Full wwPDB X-ray Structure Validation Report ⓘ

Aug 10, 2020 – 08:31 AM BST

PDB ID	:	4ZPW
Title	:	Structure of unbound MERS-CoV spike receptor-binding domain (England1 strain).
Authors	:	Joyce, M.G.; Mascola, J.R.; Graham, B.S.; Kwong, P.D.
Deposited on	:	2015-05-08
Resolution	:	3.02 Å(reported)

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

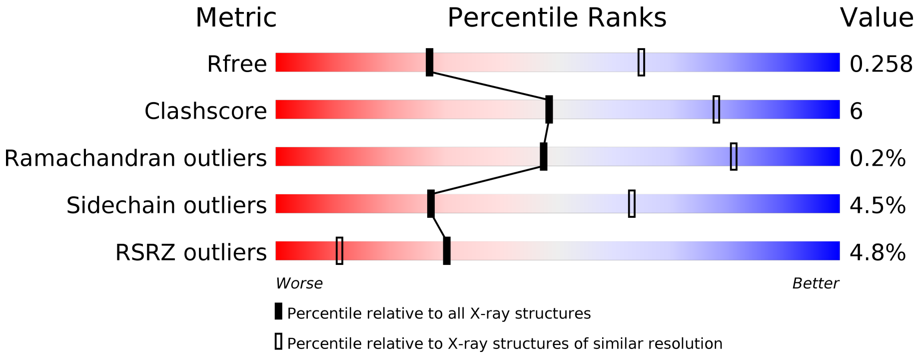
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 3.02 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	2399 (3.04-3.00)
Clashscore	141614	2734 (3.04-3.00)
Ramachandran outliers	138981	2640 (3.04-3.00)
Sidechain outliers	138945	2643 (3.04-3.00)
RSRZ outliers	127900	2287 (3.04-3.00)

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	R	208	
1	S	208	

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
2	NAG	R	601	-	-	-	X
2	NAG	S	603	-	-	-	X

2 Entry composition ⓘ

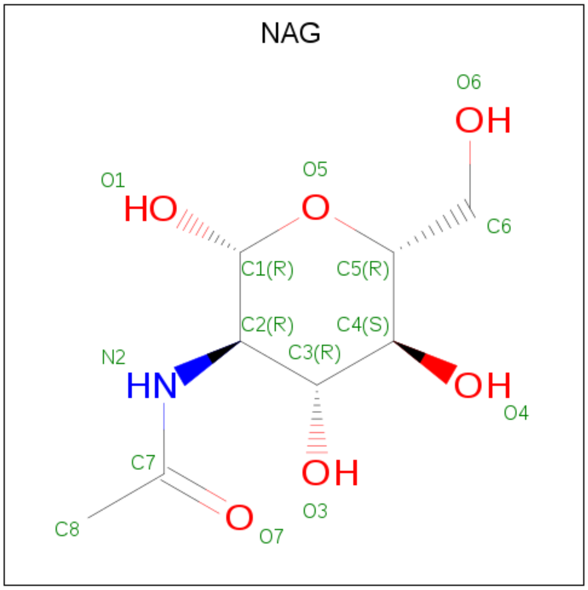
There are 4 unique types of molecules in this entry. The entry contains 3388 atoms, of which 0 are hydrogens and 0 are deuteriums.

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.

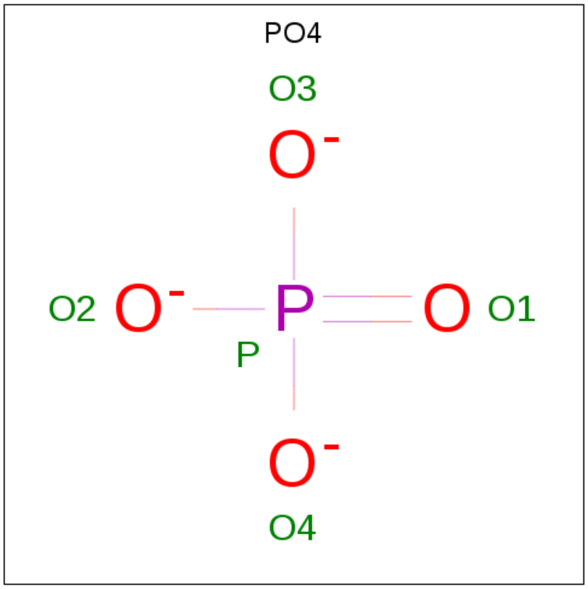
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	R	208	Total 1611	C 1029	N 256	O 315	S 11	0	0	0
1	S	208	Total 1611	C 1029	N 256	O 315	S 11	0	0	0

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



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	R	1	Total 14	C 8	N 1	O 5	0	0
2	R	1	Total 14	C 8	N 1	O 5	0	0
2	S	1	Total 14	C 8	N 1	O 5	0	0
2	S	1	Total 14	C 8	N 1	O 5	0	0

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



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

- Molecule 4 is water.

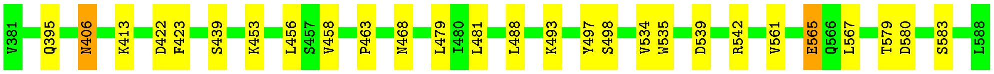
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	R	65	Total O 65 65	0	0
4	S	40	Total O 40 40	0	0

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.

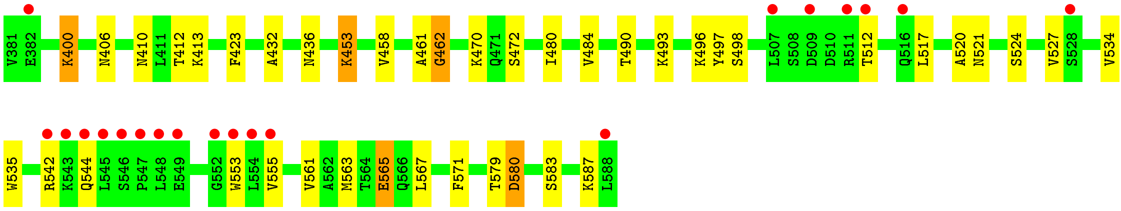
- Molecule 1: Spike glycoprotein

Chain R:



- Molecule 1: Spike glycoprotein

Chain S:



4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	46.67Å 109.89Å 125.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.73 – 3.02 42.95 – 3.02	Depositor EDS
% Data completeness (in resolution range)	90.5 (43.73-3.02) 90.6 (42.95-3.02)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.68 (at 3.01Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.217 , 0.259 0.221 , 0.258	Depositor DCC
R_{free} test set	609 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	57.7	Xtriage
Anisotropy	0.374	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 56.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	3388	wwPDB-VP
Average B, all atoms (Å^2)	61.0	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.*

---

[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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 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	R	0.24	0/1651	0.42	0/2254
1	S	0.24	0/1651	0.46	0/2254
All	All	0.24	0/3302	0.44	0/4508

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	R	1611	0	1570	16	0
1	S	1611	0	1570	28	0
2	R	28	0	26	1	0
2	S	28	0	26	2	0
3	S	5	0	0	0	0
4	R	65	0	0	1	0
4	S	40	0	0	0	0
All	All	3388	0	3192	39	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 6.

All (39) 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:S:461:ALA:N	1:S:462:GLY:HA3	2.08	0.68
1:S:542:ARG:HG2	1:S:555:VAL:HG22	1.79	0.64
1:R:422:ASP:HB3	1:R:481:LEU:HB2	1.80	0.63
1:R:456:LEU:HD13	1:R:479:LEU:HD21	1.84	0.60
1:S:480:ILE:HB	1:S:571:PHE:HB2	1.86	0.58
1:S:496:LYS:NZ	1:S:535:TRP:O	2.34	0.57
1:R:395:GLN:NE2	4:R:701:HOH:O	2.38	0.57
1:R:539:ASP:OD1	1:S:470:LYS:NZ	2.39	0.56
1:S:458:VAL:HB	1:S:461:ALA:HB3	1.87	0.56
1:R:406:ASN:HA	1:R:583:SER:HB2	1.89	0.55
1:S:472:SER:HA	1:S:520:ALA:HB1	1.88	0.54
1:R:535:TRP:CE2	1:S:517:LEU:HB2	2.43	0.53
1:S:410:ASN:HB2	1:S:587:LYS:HE3	1.91	0.53
1:S:432:ALA:O	1:S:436:ASN:ND2	2.39	0.52
1:S:498:SER:HB3	1:S:534:VAL:HG23	1.93	0.50
1:R:493:LYS:NZ	1:R:565:GLU:O	2.42	0.50
1:R:463:PRO:O	1:R:468:ASN:ND2	2.44	0.50
1:S:400:LYS:HE3	1:S:400:LYS:HA	1.93	0.49
1:R:498:SER:HB3	1:R:534:VAL:HG23	1.94	0.49
1:S:406:ASN:HA	1:S:583:SER:HB2	1.95	0.49
1:R:413:LYS:HE3	2:R:602:NAG:H82	1.95	0.48
1:R:542:ARG:HB2	1:S:458:VAL:HA	1.95	0.47
1:S:410:ASN:HB3	1:S:413:LYS:HB2	1.98	0.46
1:S:497:TYR:HB2	1:S:561:VAL:HB	1.98	0.46
1:S:493:LYS:NZ	1:S:565:GLU:O	2.37	0.46
1:R:493:LYS:HG2	1:R:567:LEU:HB2	1.97	0.46
1:S:472:SER:HB3	1:S:521:ASN:OD1	2.16	0.45
1:S:453:LYS:HD3	1:S:453:LYS:O	2.17	0.45
1:R:535:TRP:HB2	1:S:517:LEU:HD12	1.98	0.44
1:S:412:THR:HG1	2:S:603:NAG:HO6	1.66	0.44
1:R:535:TRP:CD2	1:S:517:LEU:HB2	2.52	0.43
1:R:453:LYS:HD3	1:R:481:LEU:HD22	1.99	0.43
1:S:484:VAL:HG13	1:S:567:LEU:O	2.18	0.43
1:S:524:SER:HB3	1:S:527:VAL:HG13	2.00	0.43
1:S:587:LYS:HD2	2:S:603:NAG:H82	2.02	0.42
1:R:497:TYR:HB2	1:R:561:VAL:HB	2.03	0.41
1:S:580:ASP:N	1:S:580:ASP:OD1	2.53	0.41
1:S:453:LYS:HD3	1:S:453:LYS:C	2.40	0.41
1:S:544:GLN:HB2	1:S:553:TRP:CH2	2.56	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	R	206/208 (99%)	201 (98%)	5 (2%)	0	100	100
1	S	206/208 (99%)	197 (96%)	8 (4%)	1 (0%)	29	66
All	All	412/416 (99%)	398 (97%)	13 (3%)	1 (0%)	47	81

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	S	462	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	R	190/190 (100%)	182 (96%)	8 (4%)	30	65
1	S	190/190 (100%)	181 (95%)	9 (5%)	26	61
All	All	380/380 (100%)	363 (96%)	17 (4%)	27	63

All (17) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	R	406	ASN
1	R	423	PHE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	R	439	SER
1	R	458	VAL
1	R	488	LEU
1	R	565	GLU
1	R	579	THR
1	R	580	ASP
1	S	400	LYS
1	S	423	PHE
1	S	453	LYS
1	S	490	THR
1	S	512	THR
1	S	563	MET
1	S	565	GLU
1	S	579	THR
1	S	580	ASP

Some 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 ⓘ

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

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	$\#\|Z\| > 2$
3	PO4	S	601	-	4,4,4	0.92	0	6,6,6	0.43	0
2	NAG	S	602	1	14,14,15	0.33	0	17,19,21	0.49	0
2	NAG	R	601	1	14,14,15	0.44	0	17,19,21	0.51	0
2	NAG	S	603	1	14,14,15	0.17	0	17,19,21	0.40	0
2	NAG	R	602	1	14,14,15	0.71	1 (7%)	17,19,21	0.75	1 (5%)

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	S	602	1	-	2/6/23/26	0/1/1/1
2	NAG	R	601	1	-	0/6/23/26	0/1/1/1
2	NAG	S	603	1	-	1/6/23/26	0/1/1/1
2	NAG	R	602	1	-	2/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	R	602	NAG	C1-C2	2.42	1.55	1.52

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	R	602	NAG	C1-O5-C5	2.33	115.35	112.19

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	R	602	NAG	C1-C2-N2-C7
2	S	603	NAG	O5-C5-C6-O6
2	S	602	NAG	O5-C5-C6-O6
2	S	602	NAG	C3-C2-N2-C7
2	R	602	NAG	C3-C2-N2-C7

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	S	603	NAG	2	0
2	R	602	NAG	1	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 ⓘ

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	R	208/208 (100%)	-0.25	0	100	100	29, 44, 82, 152	0
1	S	208/208 (100%)	0.23	20 (9%)	8	2	34, 61, 153, 216	0
All	All	416/416 (100%)	-0.01	20 (4%)	30	11	29, 51, 112, 216	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	S	511	ARG	5.8
1	S	546	SER	4.6
1	S	512	THR	4.5
1	S	547	PRO	4.5
1	S	553	TRP	4.0
1	S	588	LEU	3.7
1	S	552	GLY	3.3
1	S	544	GLN	3.2
1	S	549	GLU	2.9
1	S	548	LEU	2.9
1	S	507	LEU	2.7
1	S	382	GLU	2.7
1	S	543	LYS	2.6
1	S	516	GLN	2.5
1	S	509	ASP	2.5
1	S	545	LEU	2.2
1	S	528	SER	2.2
1	S	542	ARG	2.2
1	S	555	VAL	2.1
1	S	554	LEU	2.0

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 ⓘ

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, 95th 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(Å^2)	Q<0.9
2	NAG	S	602	14/15	0.57	0.30	99,120,127,128	0
2	NAG	S	603	14/15	0.63	0.53	146,166,170,171	0
2	NAG	R	601	14/15	0.67	0.50	145,161,170,175	0
2	NAG	R	602	14/15	0.73	0.28	126,136,138,139	0
3	PO4	S	601	5/5	0.85	0.32	157,158,159,159	0

6.5 Other polymers ⓘ

There are no such residues in this entry.