



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 06:46 pm GMT

PDB ID : 5F8Q  
Title : Blood group antigen binding adhesin BabA of Helicobacter pylori strain S831  
in complex with Nanobody Nb-ER19  
Authors : Moonens, K.; Gideonsson, P.; Subedi, S.; Romao, E.; Oscarson, S.; Muylder-  
mans, S.; Boren, T.; Remaut, H.  
Deposited on : 2015-12-09  
Resolution : 2.59 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	:	trunk28620
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)	:	recalc28949

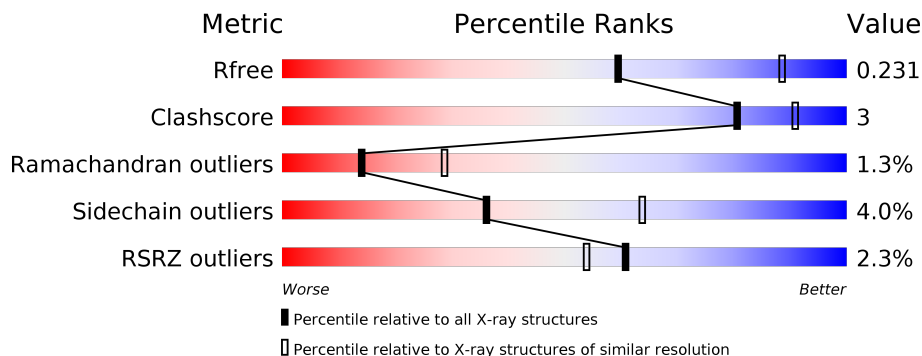
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.59 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	2542 (2.60-2.60)
Clashscore	112137	2895 (2.60-2.60)
Ramachandran outliers	110173	2848 (2.60-2.60)
Sidechain outliers	110143	2848 (2.60-2.60)
RSRZ outliers	101464	2550 (2.60-2.60)

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	A	466	<div> <div>3%</div> <div> <div></div> <div>82%</div> <div>7%</div> <div>10%</div> </div> </div>
1	B	466	<div> <div>3%</div> <div> <div></div> <div>80%</div> <div>8%</div> <div>10%</div> </div> </div>
2	C	120	<div> <div></div> <div> <div></div> <div>87%</div> <div>6%</div> <div>6%</div> </div> </div>
2	D	120	<div> <div></div> <div> <div></div> <div>92%</div> <div></div> <div>5%</div> </div> </div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 8024 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 Adhesin binding fucosylated histo-blood group antigen,Adhesin,Adhesin binding fucosylated histo-blood group antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	419	Total	C	N	O	S	0	0	0
			3118	1925	537	643	13			
1	B	419	Total	C	N	O	S	0	0	0
			3118	1925	537	643	13			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	3	ALA	-	expression tag	UNP O52269
A	4	SER	-	expression tag	UNP O52269
A	5	TRP	-	expression tag	UNP O52269
A	6	SER	-	expression tag	UNP O52269
A	7	HIS	-	expression tag	UNP O52269
A	8	PRO	-	expression tag	UNP O52269
A	9	GLN	-	expression tag	UNP O52269
A	10	PHE	-	expression tag	UNP O52269
A	11	GLU	-	expression tag	UNP O52269
A	12	LYS	-	expression tag	UNP O52269
A	13	SER	-	expression tag	UNP O52269
A	14	GLY	-	expression tag	UNP O52269
A	15	GLY	-	expression tag	UNP O52269
A	16	GLY	-	expression tag	UNP O52269
A	17	GLY	-	expression tag	UNP O52269
A	18	GLY	-	expression tag	UNP O52269
A	19	LEU	-	expression tag	UNP O52269
A	20	VAL	-	expression tag	UNP O52269
A	21	PRO	-	expression tag	UNP O52269
A	22	ARG	-	expression tag	UNP O52269
A	23	GLY	-	expression tag	UNP O52269
A	24	SER	-	expression tag	UNP O52269
A	461	GLY	-	expression tag	UNP O52269
A	462	SER	-	expression tag	UNP O52269

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	463	HIS	-	expression tag	UNP O52269
A	464	HIS	-	expression tag	UNP O52269
A	465	HIS	-	expression tag	UNP O52269
A	466	HIS	-	expression tag	UNP O52269
A	467	HIS	-	expression tag	UNP O52269
A	468	HIS	-	expression tag	UNP O52269
B	3	ALA	-	expression tag	UNP O52269
B	4	SER	-	expression tag	UNP O52269
B	5	TRP	-	expression tag	UNP O52269
B	6	SER	-	expression tag	UNP O52269
B	7	HIS	-	expression tag	UNP O52269
B	8	PRO	-	expression tag	UNP O52269
B	9	GLN	-	expression tag	UNP O52269
B	10	PHE	-	expression tag	UNP O52269
B	11	GLU	-	expression tag	UNP O52269
B	12	LYS	-	expression tag	UNP O52269
B	13	SER	-	expression tag	UNP O52269
B	14	GLY	-	expression tag	UNP O52269
B	15	GLY	-	expression tag	UNP O52269
B	16	GLY	-	expression tag	UNP O52269
B	17	GLY	-	expression tag	UNP O52269
B	18	GLY	-	expression tag	UNP O52269
B	19	LEU	-	expression tag	UNP O52269
B	20	VAL	-	expression tag	UNP O52269
B	21	PRO	-	expression tag	UNP O52269
B	22	ARG	-	expression tag	UNP O52269
B	23	GLY	-	expression tag	UNP O52269
B	24	SER	-	expression tag	UNP O52269
B	461	GLY	-	expression tag	UNP O52269
B	462	SER	-	expression tag	UNP O52269
B	463	HIS	-	expression tag	UNP O52269
B	464	HIS	-	expression tag	UNP O52269
B	465	HIS	-	expression tag	UNP O52269
B	466	HIS	-	expression tag	UNP O52269
B	467	HIS	-	expression tag	UNP O52269
B	468	HIS	-	expression tag	UNP O52269

- Molecule 2 is a protein called Nanobody Nb-ER19.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	113	Total	C	N	O	S	0	0	0
			863	540	155	163	5			

*Continued on next page...*

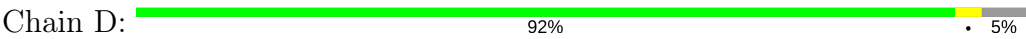
*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	114	Total	C	N	O	S	0	0	0
			873	546	158	164	5			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	21	Total	O	0	0
			21	21		
3	C	4	Total	O	0	0
			4	4		
3	B	17	Total	O	0	0
			17	17		
3	D	10	Total	O	0	0
			10	10		





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	51.03Å 132.14Å 123.48Å 90.00° 94.90° 90.00°	Depositor
Resolution (Å)	48.50 – 2.59 47.46 – 2.59	Depositor EDS
% Data completeness (in resolution range)	97.5 (48.50-2.59) 97.7 (47.46-2.59)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.18 (at 2.58Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.195 , 0.230 0.198 , 0.231	Depositor DCC
$R_{free}$ test set	2487 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	65.9	Xtriage
Anisotropy	0.153	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 33.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	8024	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.71% 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

### 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	A	0.71	0/3165	0.76	0/4305
1	B	0.64	0/3165	0.76	1/4305 (0.0%)
2	C	0.66	0/882	0.83	2/1195 (0.2%)
2	D	0.71	0/893	0.82	1/1210 (0.1%)
All	All	0.68	0/8105	0.78	4/11015 (0.0%)

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	B	0	1
2	D	0	1
All	All	0	2

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	198	LEU	CA-CB-CG	9.12	136.27	115.30
2	C	72	ARG	NE-CZ-NH1	7.26	123.93	120.30
2	C	72	ARG	NE-CZ-NH2	-5.10	117.75	120.30
2	D	72	ARG	NE-CZ-NH1	5.09	122.85	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	286	SER	Peptide
2	D	3	VAL	Peptide

## 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	3118	0	3049	20	0
1	B	3118	0	3049	20	0
2	C	863	0	845	9	0
2	D	873	0	852	0	0
3	A	21	0	0	1	0
3	B	17	0	0	1	0
3	C	4	0	0	0	0
3	D	10	0	0	0	0
All	All	8024	0	7795	49	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 49 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:284:ASN:HB2	3:A:504:HOH:O	1.56	1.04
2:C:3:VAL:HG13	2:C:102:TYR:CZ	2.14	0.83
2:C:3:VAL:HG11	2:C:102:TYR:OH	1.81	0.80
2:C:3:VAL:N	2:C:26:SER:HG	1.80	0.80
2:C:3:VAL:HG13	2:C:102:TYR:CE2	2.19	0.77

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	415/466 (89%)	391 (94%)	18 (4%)	6 (1%)	13	26
1	B	415/466 (89%)	393 (95%)	14 (3%)	8 (2%)	9	18
2	C	111/120 (92%)	108 (97%)	3 (3%)	0	100	100
2	D	112/120 (93%)	110 (98%)	2 (2%)	0	100	100
All	All	1053/1172 (90%)	1002 (95%)	37 (4%)	14 (1%)	14	29

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	237	LYS
1	A	295	ARG
1	A	288	GLY
1	B	219	ASP
1	B	242	ARG

### 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	A	346/382 (91%)	331 (96%)	15 (4%)	33	61
1	B	346/382 (91%)	331 (96%)	15 (4%)	33	61
2	C	92/99 (93%)	89 (97%)	3 (3%)	43	70
2	D	93/99 (94%)	91 (98%)	2 (2%)	57	81
All	All	877/962 (91%)	842 (96%)	35 (4%)	36	64

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	28	SER
1	B	145	TYR
1	B	462	SER
2	C	40	GLN
2	C	115	SER

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

Mol	Chain	Res	Type
1	A	352	GLN

### 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](#)

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	A	419/466 (89%)	0.12	12 (2%) 52 45	40, 61, 109, 135	0
1	B	419/466 (89%)	0.24	13 (3%) 49 41	45, 68, 120, 153	0
2	C	113/120 (94%)	-0.01	0 100 100	49, 63, 84, 102	0
2	D	114/120 (95%)	0.05	0 100 100	45, 55, 74, 99	0
All	All	1065/1172 (90%)	0.15	25 (2%) 61 54	40, 63, 110, 153	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	219	ASP	3.7
1	B	232	VAL	3.5
1	B	241	THR	3.2
1	B	243	VAL	3.2
1	B	209	GLY	3.2

### 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 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.