wwPDB X-ray Structure Validation Summary Report (i)

Nov 21, 2023 – 01:06 AM JST

PDB ID	:	7E7Y
Title	:	Crystal structure of the SARS-CoV-2 S RBD in complex with BD-623 Fab
Authors	:	Wei, Y.; Xiao, J.
Deposited on	:	2021-02-28
Resolution	:	2.41 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

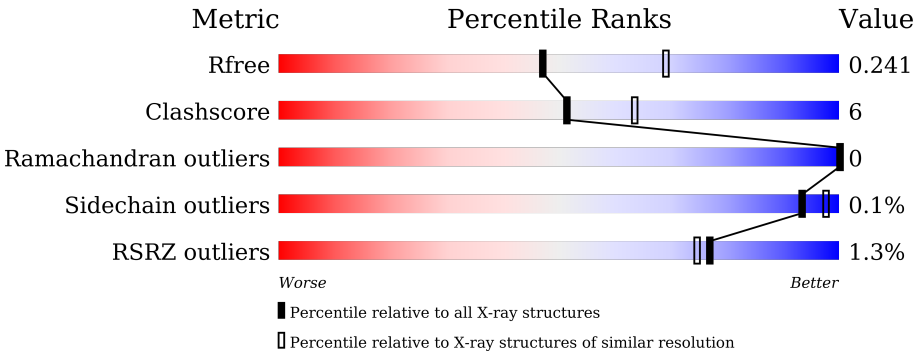
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.41 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	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)
RSRZ outliers	127900	4543 (2.44-2.40)

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	228	84% 11% .
1	C	228	86% 8% 5%
2	B	209	87% 13%
2	D	209	87% 12%
3	E	223	5% 77% 10% 13%
3	R	223	2% 74% 10% 15%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 9916 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 BD-623 Fab H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	218	Total	C	N	O	S	0	0	0
			1618	1021	273	317	7			
1	C	216	Total	C	N	O	S	0	0	0
			1601	1010	271	313	7			

- Molecule 2 is a protein called BD-623 Fab L.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	209	Total	C	N	O	S	0	0	0
			1543	965	257	316	5			
2	D	209	Total	C	N	O	S	0	0	0
			1543	965	257	316	5			

- Molecule 3 is a protein called Spike protein S1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	R	189	Total	C	N	O	S	0	0	0
			1498	958	249	283	8			
3	E	194	Total	C	N	O	S	0	0	0
			1536	984	256	288	8			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	126	Total	O	0	0
			126	126		
4	B	118	Total	O	0	0
			118	118		
4	R	85	Total	O	0	0
			85	85		
4	C	101	Total	O	0	0
			101	101		

Continued on next page...

Continued from previous page...

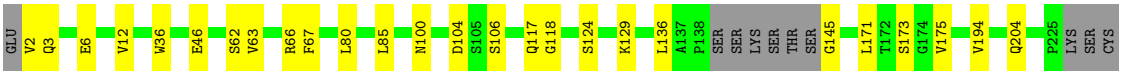
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	D	86	Total O 86 86	0	0
4	E	61	Total O 61 61	0	0

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.

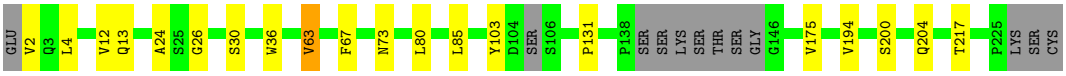
- Molecule 1: BD-623 Fab H

Chain A:



- Molecule 1: BD-623 Fab H

Chain C:



- Molecule 2: BD-623 Fab L

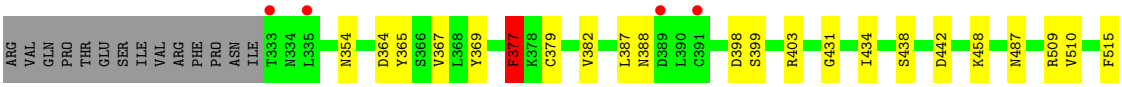
Chain B:

- Molecule 2: BD-623 Fab L

Chain D:

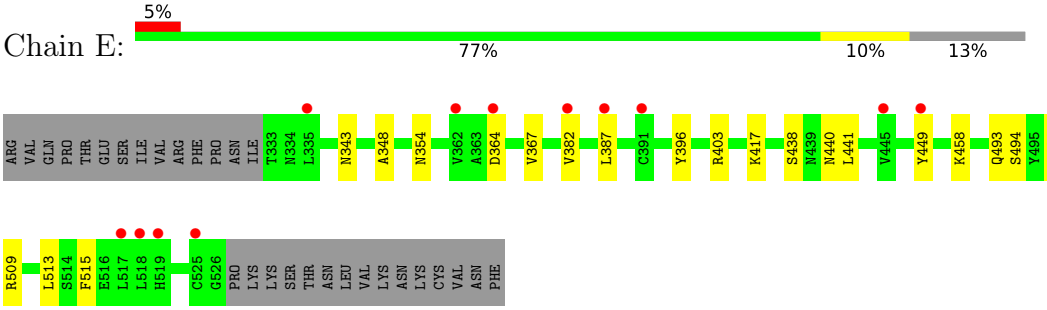
- Molecule 3: Spike protein S1

Chain R:



• Molecule 3: Spike protein S1

Chain E:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	93.82Å 126.63Å 138.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.09 – 2.41 49.09 – 2.41	Depositor EDS
% Data completeness (in resolution range)	98.1 (49.09-2.41) 92.1 (49.09-2.41)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.96 (at 2.42Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.197 , 0.241 0.198 , 0.241	Depositor DCC
R_{free} test set	2000 reflections (3.13%)	wwPDB-VP
Wilson B-factor (Å^2)	33.9	Xtriage
Anisotropy	0.122	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 36.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	9916	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.66% 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 (i)

5.1 Standard geometry (i)

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	0.48	0/1657	0.64	0/2257
1	C	0.46	1/1638 (0.1%)	0.63	0/2230
2	B	0.47	0/1580	0.60	0/2157
2	D	0.45	0/1580	0.64	1/2157 (0.0%)
3	E	0.46	0/1579	0.58	0/2149
3	R	0.51	1/1538 (0.1%)	0.64	2/2090 (0.1%)
All	All	0.47	2/9572 (0.0%)	0.62	3/13040 (0.0%)

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
3	R	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	R	516	GLU	CB-CG	5.68	1.62	1.52
1	C	63	VAL	CB-CG2	-5.62	1.41	1.52

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	R	458	LYS	CD-CE-NZ	-6.30	97.21	111.70
2	D	171	GLN	CA-CB-CG	5.34	125.16	113.40
3	R	377	PHE	CB-CG-CD2	-5.29	117.10	120.80

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	R	377	PHE	Sidechain

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	A	1618	0	1579	21	0
1	C	1601	0	1563	13	0
2	B	1543	0	1495	21	0
2	D	1543	0	1495	19	0
3	E	1536	0	1453	14	0
3	R	1498	0	1411	17	0
4	A	126	0	0	7	0
4	B	118	0	0	7	1
4	C	101	0	0	3	0
4	D	86	0	0	8	1
4	E	61	0	0	4	0
4	R	85	0	0	6	2
All	All	9916	0	8996	103	2

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.

The worst 5 of 103 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)
2:D:4:LEU:N	4:D:303:HOH:O	1.94	1.00
2:D:155:ASP:O	4:D:301:HOH:O	1.84	0.95
2:D:135:THR:OG1	4:D:302:HOH:O	1.86	0.94
2:B:138:CYS:SG	4:B:413:HOH:O	2.26	0.94
1:C:200:SER:HB2	1:C:204:GLN:HG2	1.51	0.93

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:R:658:HOH:O	4:D:372:HOH:O[3_544]	1.97	0.23
4:B:405:HOH:O	4:R:671:HOH:O[4_545]	2.16	0.04

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	214/228 (94%)	207 (97%)	7 (3%)	0	100	100
1	C	210/228 (92%)	203 (97%)	7 (3%)	0	100	100
2	B	207/209 (99%)	198 (96%)	9 (4%)	0	100	100
2	D	207/209 (99%)	201 (97%)	6 (3%)	0	100	100
3	E	192/223 (86%)	183 (95%)	9 (5%)	0	100	100
3	R	185/223 (83%)	175 (95%)	10 (5%)	0	100	100
All	All	1215/1320 (92%)	1167 (96%)	48 (4%)	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	A	180/190 (95%)	180 (100%)	0	100	100
1	C	178/190 (94%)	178 (100%)	0	100	100
2	B	173/174 (99%)	172 (99%)	1 (1%)	86	93

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	D	173/174 (99%)	173 (100%)	0	100 100
3	E	167/196 (85%)	167 (100%)	0	100 100
3	R	163/196 (83%)	163 (100%)	0	100 100
All	All	1034/1120 (92%)	1033 (100%)	1 (0%)	93 98

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	197	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	117	GLN
2	B	171	GLN
1	C	13	GLN
2	D	198	GLN
3	E	493	GLN

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)

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^{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)	Q<0.9
1	A	218/228 (95%)	-0.25	0 100 100	21, 32, 50, 56	0
1	C	216/228 (94%)	-0.32	0 100 100	27, 34, 52, 67	0
2	B	209/209 (100%)	-0.31	0 100 100	23, 35, 45, 52	0
2	D	209/209 (100%)	-0.19	0 100 100	29, 41, 59, 67	0
3	E	194/223 (86%)	0.14	12 (6%) 20 19	26, 42, 67, 99	0
3	R	189/223 (84%)	-0.03	4 (2%) 63 60	23, 37, 68, 97	0
All	All	1235/1320 (93%)	-0.17	16 (1%) 77 75	21, 36, 62, 99	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	R	333	THR	9.5
3	E	519	HIS	5.7
3	E	518	LEU	5.4
3	E	391	CYS	3.4
3	E	449	TYR	3.3

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.