



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 16, 2017 – 08:07 AM EDT

PDB ID : 3ET9  
Title : Crystal structure of the engineered neutralizing antibody 1H  
Authors : Leysath, C.E.; Monzingo, A.F.; Barnett, J.; Iverson, B.L.; Georgiou, G.; Robertus, J.D.  
Deposited on : unknown  
Resolution : 2.80 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

<http://wwpdb.org/validation/2016/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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

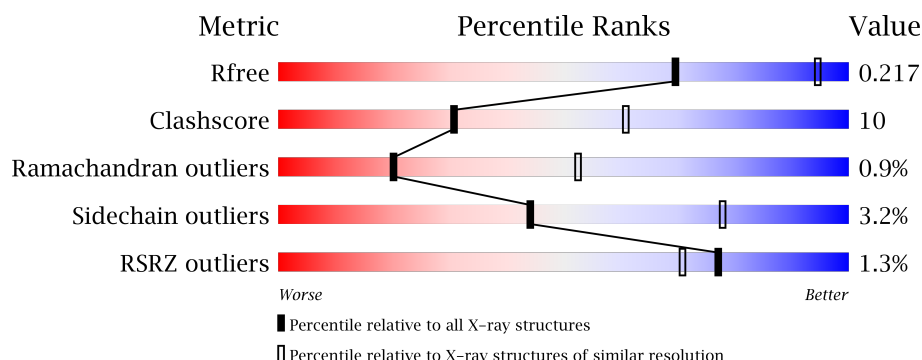
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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))
$R_{free}$	100719	2583 (2.80-2.80)
Clashscore	112137	3033 (2.80-2.80)
Ramachandran outliers	110173	2983 (2.80-2.80)
Sidechain outliers	110143	2985 (2.80-2.80)
RSRZ outliers	101464	2610 (2.80-2.80)

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. 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	F	252	<div> <div style="display: flex; align-items: center;"> <div style="width: 100%; height: 10px; position: relative;"> <div style="position: absolute; top: -5px; left: 0; right: 0; height: 1px; background-color: red;"></div> <div style="position: absolute; bottom: -5px; left: 0; right: 0; height: 1px; background-color: green;"></div> <div style="position: absolute; top: 0; left: 0; width: 100%; height: 100%; background: linear-gradient(to right, red 0%, orange 10%, yellow 22%, green 67%, grey 100%);"></div> </div> <div style="margin-left: 5px;"> <div style="font-size: 0.8em;">%</div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>67%</span> <span>22%</span> <span>• 10%</span> </div> </div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1740 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 Antibody 1H light chain and antibody 1H heavy chain linked with a synthetic (GGGGS)<sub>4</sub> linker.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	F	227	Total	C	N	O	S	0	0	1
			1724	1084	291	342	7			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-4	MET	-	EXPRESSION TAG	PDB 3ET9
F	-3	ALA	-	EXPRESSION TAG	PDB 3ET9
F	-2	ASP	-	EXPRESSION TAG	PDB 3ET9
F	-1	TYR	-	EXPRESSION TAG	PDB 3ET9
F	0	LYS	-	EXPRESSION TAG	PDB 3ET9
F	55	LEU	GLN	ENGINEERED	PDB 3ET9
F	56	PRO	SER	ENGINEERED	PDB 3ET9
F	107	ARG	LYS	ENGINEERED	PDB 3ET9
F	109	GLY	-	LINKER	PDB 3ET9
F	110	GLY	-	LINKER	PDB 3ET9
F	111	GLY	-	LINKER	PDB 3ET9
F	112	GLY	-	LINKER	PDB 3ET9
F	113	SER	-	LINKER	PDB 3ET9
F	114	GLY	-	LINKER	PDB 3ET9
F	115	GLY	-	LINKER	PDB 3ET9
F	116	GLY	-	LINKER	PDB 3ET9
F	117	GLY	-	LINKER	PDB 3ET9
F	118	SER	-	LINKER	PDB 3ET9
F	119	GLY	-	LINKER	PDB 3ET9
F	120	GLY	-	LINKER	PDB 3ET9
F	121	GLY	-	LINKER	PDB 3ET9
F	122	GLY	-	LINKER	PDB 3ET9
F	123	SER	-	LINKER	PDB 3ET9
F	124	GLY	-	LINKER	PDB 3ET9
F	125	GLY	-	LINKER	PDB 3ET9
F	126	GLY	-	LINKER	PDB 3ET9

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	127	GLY	-	LINKER	PDB 3ET9
F	128	SER	-	LINKER	PDB 3ET9

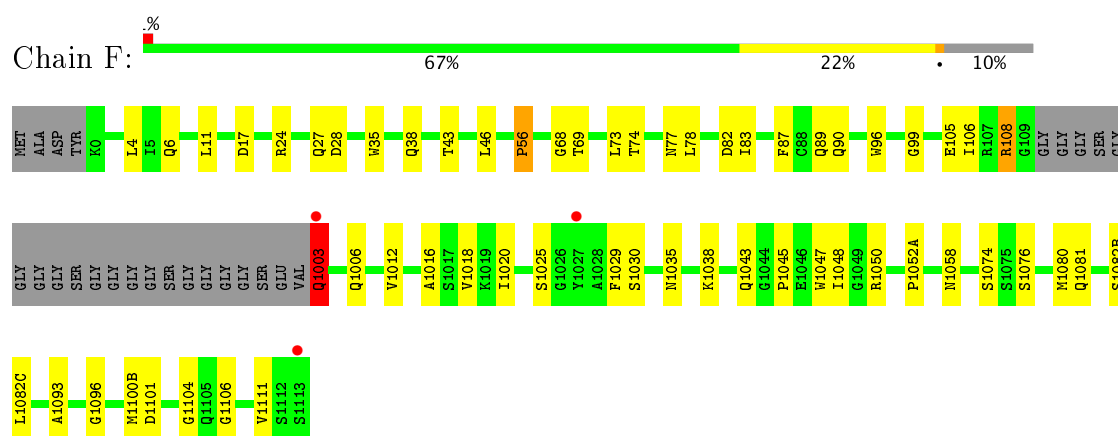
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	F	16	Total O 16 16	0	0

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

These plots are drawn for all protein, RNA and DNA 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: Antibody 1H light chain and antibody 1H heavy chain linked with a synthetic (GGGS)<sub>4</sub> linker



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.94Å 79.94Å 68.17Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.80 39.97 – 2.80	Depositor EDS
% Data completeness (in resolution range)	93.7 (20.00-2.80) 98.2 (39.97-2.80)	Depositor EDS
$R_{merge}$	0.16	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.29 (at 2.81Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.217 , 0.255 0.224 , 0.217	Depositor DCC
$R_{free}$ test set	258 reflections (4.74%)	DCC
Wilson B-factor (Å <sup>2</sup> )	21.7	Xtriage
Anisotropy	0.161	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 36.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	1740	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	13.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.13% 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](#)

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	F	0.45	0/1762	1.07	6/2389 (0.3%)

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
1	F	0	1

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	F	1003	GLN	O-C-N	-20.71	89.56	122.70
1	F	1003	GLN	CA-C-N	20.66	162.66	117.20
1	F	1003	GLN	CA-C-O	-18.05	82.20	120.10
1	F	56	PRO	CA-N-CD	-13.89	92.05	111.50
1	F	1003	GLN	C-N-CA	-5.51	107.93	121.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	F	1003	GLN	Mainchain

### 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	F	1724	0	1657	35	0
2	F	16	0	0	0	0
All	All	1740	0	1657	35	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 10.

The worst 5 of 35 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:F:1050:ARG:HE	1:F:1058:ASN:HD22	1.08	0.97
1:F:83:ILE:HD11	1:F:106:ILE:HG13	1.58	0.86
1:F:1003:GLN:HA	1:F:1025:SER:H	1.49	0.78
1:F:6:GLN:HE21	1:F:99:GLY:HA3	1.47	0.76
1:F:1012:VAL:HG21	1:F:1082(C):LEU:HD12	1.72	0.71

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	F	223/252 (88%)	209 (94%)	12 (5%)	2 (1%)	20 52

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	108	ARG
1	F	1096	GLY



### 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	F	187/201 (93%)	181 (97%)	6 (3%)	44 78

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	56	PRO
1	F	1082(B)	SER
1	F	1043	GLN
1	F	24	ARG
1	F	1074	SER

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

Mol	Chain	Res	Type
1	F	77	ASN
1	F	1058	ASN
1	F	80	GLN
1	F	31	ASN
1	F	1006	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 carbohydrates 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	F	227/252 (90%)	-0.23	3 (1%) 77 71	5, 13, 25, 33	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	1027	TYR	2.7
1	F	1003	GLN	2.1
1	F	1113	SER	2.1

### 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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.