



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 5, 2026 – 04:50 PM UTC

PDB ID : 7ZR2 / pdb_00007zr2
Title : Crystal structure of a chimeric protein mimic of SARS-CoV-2 Spike HR1 in complex with HR2
Authors : Camara-Artigas, A.; Gavira, J.A.; Cano-Munoz, M.; Polo-Megias, D.; Conejero-Lara, F.
Deposited on : 2022-05-03
Resolution : 1.45 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4-5-2 with Phenix2.0
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

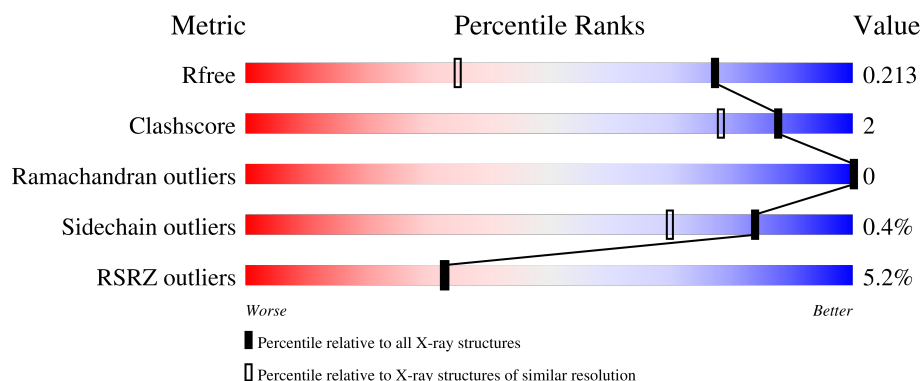
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.45 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}	180053	1756 (1.46-1.46)
Clashscore	190562	1795 (1.46-1.46)
Ramachandran outliers	187476	1776 (1.46-1.46)
Sidechain outliers	187428	1776 (1.46-1.46)
RSRZ outliers	180081	1756 (1.46-1.46)

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 $\leq 5\%$. 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	241	6% 91% 6%
2	B	44	89% 7%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4350 atoms, of which 2020 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 protein S2',Chimeric protein mimic of SARS-CoV-2 Spike HR1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	227	Total	C	H	N	O	0	2	0
			3444	1063	1721	306	354			

There are 105 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP P0DTC2
A	1	ASP	-	expression tag	UNP P0DTC2
A	13	LYS	GLN	engineered mutation	UNP P0DTC2
A	31	GLU	ALA	engineered mutation	UNP P0DTC2
A	38	GLU	VAL	engineered mutation	UNP P0DTC2
A	45	ASP	ALA	engineered mutation	UNP P0DTC2
A	59	ARG	ALA	engineered mutation	UNP P0DTC2
A	63	GLU	VAL	engineered mutation	UNP P0DTC2
A	74	GLY	VAL	engineered mutation	UNP P0DTC2
A	76	PRO	-	linker	UNP P0DTC2
A	77	ALA	-	linker	UNP P0DTC2
A	78	LYS	-	linker	UNP P0DTC2
A	79	ASP	-	linker	UNP P0DTC2
A	80	LEU	-	linker	UNP P0DTC2
A	81	ARG	-	linker	UNP P0DTC2
A	82	SER	-	linker	UNP P0DTC2
A	83	ASP	-	linker	UNP P0DTC2
A	84	ILE	-	linker	UNP P0DTC2
A	85	ASP	-	linker	UNP P0DTC2
A	86	ASN	-	linker	UNP P0DTC2
A	87	LEU	-	linker	UNP P0DTC2
A	88	GLU	-	linker	UNP P0DTC2
A	89	SER	-	linker	UNP P0DTC2
A	90	LYS	-	linker	UNP P0DTC2
A	91	ILE	-	linker	UNP P0DTC2
A	92	ALA	-	linker	UNP P0DTC2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	93	GLY	-	linker	UNP P0DTC2
A	94	PHE	-	linker	UNP P0DTC2
A	95	ASN	-	linker	UNP P0DTC2
A	96	SER	-	linker	UNP P0DTC2
A	97	SER	-	linker	UNP P0DTC2
A	98	LEU	-	linker	UNP P0DTC2
A	99	GLN	-	linker	UNP P0DTC2
A	100	LYS	-	linker	UNP P0DTC2
A	101	VAL	-	linker	UNP P0DTC2
A	102	LEU	-	linker	UNP P0DTC2
A	103	THR	-	linker	UNP P0DTC2
A	104	ASN	-	linker	UNP P0DTC2
A	105	LEU	-	linker	UNP P0DTC2
A	106	ALA	-	linker	UNP P0DTC2
A	107	GLN	-	linker	UNP P0DTC2
A	108	LYS	-	linker	UNP P0DTC2
A	109	ASN	-	linker	UNP P0DTC2
A	110	GLN	-	linker	UNP P0DTC2
A	111	ASN	-	linker	UNP P0DTC2
A	112	VAL	-	linker	UNP P0DTC2
A	113	GLU	-	linker	UNP P0DTC2
A	114	ASP	-	linker	UNP P0DTC2
A	115	LYS	-	linker	UNP P0DTC2
A	116	LEU	-	linker	UNP P0DTC2
A	117	LYS	-	linker	UNP P0DTC2
A	118	GLY	-	linker	UNP P0DTC2
A	119	LEU	-	linker	UNP P0DTC2
A	120	GLU	-	linker	UNP P0DTC2
A	121	SER	-	linker	UNP P0DTC2
A	122	ARG	-	linker	UNP P0DTC2
A	123	THR	-	linker	UNP P0DTC2
A	124	SER	-	linker	UNP P0DTC2
A	125	SER	-	linker	UNP P0DTC2
A	126	LEU	-	linker	UNP P0DTC2
A	127	GLU	-	linker	UNP P0DTC2
A	128	LYS	-	linker	UNP P0DTC2
A	129	GLN	-	linker	UNP P0DTC2
A	130	ILE	-	linker	UNP P0DTC2
A	131	LYS	-	linker	UNP P0DTC2
A	132	GLY	-	linker	UNP P0DTC2
A	133	ILE	-	linker	UNP P0DTC2
A	134	ALA	-	linker	UNP P0DTC2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	135	SER	-	linker	UNP P0DTC2
A	136	ASN	-	linker	UNP P0DTC2
A	137	PHE	-	linker	UNP P0DTC2
A	138	GLN	-	linker	UNP P0DTC2
A	139	ASN	-	linker	UNP P0DTC2
A	140	GLU	-	linker	UNP P0DTC2
A	141	ILE	-	linker	UNP P0DTC2
A	142	LEU	-	linker	UNP P0DTC2
A	143	LYS	-	linker	UNP P0DTC2
A	144	GLN	-	linker	UNP P0DTC2
A	145	ARG	-	linker	UNP P0DTC2
A	146	GLU	-	linker	UNP P0DTC2
A	147	TYR	-	linker	UNP P0DTC2
A	148	LEU	-	linker	UNP P0DTC2
A	149	VAL	-	linker	UNP P0DTC2
A	150	ASN	-	linker	UNP P0DTC2
A	151	LYS	-	linker	UNP P0DTC2
A	152	GLY	-	linker	UNP P0DTC2
A	153	SER	-	linker	UNP P0DTC2
A	154	GLY	-	linker	UNP P0DTC2
A	155	ASN	-	linker	UNP P0DTC2
A	165	GLU	ALA	engineered mutation	UNP P0DTC2
A	183	LYS	ALA	engineered mutation	UNP P0DTC2
A	190	LYS	GLN	engineered mutation	UNP P0DTC2
A	197	LYS	ALA	engineered mutation	UNP P0DTC2
A	222	LYS	LEU	engineered mutation	UNP P0DTC2
A	230	GLY	-	expression tag	UNP P0DTC2
A	231	GLY	-	expression tag	UNP P0DTC2
A	232	GLY	-	expression tag	UNP P0DTC2
A	233	GLY	-	expression tag	UNP P0DTC2
A	234	SER	-	expression tag	UNP P0DTC2
A	235	HIS	-	expression tag	UNP P0DTC2
A	236	HIS	-	expression tag	UNP P0DTC2
A	237	HIS	-	expression tag	UNP P0DTC2
A	238	HIS	-	expression tag	UNP P0DTC2
A	239	HIS	-	expression tag	UNP P0DTC2
A	240	HIS	-	expression tag	UNP P0DTC2

- Molecule 2 is a protein called Spike protein S2'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	41	Total	C	H	N	O	0	0	0
			602	186	299	52	65			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1163	ACE	-	acetylation	UNP P0DTC2
B	1203	SER	-	expression tag	UNP P0DTC2
B	1204	GLY	-	expression tag	UNP P0DTC2
B	1205	GLY	-	expression tag	UNP P0DTC2
B	1206	TYR	-	expression tag	UNP P0DTC2

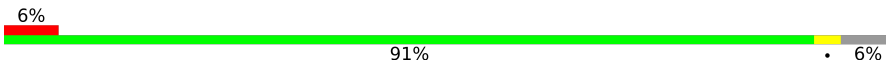
- Molecule 3 is water.

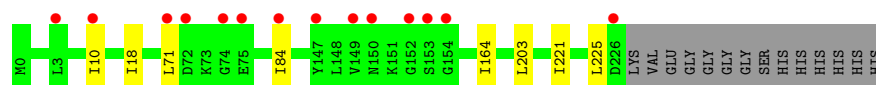
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	243	Total O 243 243	0	0
3	B	61	Total O 61 61	0	0

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.

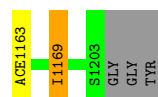
- Molecule 1: Spike protein S2',Chimeric protein mimic of SARS-CoV-2 Spike HR1

Chain A: 



- Molecule 2: Spike protein S2'

Chain B: 



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	166.78Å 37.37Å 43.91Å 90.00° 95.67° 90.00°	Depositor
Resolution (Å)	18.80 – 1.45 18.80 – 1.45	Depositor EDS
% Data completeness (in resolution range)	96.0 (18.80-1.45) 96.0 (18.80-1.45)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.72 (at 1.45Å)	Xtriage
Refinement program	PHENIX 1.20_4459	Depositor
R, R_{free}	0.174 , 0.209 0.178 , 0.213	Depositor DCC
R_{free} test set	2266 reflections (4.70%)	wwPDB-VP
Wilson B-factor (Å ²)	15.1	Xtriage
Anisotropy	0.245	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 37.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	4350	wwPDB-VP
Average B, all atoms (Å ²)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.15% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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.

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: ACE

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	0.44	1/1732 (0.1%)	0.48	0/2331
2	B	0.44	1/300 (0.3%)	0.47	0/406
All	All	0.44	2/2032 (0.1%)	0.48	0/2737

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	18	ILE	C-O	-6.34	1.16	1.24
2	B	1163	ACE	C-N	5.01	1.43	1.33

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	A	1723	1721	1722	8	0
2	B	303	299	303	1	0
3	A	243	0	0	0	0
3	B	61	0	0	0	0
All	All	2330	2020	2025	8	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 (8) 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:10:ILE:HD11	1:A:164:ILE:HD13	1.55	0.87
1:A:84:ILE:HD12	1:A:225:LEU:HD11	1.67	0.75
1:A:10:ILE:CD1	1:A:164:ILE:HD13	2.27	0.61
1:A:84:ILE:HD13	1:A:221:ILE:HG22	1.93	0.51
1:A:84:ILE:CD1	1:A:225:LEU:HD11	2.39	0.49
1:A:84:ILE:HD13	1:A:221:ILE:CG2	2.45	0.47
1:A:71:LEU:HD11	1:A:221:ILE:HG23	1.98	0.45
1:A:203:LEU:HD12	2:B:1169:ILE:HD11	1.98	0.45

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	227/241 (94%)	225 (99%)	2 (1%)	0	100	100
2	B	39/44 (89%)	39 (100%)	0	0	100	100
All	All	266/285 (93%)	264 (99%)	2 (1%)	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	191/213 (90%)	191 (100%)	0	100	100
2	B	34/37 (92%)	33 (97%)	1 (3%)	37	7
All	All	225/250 (90%)	224 (100%)	1 (0%)	84	70

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

Mol	Chain	Res	Type
2	B	1169	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	41	GLN
1	A	109	ASN
1	A	110	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 [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, 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	227/241 (94%)	0.13	14 (6%) 26 26	8, 24, 44, 61	2 (0%)
2	B	40/44 (90%)	-0.22	0 100 100	13, 21, 34, 41	0
All	All	267/285 (93%)	0.08	14 (5%) 33 33	8, 23, 44, 61	2 (0%)

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	147	TYR	4.5
1	A	149	VAL	4.3
1	A	10	ILE	4.1
1	A	226	ASP	3.7
1	A	153	SER	3.7
1	A	74	GLY	2.8
1	A	154	GLY	2.7
1	A	3	LEU	2.5
1	A	150	ASN	2.5
1	A	72	ASP	2.4
1	A	152	GLY	2.3
1	A	84	ILE	2.3
1	A	75	GLU	2.3
1	A	71	LEU	2.2

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](#)

There are no oligosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.