



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 9, 2023 – 11:30 AM EDT

PDB ID : 8DFG  
Title : Crystal structure of potently neutralizing human monoclonal antibody 42D6 Fab in complex with MSP1-19  
Authors : Patel, P.N.; Tang, W.K.; Tolia, N.H.  
Deposited on : 2022-06-22  
Resolution : 2.00 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](mailto: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 : 4.02b-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

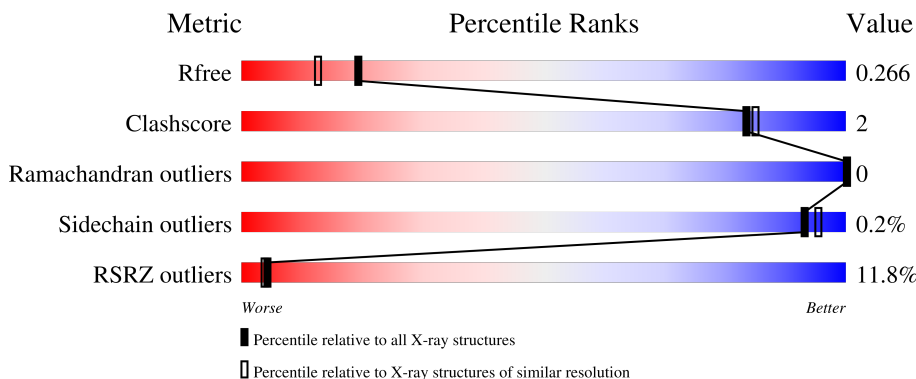
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

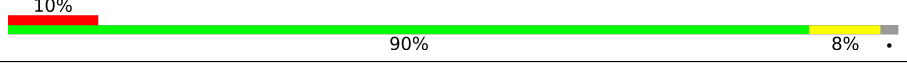
The reported resolution of this entry is 2.00 Å.

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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	H	242	
1	I	242	
2	L	217	
2	M	217	
3	A	105	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	B	105	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into three segments: a red segment on the left labeled '25%', a green segment in the middle labeled '84%', and a grey segment on the right labeled '14%'. A small yellow dot is visible at the end of the green segment.</p>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 15897 atoms, of which 7571 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 42D6 Fab Heavy Chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	H	218	3258	1063	1587	280	324	4	0	0	0
1	I	216	3274	1056	1616	277	321	4	0	0	0

- Molecule 2 is a protein called 42D6 Fab Light Chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	L	213	3221	1025	1585	278	327	6	0	0	0
2	M	213	3221	1025	1585	278	327	6	0	0	0

- Molecule 3 is a protein called Merozoite surface protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	B	90	1291	411	599	122	147	12	0	0	0
3	A	90	1291	411	599	122	147	12	0	0	0

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	GLU	-	expression tag	UNP Q8I0U8
B	-1	THR	-	expression tag	UNP Q8I0U8
B	0	GLY	-	expression tag	UNP Q8I0U8
B	3	ALA	SER	engineered mutation	UNP Q8I0U8
B	48	ALA	THR	engineered mutation	UNP Q8I0U8
B	94	GLY	-	expression tag	UNP Q8I0U8
B	95	THR	-	expression tag	UNP Q8I0U8
B	96	LYS	-	expression tag	UNP Q8I0U8

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	97	HIS	-	expression tag	UNP Q8I0U8
B	98	HIS	-	expression tag	UNP Q8I0U8
B	99	HIS	-	expression tag	UNP Q8I0U8
B	100	HIS	-	expression tag	UNP Q8I0U8
B	101	HIS	-	expression tag	UNP Q8I0U8
B	102	HIS	-	expression tag	UNP Q8I0U8
A	-2	GLU	-	expression tag	UNP Q8I0U8
A	-1	THR	-	expression tag	UNP Q8I0U8
A	0	GLY	-	expression tag	UNP Q8I0U8
A	3	ALA	SER	engineered mutation	UNP Q8I0U8
A	48	ALA	THR	engineered mutation	UNP Q8I0U8
A	94	GLY	-	expression tag	UNP Q8I0U8
A	95	THR	-	expression tag	UNP Q8I0U8
A	96	LYS	-	expression tag	UNP Q8I0U8
A	97	HIS	-	expression tag	UNP Q8I0U8
A	98	HIS	-	expression tag	UNP Q8I0U8
A	99	HIS	-	expression tag	UNP Q8I0U8
A	100	HIS	-	expression tag	UNP Q8I0U8
A	101	HIS	-	expression tag	UNP Q8I0U8
A	102	HIS	-	expression tag	UNP Q8I0U8

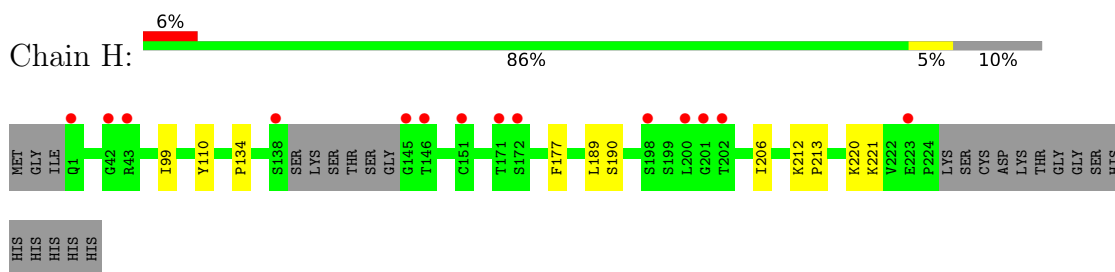
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	H	89	Total O 89 89	0	0
4	L	81	Total O 81 81	0	0
4	I	76	Total O 76 76	0	0
4	M	69	Total O 69 69	0	0
4	B	12	Total O 12 12	0	0
4	A	14	Total O 14 14	0	0

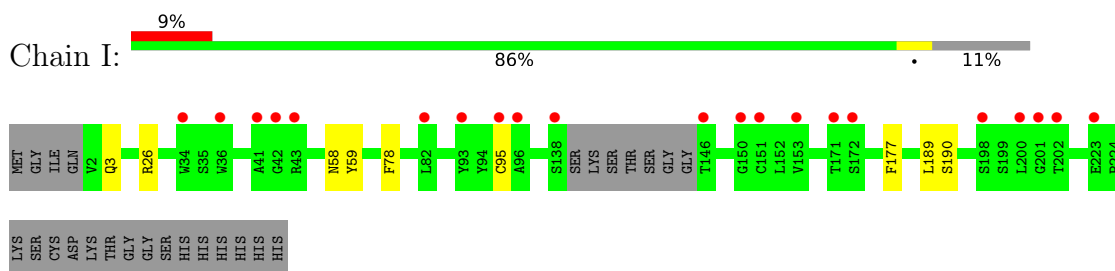
### 3 Residue-property plots [i](#)

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.

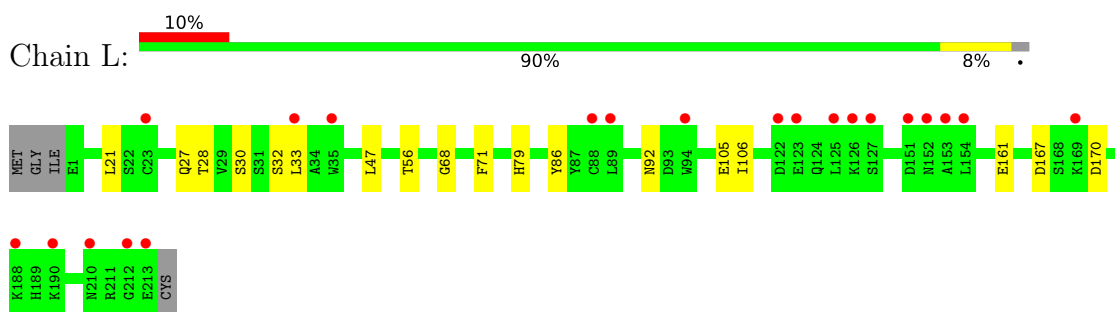
- Molecule 1: 42D6 Fab Heavy Chain



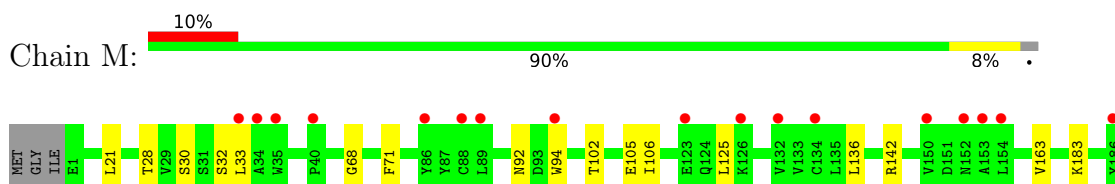
- Molecule 1: 42D6 Fab Heavy Chain



- Molecule 2: 42D6 Fab Light Chain

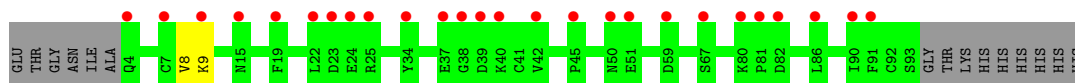
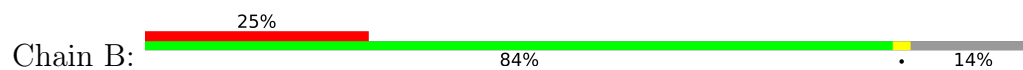


- Molecule 2: 42D6 Fab Light Chain

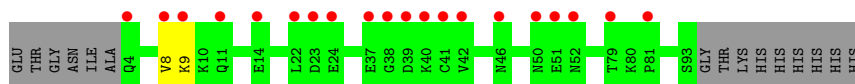
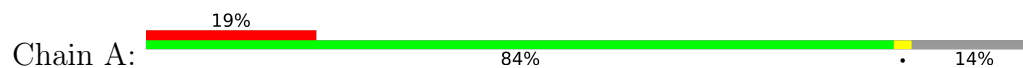




- Molecule 3: Merozoite surface protein 1



- Molecule 3: Merozoite surface protein 1



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	41.68Å 71.37Å 196.50Å 90.00° 94.64° 90.00°	Depositor
Resolution (Å)	19.92 – 2.00 19.92 – 2.00	Depositor EDS
% Data completeness (in resolution range)	96.3 (19.92-2.00) 96.4 (19.92-2.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.00 (at 1.99Å)	Xtrriage
Refinement program	PHENIX 1.16_3549	Depositor
R, $R_{free}$	0.225 , 0.266 0.225 , 0.266	Depositor DCC
$R_{free}$ test set	3760 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.7	Xtrriage
Anisotropy	0.336	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.44 , 50.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.56$ , $\langle L^2 \rangle = 0.40$	Xtrriage
Estimated twinning fraction	0.000 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	15897	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 61.31 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.3346e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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](#)

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	0.66	0/1715	0.67	0/2340
1	I	0.49	0/1702	0.59	0/2323
2	L	0.60	0/1674	0.64	0/2276
2	M	0.49	0/1674	0.60	0/2276
3	A	0.40	0/703	0.53	0/944
3	B	0.38	0/703	0.48	0/944
All	All	0.54	0/8171	0.61	0/11103

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	H	1671	1587	1629	9	0
1	I	1658	1616	1615	5	0
2	L	1636	1585	1588	12	0
2	M	1636	1585	1588	11	0
3	A	692	599	621	1	0
3	B	692	599	621	1	0
4	A	14	0	0	0	0
4	B	12	0	0	0	0
4	H	89	0	0	0	0
4	I	76	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	L	81	0	0	3	0
4	M	69	0	0	0	0
All	All	8326	7571	7662	38	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.

The worst 5 of 38 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:M:21:LEU:HD22	2:M:102:THR:HG21	1.64	0.79
2:M:33:LEU:HD22	2:M:71:PHE:CG	2.27	0.70
2:L:28:THR:HG23	2:L:68:GLY:HA2	1.78	0.66
1:H:177:PHE:O	1:H:189:LEU:HD13	2.01	0.60
1:H:189:LEU:HD12	1:H:190:SER:N	2.17	0.58

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	H	214/242 (88%)	210 (98%)	4 (2%)	0	100	100
1	I	212/242 (88%)	209 (99%)	3 (1%)	0	100	100
2	L	211/217 (97%)	207 (98%)	4 (2%)	0	100	100
2	M	211/217 (97%)	206 (98%)	5 (2%)	0	100	100
3	A	88/105 (84%)	84 (96%)	4 (4%)	0	100	100
3	B	88/105 (84%)	85 (97%)	3 (3%)	0	100	100
All	All	1024/1128 (91%)	1001 (98%)	23 (2%)	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	H	188/208 (90%)	188 (100%)	0	100	100
1	I	187/208 (90%)	186 (100%)	1 (0%)	88	92
2	L	185/188 (98%)	184 (100%)	1 (0%)	88	92
2	M	185/188 (98%)	185 (100%)	0	100	100
3	A	81/93 (87%)	81 (100%)	0	100	100
3	B	81/93 (87%)	81 (100%)	0	100	100
All	All	907/978 (93%)	905 (100%)	2 (0%)	93	95

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	L	79	HIS
1	I	58	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	M	79	HIS

### 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 [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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	H	218/242 (90%)	0.33	14 (6%) 19 18	18, 29, 50, 64	0
1	I	216/242 (89%)	0.53	21 (9%) 7 7	27, 38, 62, 81	0
2	L	213/217 (98%)	0.47	21 (9%) 7 6	19, 30, 51, 67	0
2	M	213/217 (98%)	0.68	21 (9%) 7 6	25, 39, 66, 93	0
3	A	90/105 (85%)	1.14	20 (22%) 0 0	27, 48, 76, 89	0
3	B	90/105 (85%)	1.47	26 (28%) 0 0	35, 57, 84, 92	0
All	All	1040/1128 (92%)	0.64	123 (11%) 4 4	18, 37, 69, 93	0

The worst 5 of 123 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	39	ASP	7.1
3	B	39	ASP	6.4
2	L	213	GLU	6.3
3	B	22	LEU	5.9
2	M	153	ALA	5.2

### 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 monosaccharides 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.