

Full wwPDB X-ray Structure Validation Report (i)

Aug 26, 2020 – 12:07 PM BST

PDB ID	:	5X4S
Title	:	Structure of the N-terminal domain (NTD) of SARS-CoV spike protein
Authors	:	Yuan, Y.; Zhang, Y.; Qi, J.; Shi, Y.; Gao, G.F.
Deposited on		
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 A user guide is available at https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

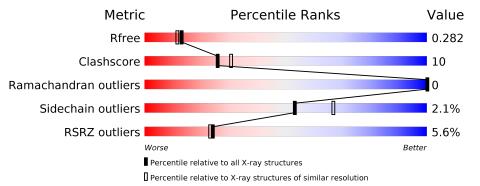
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	$7.0.044 (\mathrm{Gargrove})$
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13

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	$egin{array}{c} { m Whole \ archive} \ (\#{ m Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries},{ m resolution\ range}({ m \AA}))$
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 on 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	А	285	76%	18%	6%
2	В	2	100%		



2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2278 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 glycoprotein.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	А	269	Total 2152	C 1390	N 347	O 404	S 11	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

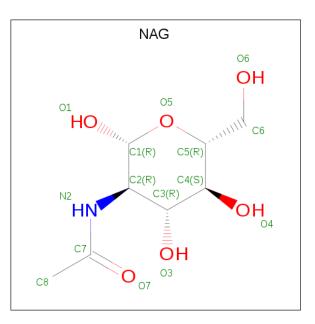
Chain	Residue	Modelled	Actual Comment		Reference
A	77	ASP	GLY	See Sequence details	UNP P59594
A	244	THR	ILE	See Sequence details	UNP P59594
A	293	HIS	-	expression tag	UNP P59594
A	294	HIS	-	expression tag	UNP P59594
А	295	HIS	-	expression tag	UNP P59594
А	296	HIS	-	expression tag	UNP P59594
A	297	HIS	-	expression tag	UNP P59594
А	298	HIS	-	expression tag	UNP P59594

• Molecule 2 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-a cetamido-2-deoxy-beta-D-glucopyranose.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	В	2	Total 28	C 16	N 2	O 10	0	0	0

• Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).





Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	Δ	1	Total	С	Ν	Ο	0	Ο
	11	L	14	8	1	5	0	0

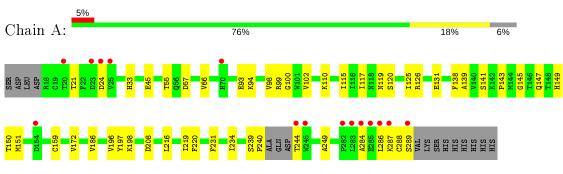
• Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	84	Total O 84 84	0	0



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 glycoprotein

• Molecule 2: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain B:

100%

NAG1 NAG2



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants	73.29Å 73.29Å 240.26Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 120.00°	Depositor
Resolution (Å)	38.31 - 2.20	Depositor
Resolution (A)	38.31 - 2.20	EDS
% Data completeness	99.5 (38.31-2.20)	Depositor
(in resolution range)	99.5(38.31 - 2.20)	EDS
R _{merge}	(Not available)	Depositor
R _{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$5.22 (at 2.20 \text{\AA})$	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
D D.	0.252 , 0.282	Depositor
R, R_{free}	0.252 , 0.282	DCC
R_{free} test set	1040 reflections (5.14%)	wwPDB-VP
Wilson B-factor $(Å^2)$	45.4	Xtriage
Anisotropy	0.142	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.33 , 53.8	EDS
L-test for twinning ²	$ L > = 0.44, < L^2 > = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	2278	wwPDB-VP
Average B, all atoms $(Å^2)$	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.09% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



 $^{^1 {\}rm Intensities}$ estimated from amplitudes.

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: 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		
	Unam	RMSZ	# Z > 5	RMSZ	# Z > 5	
1	А	0.29	0/2212	0.50	0/3008	

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	А	2152	0	2060	45	0
2	В	28	0	25	0	0
3	А	14	0	13	1	0
4	А	84	0	0	16	0
All	All	2278	0	2098	45	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 10.

All (45) 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:A:289:SER:N	4:A:701:HOH:O	1.91	1.04
1:A:286:LEU:CA	4:A:701:HOH:O	2.12	0.97
1:A:286:LEU:C	4:A:701:HOH:O	2.08	0.91
1:A:286:LEU:CB	4:A:701:HOH:O	2.22	0.84
1:A:244:THR:N	4:A:705:HOH:O	2.13	0.81
1:A:286:LEU:HB2	4:A:701:HOH:O	1.83	0.77
1:A:151:MET:SD	4:A:707:HOH:O	2.44	0.75
1:A:139:ALA:HB3	1:A:149:HIS:HB2	1.71	0.71
1:A:45:GLU:O	4:A:703:HOH:O	2.08	0.70
1:A:126:ARG:NH1	1:A:159:CYS:SG	2.66	0.69
1:A:57:ASP:OD2	4:A:704:HOH:O	2.10	0.69
1:A:286:LEU:N	4:A:701:HOH:O	2.27	0.65
1:A:288:CYS:C	4:A:701:HOH:O	2.26	0.65
1:A:99:ARG:HH22	1:A:172:VAL:HG13	1.64	0.63
1:A:33:HIS:ND1	1:A:208:ASP:OD1	2.29	0.63
1:A:287:LYS:N	1:A:288:CYS:HA	2.13	0.63
1:A:196:VAL:HB	1:A:220:PHE:HB2	1.82	0.61
1:A:126:ARG:NH1	4:A:708:HOH:O	2.26	0.59
1:A:102:VAL:HG22	1:A:234:ILE:HD11	1.89	0.55
1:A:287:LYS:NZ	4:A:713:HOH:O	2.39	0.55
1:A:99:ARG:HH21	1:A:120:SER:HB3	1.74	0.53
1:A:186:VAL:HG23	1:A:216:LEU:HD12	1.92	0.51
1:A:99:ARG:NH2	1:A:120:SER:HB3	2.27	0.49
1:A:131:GLU:HG3	3:A:601:NAG:HN2	1.77	0.49
1:A:99:ARG:HD3	1:A:138:PHE:CD2	2.48	0.49
1:A:286:LEU:HB2	1:A:289:SER:HB3	1.95	0.48
1:A:288:CYS:HA	4:A:701:HOH:O	2.14	0.47
1:A:288:CYS:CA	4:A:701:HOH:O	2.64	0.46
1:A:102:VAL:HG13	1:A:115:ILE:HG12	1.97	0.45
1:A:284:ALA:O	1:A:288:CYS:HB3	2.17	0.44
1:A:119:ASN:O	1:A:120:SER:OG	2.31	0.44
1:A:94:LYS:NZ	1:A:249:ALA:HB3	2.34	0.43
1:A:147:GLN:HB2	1:A:172:VAL:HG11	2.00	0.43
1:A:99:ARG:HH12	1:A:172:VAL:HG13	1.84	0.42
1:A:93:GLU:OE2	1:A:98:VAL:N	2.40	0.42
1:A:143:PRO:HD3	1:A:240:PRO:HA	2.02	0.41
1:A:21:THR:HA	4:A:709:HOH:O	2.20	0.41
1:A:102:VAL:O	1:A:231:PHE:HA	2.20	0.41
1:A:115:ILE:O	1:A:125:ILE:HA	2.20	0.41
1:A:239:SER:HA	1:A:240:PRO:HD3	1.91	0.41
1:A:94:LYS:HZ1	1:A:249:ALA:HB3	1.85	0.41
1:A:100:GLY:HA3	1:A:117:ILE:HG22	2.03	0.41



Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:141:SER:O	1:A:145:GLY:N	2.52	0.40
1:A:198:LYS:HD2	1:A:219:ILE:HD13	2.02	0.40
1:A:186:VAL:HB	1:A:197:TYR:HB2	2.02	0.40

Continued from previous page...

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	Perce	ntiles
1	А	265/285~(93%)	246~(93%)	19 (7%)	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	А	241/256~(94%)	236~(98%)	5(2%)	53 67	

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

Mol	Chain	Res	Type
1	А	24	ASP
1	А	55	THR
1	А	66	VAL

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type
1	А	110	LYS
1	А	150	THR

Some 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)

2 monosaccharides 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		les	
WIOI	туре	Chain	Ites		Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z >2
2	NAG	В	1	1,2	14, 14, 15	0.20	0	$17,\!19,\!21$	0.41	0
2	NAG	В	2	2	14,14,15	0.24	0	17,19,21	0.38	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
2	NAG	В	1	1,2	-	0/6/23/26	0/1/1/1
2	NAG	В	2	2	-	0/6/23/26	0/1/1/1

There are no bond length outliers.



There are no bond angle outliers.

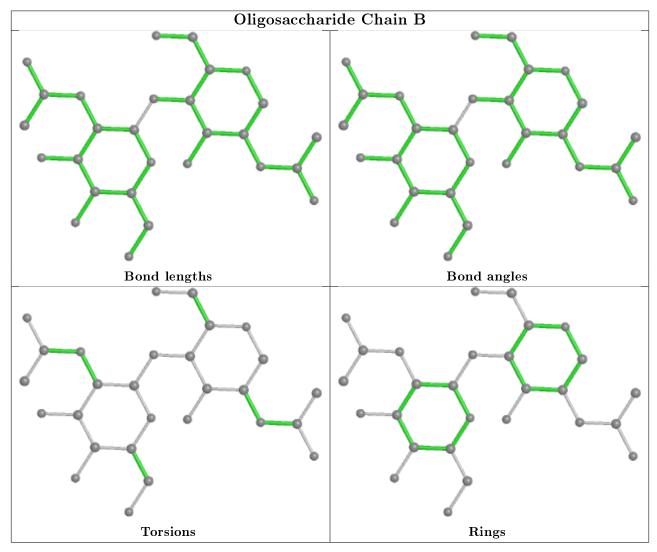
There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.



5.6 Ligand geometry (i)

1 ligand is 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			
	WIOI	туре	Ullalli	nes		Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
	3	NAG	А	601	1	14,14,15	0.38	0	17,19,21	0.38	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	\mathbf{Res}	Link	Chirals	Torsions	Rings
3	NAG	А	601	1	-	4/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	А	601	NAG	C1-C2-N2-C7
3	А	601	NAG	O5-C5-C6-O6
3	А	601	NAG	C3-C2-N2-C7
3	А	601	NAG	C4-C5-C6-O6

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	А	601	NAG	1	0

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^{th} 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	$OWAB(Å^2)$	$\mathbf{Q}{<}0.9$
1	А	269/285~(94%)	0.20	15 (5%) 24 23	35, 51, 80, 121	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ	
1	А	244	THR	7.4	
1	А	284	ALA	4.3	
1	А	70	HIS	3.8	
1	А	286	LEU	3.7	
1	А	245	TRP	3.6	
1	А	283	LEU	3.0	
1	А	24	ASP	2.8	
1	А	23	ASP	2.8	
1	А	25	VAL	2.7	
1	А	287	LYS	2.5	
1	А	20	THR	2.2	
1	А	285	GLU	2.2	
1	А	289	SER	2.1	
1	А	282	PRO	2.0	
1	А	154	ASP	2.0	

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

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

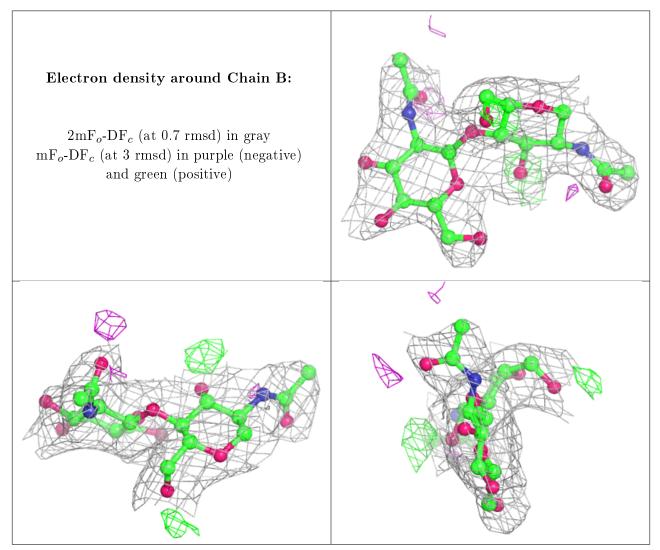
6.3 Carbohydrates (i)

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^{th} 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	$\mathbf{B} ext{-factors}(\mathbf{A}^2)$	Q<0.9
2	NAG	В	2	14/15	0.89	0.21	$49,\!60,\!76,\!82$	0
2	NAG	В	1	14/15	0.92	0.12	42,49,55,58	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.



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^{th} 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	${f B} ext{-factors}({f A}^2)$	Q<0.9
3	NAG	А	601	14/15	0.87	0.18	$69,\!85,\!89,\!94$	0



6.5 Other polymers (i)

There are no such residues in this entry.

