



# Full wwPDB X-ray Structure Validation Report ⓘ

Apr 5, 2016 – 09:13 AM EDT

PDB ID : 4Z3J  
Title : Crystal structure of the lectin domain of PapG from E. coli BI47 in space group P1  
Authors : Jakob, R.P.; Navarra, G.; Zihlmann, P.; Stangier, K.; Preston, R.C.; Rabbani, S.; Maier, T.; Ernst, B.  
Deposited on : 2015-03-31  
Resolution : 2.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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  
Mogul : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027107  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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-20027107

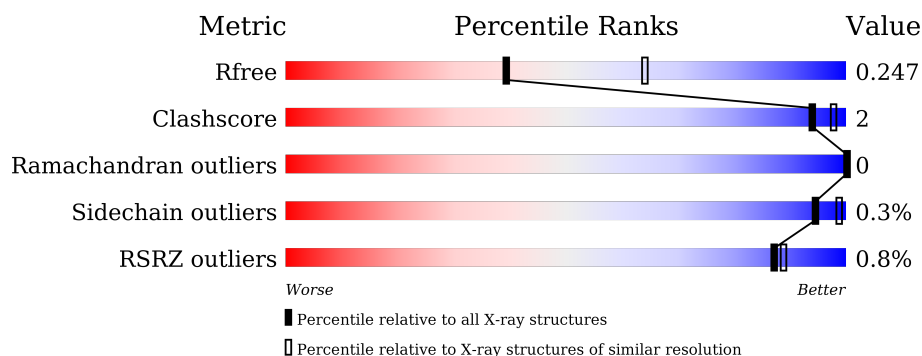
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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))
$R_{free}$	91344	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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 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	198	<div> <div style="width: 100%; height: 10px; background-color: red; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">%</span> </div> <div style="width: 95%; height: 10px; background-color: green; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">95%</span> </div> <div style="width: 5%; height: 10px; background-color: yellow; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">5%</span> </div> <div style="width: 5%; height: 10px; background-color: grey; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">5%</span> </div> </div>
1	B	198	<div> <div style="width: 100%; height: 10px; background-color: red; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">%</span> </div> <div style="width: 94%; height: 10px; background-color: green; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">94%</span> </div> <div style="width: 5%; height: 10px; background-color: yellow; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">5%</span> </div> <div style="width: 1%; height: 10px; background-color: grey; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">1%</span> </div> </div>
1	C	198	<div> <div style="width: 100%; height: 10px; background-color: red; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">%</span> </div> <div style="width: 95%; height: 10px; background-color: green; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">95%</span> </div> <div style="width: 5%; height: 10px; background-color: yellow; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">5%</span> </div> <div style="width: 5%; height: 10px; background-color: grey; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">5%</span> </div> </div>
1	D	198	<div> <div style="width: 100%; height: 10px; background-color: red; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">%</span> </div> <div style="width: 94%; height: 10px; background-color: green; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">94%</span> </div> <div style="width: 5%; height: 10px; background-color: yellow; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">5%</span> </div> <div style="width: 1%; height: 10px; background-color: grey; position: relative;"> <span style="position: absolute; right: 0; top: -10px;">1%</span> </div> </div>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 12732 atoms, of which 6119 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 PapG, lectin domain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	196	Total	C	H	N	O	S	0	0	0
			3131	1042	1531	264	290	4			
1	B	196	Total	C	H	N	O	S	0	0	0
			3131	1042	1531	264	290	4			
1	C	196	Total	C	H	N	O	S	0	0	0
			3131	1042	1531	264	290	4			
1	D	195	Total	C	H	N	O	S	0	0	0
			3121	1039	1526	263	289	4			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	Expression tag	UNP T7DCJ2
A	109	GLN	GLU	conflict	UNP T7DCJ2
B	-1	MET	-	Expression tag	UNP T7DCJ2
B	109	GLN	GLU	conflict	UNP T7DCJ2
C	-1	MET	-	Expression tag	UNP T7DCJ2
C	109	GLN	GLU	conflict	UNP T7DCJ2
D	-1	MET	-	Expression tag	UNP T7DCJ2
D	109	GLN	GLU	conflict	UNP T7DCJ2

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	52	Total	O	0	0
			52	52		
2	B	54	Total	O	0	0
			54	54		
2	C	52	Total	O	0	0
			52	52		
2	D	60	Total	O	0	0
			60	60		

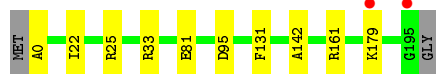
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

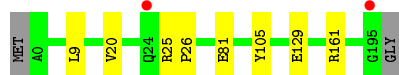
- Molecule 1: PapG, lectin domain



- Molecule 1: PapG, lectin domain



- Molecule 1: PapG, lectin domain



- Molecule 1: PapG, lectin domain



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	50.76 Å 56.56 Å 70.94 Å 112.76° 102.76° 88.12°	Depositor
Resolution (Å)	52.07 – 2.50 52.07 – 2.50	Depositor EDS
% Data completeness (in resolution range)	95.2 (52.07-2.50) 82.3 (52.07-2.50)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.71 (at 2.51 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.219 , 0.246 0.222 , 0.247	Depositor DCC
$R_{free}$ test set	1166 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	22.9	Xtriage
Anisotropy	0.666	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 1.6	EDS
Estimated twinning fraction	0.156 for -h,k,-k-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 23346 reflections	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	12732	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 48.83 % 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 8.1385e-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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.25	0/1652	0.46	0/2249
1	B	0.25	0/1652	0.44	0/2249
1	C	0.26	0/1652	0.48	0/2249
1	D	0.24	0/1647	0.44	0/2242
All	All	0.25	0/6603	0.45	0/8989

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	A	1600	1531	1535	4	0
1	B	1600	1531	1535	7	0
1	C	1600	1531	1535	5	0
1	D	1595	1526	1530	7	0
2	A	52	0	0	2	0
2	B	54	0	0	4	0
2	C	52	0	0	2	0
2	D	60	0	0	5	0
All	All	6613	6119	6135	23	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.

All (23) 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:120:LYS:NZ	2:A:201:HOH:O	2.06	0.88
1:C:25:ARG:NH1	1:C:81:GLU:OE1	2.09	0.85
1:D:13:ASN:ND2	2:D:201:HOH:O	2.10	0.84
1:B:33:ARG:NH1	2:B:202:HOH:O	2.26	0.69
1:B:25:ARG:NH1	1:B:81:GLU:OE2	2.31	0.63
1:D:25:ARG:NH1	2:D:206:HOH:O	2.33	0.61
1:B:0:ALA:N	2:B:207:HOH:O	2.36	0.57
1:B:95:ASP:OD1	1:B:179:LYS:NZ	2.39	0.55
1:A:157:SER:OG	1:A:185:ASN:ND2	2.40	0.54
1:C:9:LEU:O	2:C:201:HOH:O	2.19	0.53
1:D:25:ARG:NH1	2:D:214:HOH:O	2.42	0.52
1:B:131:PHE:O	2:B:201:HOH:O	2.19	0.52
1:D:20:VAL:HG11	1:D:26:PRO:HB3	1.94	0.49
1:A:83:HIS:ND1	1:A:137:LYS:NZ	2.64	0.46
1:A:33:ARG:NH2	2:A:214:HOH:O	2.49	0.45
1:C:161:ARG:NE	2:C:208:HOH:O	2.47	0.45
1:D:22:ILE:HG21	1:D:142:ALA:HA	1.99	0.45
1:D:120:LYS:NZ	2:D:210:HOH:O	2.37	0.45
1:B:161:ARG:NH2	2:B:214:HOH:O	2.50	0.44
1:B:22:ILE:HG21	1:B:142:ALA:HA	2.00	0.43
1:C:20:VAL:HG11	1:C:26:PRO:HB3	2.01	0.42
1:C:105:TYR:CE2	1:C:129:GLU:HG2	2.55	0.42
1:D:73:LYS:NZ	2:D:211:HOH:O	2.37	0.40

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	194/198 (98%)	188 (97%)	6 (3%)	0	<b>100</b> <b>100</b>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	194/198 (98%)	190 (98%)	4 (2%)	0	100	100
1	C	194/198 (98%)	189 (97%)	5 (3%)	0	100	100
1	D	193/198 (98%)	190 (98%)	3 (2%)	0	100	100
All	All	775/792 (98%)	757 (98%)	18 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	169/170 (99%)	168 (99%)	1 (1%)	90	97
1	B	169/170 (99%)	169 (100%)	0	100	100
1	C	169/170 (99%)	169 (100%)	0	100	100
1	D	169/170 (99%)	168 (99%)	1 (1%)	90	97
All	All	676/680 (99%)	674 (100%)	2 (0%)	94	99

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

Mol	Chain	Res	Type
1	A	126	ARG
1	D	14	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	185	ASN

### 5.3.3 RNA ⓘ

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	196/198 (98%)	-0.32	1 (0%) 91 92	14, 28, 44, 56	0
1	B	196/198 (98%)	-0.34	2 (1%) 84 86	17, 26, 46, 62	0
1	C	196/198 (98%)	-0.34	2 (1%) 84 86	15, 24, 49, 67	0
1	D	195/198 (98%)	-0.32	1 (0%) 91 92	16, 27, 53, 67	0
All	All	783/792 (98%)	-0.33	6 (0%) 87 89	14, 27, 47, 67	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	179	LYS	3.0
1	B	195	GLY	2.8
1	D	195	GLY	2.5
1	C	24	GLN	2.2
1	A	176	ASN	2.1
1	C	195	GLY	2.0

### 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

There are no such residues in this entry.