Full wwPDB X-ray Structure Validation Report (i)

Mar 24, 2025 – 12:19 PM JST

PDB ID	:	9IN1
Title	:	Crystal Structure of C-terminal domain of nucleocapsid protein from SARS-CoV-2
Authors	:	Dhaka, P.; Mahto, J.K.; Tomar, S.; Kumar, P.
Deposited on	:	2024-07-05
Resolution	:	1.40 Å(reported)
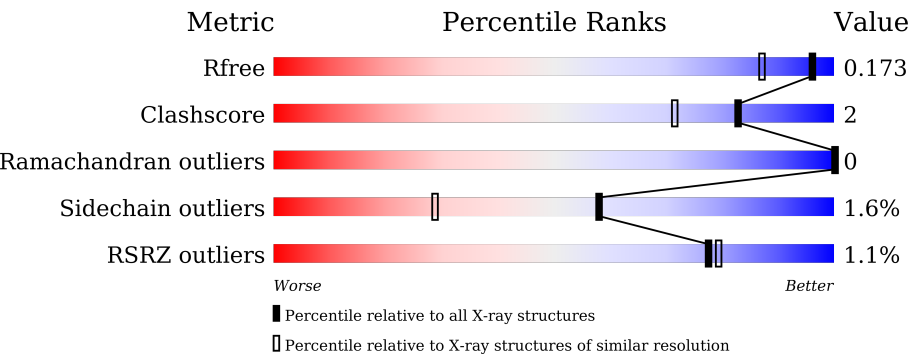
1   Overall quality at a glance ⓘ

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

The reported resolution of this entry is 1.40 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	Percentile Ranks	Value
Rfree		0.173
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		1.6%
RSRZ outliers		1.1%

Worse — *Better*

▊ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	164625	2247 (1.40-1.40)
Clashscore	180529	2446 (1.40-1.40)
Ramachandran outliers	177936	2398 (1.40-1.40)
Sidechain outliers	177891	2397 (1.40-1.40)
RSRZ outliers	164620	2246 (1.40-1.40)

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 >=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 <=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	120	83% · · · 6% · 10%
1	B	120	.% 84% 7% · 8%
1	C	120	2% 87% 9% · ·
1	D	120	2% 86% 7% 8%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 4504 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 Nucleoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	108	Total	C	N	O	S	0	3	0
			885	565	157	161	2			
1	B	110	Total	C	N	O	S	0	4	0
			904	575	160	166	3			
1	C	116	Total	C	N	O	S	0	7	0
			970	617	175	176	2			
1	D	111	Total	C	N	O	S	0	4	0
			914	579	161	171	3			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	245	MET	-	initiating methionine	UNP P0DTC9
A	246	GLY	-	expression tag	UNP P0DTC9
B	245	MET	-	initiating methionine	UNP P0DTC9
B	246	GLY	-	expression tag	UNP P0DTC9
C	245	MET	-	initiating methionine	UNP P0DTC9
C	246	GLY	-	expression tag	UNP P0DTC9
D	245	MET	-	initiating methionine	UNP P0DTC9
D	246	GLY	-	expression tag	UNP P0DTC9

- Molecule 2 is water.
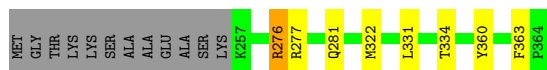
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	203	Total	O	0	0
			203	203		
2	B	223	Total	O	0	0
			223	223		
2	C	225	Total	O	0	0
			225	225		
2	D	180	Total	O	0	0
			180	180		

3 Residue-property plots ⓘ

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: Nucleoprotein

Chain A:

• Molecule 1: Nucleoprotein

Chain B:

• Molecule 1: Nucleoprotein

Chain C:

• Molecule 1: Nucleoprotein

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	71.05Å 43.49Å 74.78Å 90.00° 94.45° 90.00°	Depositor
Resolution (Å)	24.05 – 1.40 24.05 – 1.40	Depositor EDS
% Data completeness (in resolution range)	94.3 (24.05-1.40) 94.4 (24.05-1.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.24 (at 1.40Å)	Xtriage
Refinement program	REFMAC 5.8.0419	Depositor
R, R_{free}	0.132 , 0.170 0.142 , 0.173	Depositor DCC
R_{free} test set	4413 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å²)	8.1	Xtriage
Anisotropy	0.511	Xtriage
Bulk solvent k_{sol}(e/Å³), B_{sol}(Å²)	0.33 , 43.0	EDS
L-test for twinning [2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.026 for l,-k,h	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4504	wwPDB-VP
Average B, all atoms (Å²)	13.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.74% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.
[2] Theoretical values of $< |L| >$, $< L^2 >$ 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	0.58	0/914	1.05	2/1234 (0.2%)
1	B	0.65	0/936	1.04	5/1263 (0.4%)
1	C	0.59	0/1008	0.96	4/1356 (0.3%)
1	D	0.55	0/940	1.03	3/1269 (0.2%)
All	All	0.60	0/3798	1.02	14/5122 (0.3%)

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	0	1
1	B	0	1
1	C	0	2
All	All	0	4

There are no bond length outliers.

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	259	ARG	NE-CZ-NH1	11.69	126.14	120.30
1	C	259	ARG	NE-CZ-NH2	8.54	124.57	120.30
1	B	276	ARG	NE-CZ-NH1	-7.34	116.63	120.30
1	C	259	ARG	NE-CZ-NH1	-7.14	116.73	120.30
1	B	322[A]	MET	CG-SD-CE	-6.71	89.46	100.20
1	B	322[B]	MET	CG-SD-CE	-6.71	89.46	100.20
1	B	293	ARG	NE-CZ-NH1	6.62	123.61	120.30
1	D	259	ARG	NE-CZ-NH2	-6.41	117.10	120.30
1	D	276	ARG	NE-CZ-NH2	6.39	123.49	120.30
1	A	276	ARG	NE-CZ-NH2	-6.04	117.28	120.30
1	C	276[A]	ARG	NE-CZ-NH1	-5.54	117.53	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	276[B]	ARG	NE-CZ-NH1	-5.54	117.53	120.30
1	A	276	ARG	CB-CG-CD	5.32	125.44	111.60
1	B	276	ARG	NE-CZ-NH2	5.11	122.86	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	276	ARG	Sidechain
1	B	262	ARG	Sidechain
1	C	262	ARG	Sidechain
1	C	276[A]	ARG	Sidechain

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	885	0	877	5	0
1	B	904	0	895	5	0
1	C	970	0	978	7	0
1	D	914	0	892	6	0
2	A	203	0	0	3	0
2	B	223	0	0	2	0
2	C	225	0	0	0	0
2	D	180	0	0	0	0
All	All	4504	0	3642	17	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 2.

All (17) 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:C:334:THR:HG22	1:D:334[B]:THR:HG23	1.60	0.83
1:A:334:THR:HG22	1:B:334[B]:THR:HG23	1.65	0.78

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:281:GLN:HG2	2:A:666:HOH:O	1.96	0.65
2:A:544:HOH:O	1:B:338:LYS:HE3	1.97	0.65
1:C:334:THR:HG22	1:D:334[A]:THR:HG22	1.79	0.65
1:C:350:VAL:HA	1:D:322[A]:MET:HE1	1.84	0.60
1:A:281:GLN:CG	2:A:666:HOH:O	2.50	0.57
1:B:303[A]:GLN:HG2	2:B:513:HOH:O	2.11	0.51
1:C:322:MET:HE1	1:D:354:ASN:OD1	2.13	0.49
1:A:360:TYR:HA	1:A:363:PHE:CZ	2.48	0.48
1:C:360:TYR:HA	1:C:363:PHE:CZ	2.48	0.47
1:B:293:ARG:HG2	1:B:294[B]:GLN:HE21	1.80	0.46
1:C:322:MET:CE	1:D:354:ASN:OD1	2.66	0.44
1:B:334[B]:THR:HG21	2:B:639:HOH:O	2.20	0.41
1:D:341[B]:ASP:HA	1:D:346:PHE:CD2	2.55	0.41
1:C:341:ASP:OD2	1:C:342:LYS:NZ	2.46	0.40
1:A:331[B]:LEU:HD23	1:A:331[B]:LEU:O	2.21	0.40

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	A	109/120 (91%)	108 (99%)	1 (1%)	0	100	100
1	B	112/120 (93%)	111 (99%)	1 (1%)	0	100	100
1	C	121/120 (101%)	119 (98%)	2 (2%)	0	100	100
1	D	113/120 (94%)	112 (99%)	1 (1%)	0	100	100
All	All	455/480 (95%)	450 (99%)	5 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	93/98 (95%)	91 (98%)	2 (2%)	47	16
1	B	96/98 (98%)	95 (99%)	1 (1%)	73	49
1	C	102/98 (104%)	100 (98%)	2 (2%)	50	20
1	D	96/98 (98%)	95 (99%)	1 (1%)	73	49
All	All	387/392 (99%)	381 (98%)	6 (2%)	58	29

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	277	ARG
1	A	322	MET
1	B	277	ARG
1	C	277	ARG
1	C	281	GLN
1	D	277	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	303	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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 ⓘ

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, 95th 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^2)	Q<0.9
1	A	108/120 (90%)	-0.73	0	100	100	3, 7, 15, 27	3 (2%)
1	B	110/120 (91%)	-0.61	1 (0%)	81	83	3, 8, 20, 35	4 (3%)
1	C	116/120 (96%)	-0.60	2 (1%)	69	70	3, 7, 18, 34	7 (6%)
1	D	111/120 (92%)	-0.42	2 (1%)	67	69	2, 11, 25, 42	4 (3%)
All	All	445/480 (92%)	-0.59	5 (1%)	77	79	2, 8, 20, 42	18 (4%)

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	254	ALA	3.5
1	C	322	MET	2.5
1	B	364	PRO	2.4
1	C	249	LYS	2.3
1	D	364	PRO	2.1

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 monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.