	21, 41, 58, 63	0
1	J	108/114 (94%)	0.61	8 (7%) 14 13	24, 40, 67, 73	0
All	All	875/912 (95%)	0.07	18 (2%) 63 61	15, 30, 56, 73	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	J	115	ASP	5.3
1	J	18	TYR	4.7
1	J	20	VAL	3.8
1	B	115	ASP	3.6
1	A	115	ASP	3.5

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 6.4 Ligands

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.