



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 03:53 am BST

PDB ID : 4K2U  
Title : Crystal structure of PfEBA-175 F1 in complex with R218 antibody Fab fragment  
Authors : Tolia, N.H.  
Deposited on : 2013-04-09  
Resolution : 2.45 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 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 : 2.11  
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) : 2.11

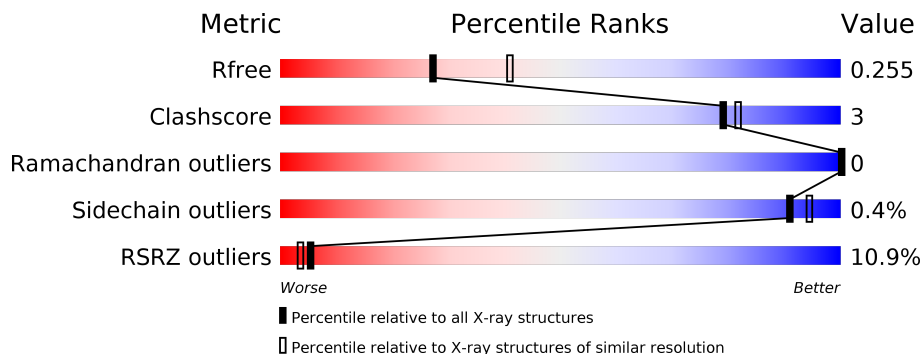
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.45 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))
$R_{free}$	130704	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

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  $\geq 3$ , 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	A	297	
1	B	297	
2	H	233	
2	I	233	
3	L	234	
3	M	234	

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:

<b>Mol</b>	<b>Type</b>	<b>Chain</b>	<b>Res</b>	<b>Chirality</b>	<b>Geometry</b>	<b>Clashes</b>	<b>Electron density</b>
4	SO4	I	301	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 20975 atoms, of which 10201 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 Erythrocyte binding antigen 175.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	A	230	Total 3869	C 1236	H 1919	N 345	O 353	S 16	0	0	0
1	B	257	Total 4313	C 1376	H 2144	N 376	O 398	S 19	0	0	0

- Molecule 2 is a protein called Antibody Heavy Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
2	H	208	Total 3143	C 1023	H 1544	N 257	O 314	S 5	0	0	0
2	I	214	Total 3235	C 1049	H 1589	N 267	O 324	S 6	0	0	0

- Molecule 3 is a protein called Antibody Light Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
3	L	208	Total 3146	C 1009	H 1526	N 273	O 332	S 6	0	0	0
3	M	200	Total 3038	C 970	H 1479	N 262	O 321	S 6	0	0	0

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).

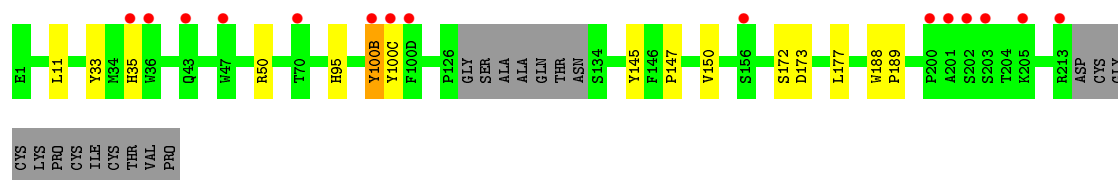


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	B	1	Total O S 5 4 1	0	0
4	H	1	Total O S 5 4 1	0	0
4	H	1	Total O S 5 4 1	0	0
4	H	1	Total O S 5 4 1	0	0
4	I	1	Total O S 5 4 1	0	0
4	I	1	Total O S 5 4 1	0	0
4	I	1	Total O S 5 4 1	0	0
4	L	1	Total O S 5 4 1	0	0
4	L	1	Total O S 5 4 1	0	0
4	M	1	Total O S 5 4 1	0	0

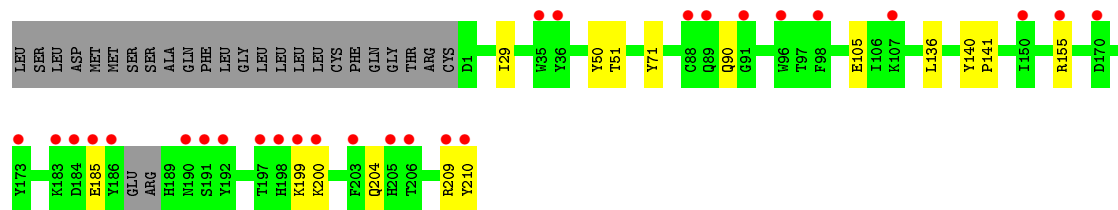
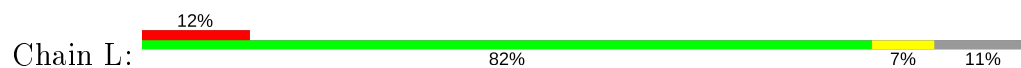
- Molecule 5 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
5	A	31	Total 31	O 31	0	0
5	B	46	Total 46	O 46	0	0
5	H	23	Total 23	O 23	0	0
5	I	36	Total 36	O 36	0	0
5	L	14	Total 14	O 14	0	0
5	M	21	Total 21	O 21	0	0

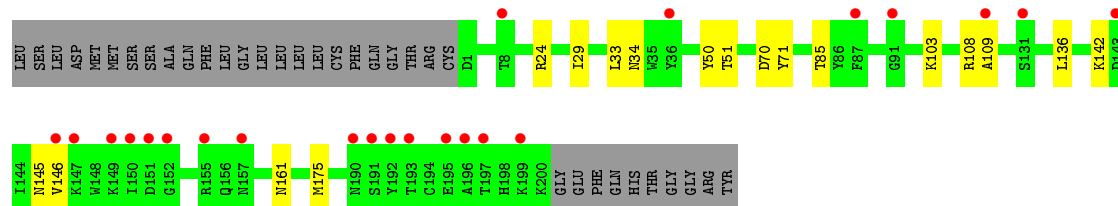
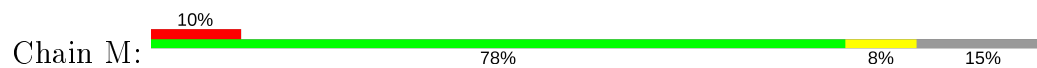




- Molecule 3: Antibody Light Chain



- Molecule 3: Antibody Light Chain





## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.95Å 53.53Å 156.79Å 90.00° 91.18° 90.00°	Depositor
Resolution (Å)	29.79 – 2.45 29.79 – 2.45	Depositor EDS
% Data completeness (in resolution range)	99.4 (29.79-2.45) 99.5 (29.79-2.45)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.48 (at 2.45Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.204 , 0.253 0.208 , 0.255	Depositor DCC
$R_{free}$ test set	1998 reflections (3.19%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	55.6	Xtrriage
Anisotropy	0.572	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 56.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.019 for h,-k,-l	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	20975	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	89.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.47% of the height of the origin peak. No significant pseudotranslation is detected.*

<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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	0.22	0/1989	0.35	0/2654
1	B	0.22	0/2209	0.34	0/2947
2	H	0.23	0/1644	0.42	0/2251
2	I	0.23	0/1692	0.42	0/2315
3	L	0.22	0/1656	0.38	0/2245
3	M	0.23	0/1593	0.41	0/2162
All	All	0.22	0/10783	0.39	0/14574

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	A	1950	1919	1926	11	0
1	B	2169	2144	2150	13	0
2	H	1599	1544	1551	8	0
2	I	1646	1589	1596	9	0
3	L	1620	1526	1533	8	0
3	M	1559	1479	1485	11	0
4	A	10	0	0	1	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	5	0	0	0	0
4	H	15	0	0	1	0
4	I	15	0	0	1	0
4	L	10	0	0	0	0
4	M	5	0	0	0	0
5	A	31	0	0	1	0
5	B	46	0	0	1	0
5	H	23	0	0	0	0
5	I	36	0	0	0	0
5	L	14	0	0	0	0
5	M	21	0	0	0	0
All	All	10774	10201	10241	58	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 3.

The worst 5 of 58 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)
3:M:108:ARG:NH1	3:M:109:ALA:O	1.98	0.97
1:A:30:ASP:OD2	1:A:33:ASN:N	2.17	0.77
1:A:194:ARG:NH2	5:A:423:HOH:O	2.24	0.70
3:L:155:ARG:NH2	3:L:185:GLU:OE2	2.33	0.62
1:B:245:VAL:HB	1:B:246:PRO:CD	2.34	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	A	220/297 (74%)	212 (96%)	8 (4%)	0	<b>100</b> <b>100</b>

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	243/297 (82%)	235 (97%)	8 (3%)	0	100	100
2	H	202/233 (87%)	197 (98%)	5 (2%)	0	100	100
2	I	210/233 (90%)	201 (96%)	9 (4%)	0	100	100
3	L	204/234 (87%)	188 (92%)	16 (8%)	0	100	100
3	M	198/234 (85%)	189 (96%)	9 (4%)	0	100	100
All	All	1277/1528 (84%)	1222 (96%)	55 (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	217/278 (78%)	217 (100%)	0	100	100
1	B	244/278 (88%)	243 (100%)	1 (0%)	91	94
2	H	181/201 (90%)	180 (99%)	1 (1%)	86	91
2	I	186/201 (92%)	185 (100%)	1 (0%)	88	93
3	L	182/205 (89%)	180 (99%)	2 (1%)	73	82
3	M	177/205 (86%)	177 (100%)	0	100	100
All	All	1187/1368 (87%)	1182 (100%)	5 (0%)	91	94

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	26	LYS
2	H	100(B)	TYR
2	I	100(B)	TYR
3	L	90	GLN
3	L	105	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

12 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	$\# Z  > 2$
4	SO4	A	302	-	4,4,4	0.14	0	6,6,6	0.05	0
4	SO4	I	301	-	4,4,4	0.14	0	6,6,6	0.05	0
4	SO4	H	301	-	4,4,4	0.14	0	6,6,6	0.06	0
4	SO4	M	301	-	4,4,4	0.14	0	6,6,6	0.07	0
4	SO4	I	302	-	4,4,4	0.14	0	6,6,6	0.09	0
4	SO4	L	301	-	4,4,4	0.14	0	6,6,6	0.05	0
4	SO4	H	302	-	4,4,4	0.14	0	6,6,6	0.04	0
4	SO4	H	303	-	4,4,4	0.13	0	6,6,6	0.12	0
4	SO4	L	302	-	4,4,4	0.14	0	6,6,6	0.06	0
4	SO4	I	303	-	4,4,4	0.13	0	6,6,6	0.11	0
4	SO4	A	301	-	4,4,4	0.13	0	6,6,6	0.06	0
4	SO4	B	301	-	4,4,4	0.14	0	6,6,6	0.05	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	302	SO4	1	0
4	H	303	SO4	1	0
4	I	303	SO4	1	0

## 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	A	230/297 (77%)	0.78	29 (12%) 3 2	51, 78, 153, 184	0
1	B	257/297 (86%)	0.83	37 (14%) 2 1	46, 72, 138, 166	0
2	H	208/233 (89%)	0.35	11 (5%) 26 23	48, 72, 112, 138	0
2	I	214/233 (91%)	0.47	15 (7%) 16 13	43, 67, 110, 136	0
3	L	208/234 (88%)	0.68	28 (13%) 3 2	51, 86, 142, 166	0
3	M	200/234 (85%)	0.68	23 (11%) 4 3	44, 78, 140, 159	0
All	All	1317/1528 (86%)	0.64	143 (10%) 5 3	43, 75, 139, 184	0

The worst 5 of 143 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	197	LEU	8.6
1	A	196	LYS	7.2
1	B	245	VAL	6.8
1	B	208	TYR	6.5
1	B	213	LYS	6.5

### 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 carbohydrates in this entry.

## 6.4 Ligands [i](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
4	SO4	I	301	5/5	0.66	0.74	199,200,201,202	0
4	SO4	L	302	5/5	0.78	0.26	145,151,152,153	0
4	SO4	A	302	5/5	0.90	0.35	117,123,125,127	0
4	SO4	A	301	5/5	0.90	0.15	113,117,123,128	0
4	SO4	H	302	5/5	0.91	0.14	136,137,138,138	0
4	SO4	I	303	5/5	0.95	0.35	76,77,103,108	0
4	SO4	B	301	5/5	0.95	0.11	123,124,128,131	0
4	SO4	I	302	5/5	0.96	0.13	80,81,84,87	0
4	SO4	L	301	5/5	0.96	0.12	76,85,95,95	0
4	SO4	M	301	5/5	0.97	0.11	62,71,82,84	0
4	SO4	H	301	5/5	0.97	0.11	64,64,76,81	0
4	SO4	H	303	5/5	0.97	0.21	76,83,92,102	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.