WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Sep 10, 2023 – 10:35 AM EDT

PDB ID	:	4JHW
Title	:	Crystal Structure of Respiratory Syncytial Virus Fusion Glycoprotein Stabilized in the Prefusion Conformation by Human Antibody D25
Authors	:	Mclellan, J.S.; Chen, M.; Leung, S.; Graepel, K.W.; Du, X.; Yang, Y.; Zhou, T.; Baxa, U.; Yasuda, E.; Beaumont, T.; Kumar, A.; Modjarrad, K.; Zheng, Z.; Zhao, M.; Xia, N.; Kwong, P.D.; Graham, B.S.
Deposited on	:	2013-03-05
Resolution	:	3.60 Å(reported)

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

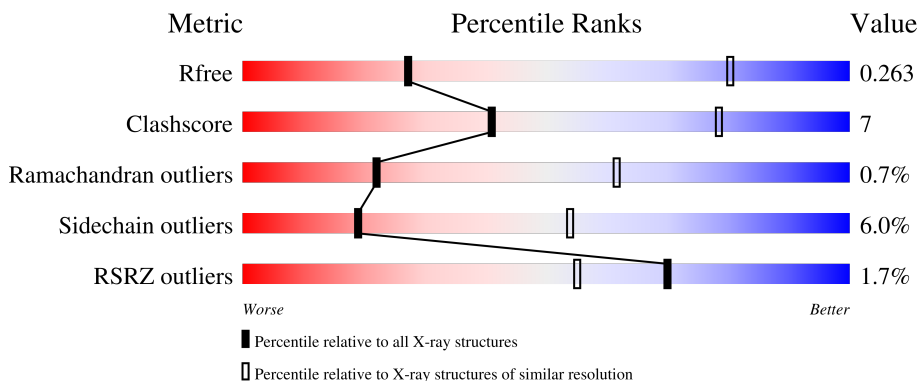|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.60 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		0.263
Clashscore		7
Ramachandran outliers		0.7%
Sidechain outliers		6.0%
RSRZ outliers		1.7%

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}	130704	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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	H	230	3% 83% 14% ··
2	L	214	3% 84% 15% ·
3	F	498	% 63% 24% · 10%

WORLDWIDE wwPDB PROTEIN DATA BANK

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 6778 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 D25 antigen-binding fragment heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	H	226	Total	C	N	O	S	0	0	0
			1683	1067	279	330	7			

- Molecule 2 is a protein called D25 light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	L	211	Total	C	N	O	S	0	0	0
			1619	1014	269	331	5			

- Molecule 3 is a protein called Fusion glycoprotein F0.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	449	Total	C	N	O	S	0	0	0
			3476	2195	573	687	21			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	129	ALA	PRO	engineered mutation	UNP P03420
F	379	VAL	ILE	engineered mutation	UNP P03420
F	447	VAL	MET	engineered mutation	UNP P03420
F	514	SER	-	expression tag	UNP P03420
F	515	ALA	-	expression tag	UNP P03420
F	516	ILE	-	expression tag	UNP P03420
F	517	GLY	-	expression tag	UNP P03420
F	518	GLY	-	expression tag	UNP P03420
F	519	TYR	-	expression tag	UNP P03420
F	520	ILE	-	expression tag	UNP P03420
F	521	PRO	-	expression tag	UNP P03420
F	522	GLU	-	expression tag	UNP P03420
F	523	ALA	-	expression tag	UNP P03420
F	524	PRO	-	expression tag	UNP P03420

Continued on next page...

Continued from previous page...

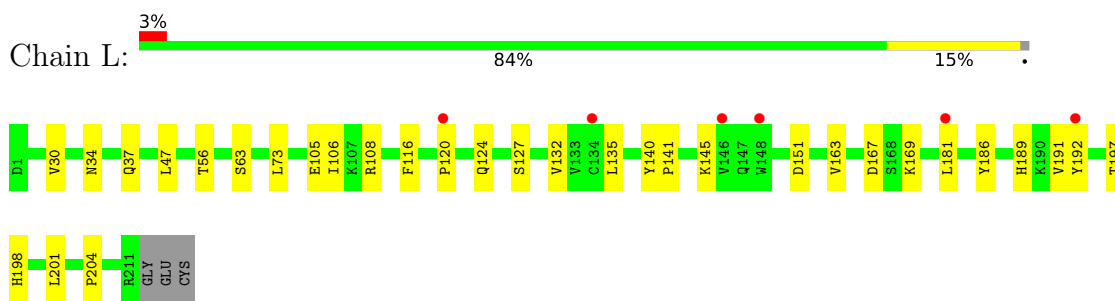| Chain	Residue	Modelled	Actual	Comment	Reference
F	525	ARG	-	expression tag	UNP P03420
F	526	ASP	-	expression tag	UNP P03420
F	527	GLY	-	expression tag	UNP P03420
F	528	GLN	-	expression tag	UNP P03420
F	529	ALA	-	expression tag	UNP P03420
F	530	TYR	-	expression tag	UNP P03420
F	531	VAL	-	expression tag	UNP P03420
F	532	ARG	-	expression tag	UNP P03420
F	533	LYS	-	expression tag	UNP P03420
F	534	ASP	-	expression tag	UNP P03420
F	535	GLY	-	expression tag	UNP P03420
F	536	GLU	-	expression tag	UNP P03420
F	537	TRP	-	expression tag	UNP P03420
F	538	VAL	-	expression tag	UNP P03420
F	539	LEU	-	expression tag	UNP P03420
F	540	LEU	-	expression tag	UNP P03420
F	541	SER	-	expression tag	UNP P03420
F	542	THR	-	expression tag	UNP P03420
F	543	PHE	-	expression tag	UNP P03420
F	544	LEU	-	expression tag	UNP P03420
F	545	GLY	-	expression tag	UNP P03420
F	546	GLY	-	expression tag	UNP P03420
F	547	LEU	-	expression tag	UNP P03420
F	548	VAL	-	expression tag	UNP P03420
F	549	PRO	-	expression tag	UNP P03420
F	550	ARG	-	expression tag	UNP P03420

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: D25 antigen-binding fragment heavy chain
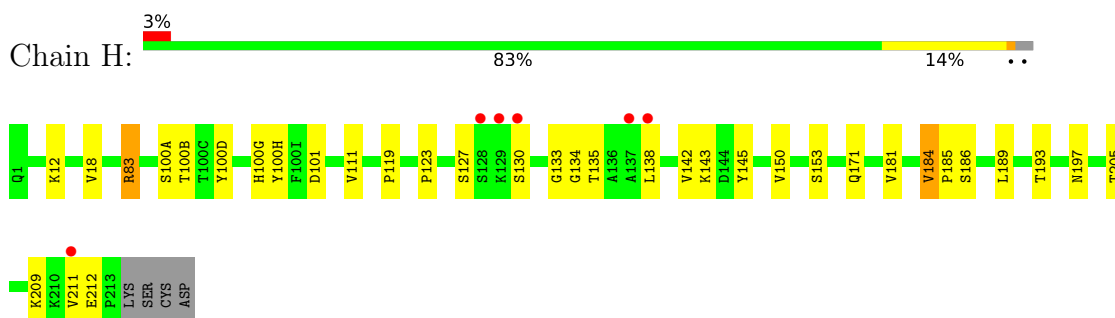
• Molecule 2: D25 light chain

• Molecule 3: Fusion glycoprotein F0

E497 K498 I499 N500 Q501 S502 L503 A504 F505 L513 SER ALA ILE GLY GLY TYR ILE PRO GLU ALA PRO ARG ASP GLY GLN ALA TYR VAL ARG LYS ASP GLY GLU TRP VAL LEU LEU SER THR PHE LEU GLY LEU VAL PRO ARG

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, α, β, γ	152.29Å 152.29Å 152.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.24 – 3.60 42.24 – 3.60	Depositor EDS
% Data completeness (in resolution range)	99.6 (42.24-3.60) 99.6 (42.24-3.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.03 (at 3.57Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.213 , 0.267 0.214 , 0.263	Depositor DCC
R_{free} test set	671 reflections (4.83%)	wwPDB-VP
Wilson B-factor (Å^2)	156.0	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 80.3	EDS
L-test for twinning [2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.039 for l,-k,h	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6778	wwPDB-VP
Average B, all atoms (Å^2)	128.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.85% 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	H	0.22	0/1726	0.48	0/2363
2	L	0.24	0/1652	0.48	0/2247
3	F	0.25	0/3525	0.50	0/4775
All	All	0.24	0/6903	0.49	0/9385

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	H	1683	0	1662	17	0
2	L	1619	0	1579	16	0
3	F	3476	0	3522	67	0
All	All	6778	0	6763	96	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 7.

The worst 5 of 96 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:L:56:THR:HG21	3:F:210:GLN:HE21	1.49	0.77

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:F:167:ILE:HD11	3:F:181:LEU:HD11	1.70	0.74
2:L:120:PRO:HD3	2:L:132:VAL:HG22	1.71	0.73
3:F:45:LEU:HB3	3:F:364:ARG:HH21	1.51	0.72
3:F:407:ILE:HD11	3:F:457:TYR:HB3	1.73	0.69

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	H	224/230 (97%)	219 (98%)	5 (2%)	0	100	100
2	L	209/214 (98%)	202 (97%)	6 (3%)	1 (0%)	29	68
3	F	445/498 (89%)	387 (87%)	53 (12%)	5 (1%)	14	53
All	All	878/942 (93%)	808 (92%)	64 (7%)	6 (1%)	22	61

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	F	182	SER
3	F	210	GLN
3	F	212	CYS
3	F	505	PHE
2	L	30	VAL

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	H	191/195 (98%)	184 (96%)	7 (4%)	34	66
2	L	186/188 (99%)	182 (98%)	4 (2%)	52	77
3	F	410/448 (92%)	374 (91%)	36 (9%)	10	40
All	All	787/831 (95%)	740 (94%)	47 (6%)	19	54

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	F	209	LYS
3	F	311	THR
3	F	210	GLN
3	F	232	GLU
3	F	364	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
3	F	210	GLN
3	F	270	GLN
3	F	354	GLN
3	F	302	GLN
3	F	26	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 monosaccharides 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	H	226/230 (98%)	0.06	6 (2%)	54	38	77, 151, 223, 256	0
2	L	211/214 (98%)	0.10	6 (2%)	53	37	62, 126, 222, 251	0
3	F	449/498 (90%)	-0.09	3 (0%)	87	78	66, 107, 166, 225	0
All	All	886/942 (94%)	-0.00	15 (1%)	70	55	62, 120, 214, 256	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	211	VAL	4.5
1	H	130	SER	4.0
1	H	138	LEU	3.4
1	H	129	LYS	2.8
3	F	356	GLU	2.7

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 monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.