



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:45 PM GMT

PDB ID : 4PUP  
Title : 2.75 Angstrom resolution crystal structure of uncharacterized protein from Burkholderia cenocepacia J2315  
Authors : Halavaty, A.S.; Filippova, E.V.; Wawrzak, Z.; Kiryukhina, O.; Minasov, G.; Jedrzejczak, R.; Shuvalova, L.; Joachimiak, A.; Anderson, W.F.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : 2014-03-13  
Resolution : 2.75 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

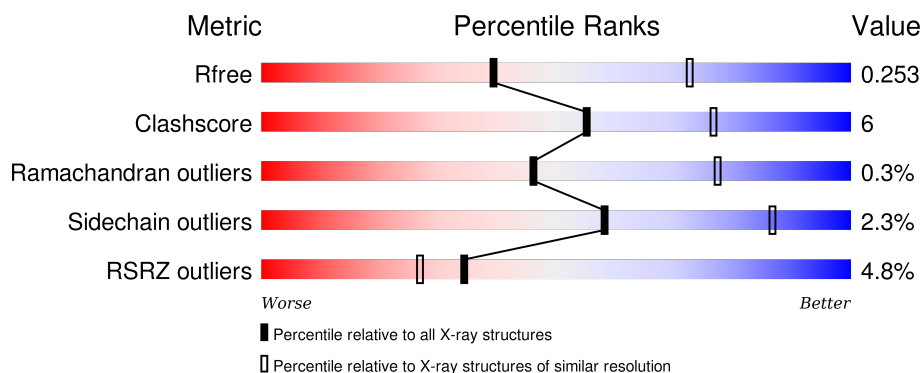
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.75 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	3340 (2.80-2.72)
Clashscore	102246	3829 (2.80-2.72)
Ramachandran outliers	100387	3767 (2.80-2.72)
Sidechain outliers	100360	3770 (2.80-2.72)
RSRZ outliers	91569	3352 (2.80-2.72)

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	131	<div> <div>2%</div> <div>74%</div> <div>9%</div> <div>•</div> <div>15%</div> </div>
1	B	131	<div> <div>2%</div> <div>76%</div> <div>9%</div> <div>•</div> <div>14%</div> </div>
1	C	131	<div> <div>8%</div> <div>73%</div> <div>14%</div> <div>•</div> <div>12%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2580 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 Uncharacterized protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	112	Total	C	N	O	S	Se	0	0	0
			830	520	146	160	1	3			
1	B	113	Total	C	N	O	S	Se	0	0	0
			840	524	152	160	1	3			
1	C	115	Total	C	N	O	S	Se	0	0	0
			854	534	152	163	2	3			

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	MSE	-	EXPRESSION TAG	UNP B4EGA9
A	-5	HIS	-	EXPRESSION TAG	UNP B4EGA9
A	-4	HIS	-	EXPRESSION TAG	UNP B4EGA9
A	-3	HIS	-	EXPRESSION TAG	UNP B4EGA9
A	-2	HIS	-	EXPRESSION TAG	UNP B4EGA9
A	-1	HIS	-	EXPRESSION TAG	UNP B4EGA9
A	0	HIS	-	EXPRESSION TAG	UNP B4EGA9
B	-6	MSE	-	EXPRESSION TAG	UNP B4EGA9
B	-5	HIS	-	EXPRESSION TAG	UNP B4EGA9
B	-4	HIS	-	EXPRESSION TAG	UNP B4EGA9
B	-3	HIS	-	EXPRESSION TAG	UNP B4EGA9
B	-2	HIS	-	EXPRESSION TAG	UNP B4EGA9
B	-1	HIS	-	EXPRESSION TAG	UNP B4EGA9
B	0	HIS	-	EXPRESSION TAG	UNP B4EGA9
C	-6	MSE	-	EXPRESSION TAG	UNP B4EGA9
C	-5	HIS	-	EXPRESSION TAG	UNP B4EGA9
C	-4	HIS	-	EXPRESSION TAG	UNP B4EGA9
C	-3	HIS	-	EXPRESSION TAG	UNP B4EGA9
C	-2	HIS	-	EXPRESSION TAG	UNP B4EGA9
C	-1	HIS	-	EXPRESSION TAG	UNP B4EGA9
C	0	HIS	-	EXPRESSION TAG	UNP B4EGA9

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	24	Total 24	O 24	0	0
2	B	16	Total 16	O 16	0	0
2	C	16	Total 16	O 16	0	0



- Molecule 1: Uncharacterized protein



Amino Acid Category	Number of Proteins
MSE	10
HIS	10
HIS	10
HIS	10
HIS	10
H+1	10
HO	10
M1	10
H21	10
D27	10
R35	10
G39	10
V40	10
P41	10
T42	10
W50	10
L60	10
D68	10
H79	10
ARG	10
ALA	10
VAL	10
PRO	10
ALA	10
ALA	10
GLY	10
GLU	10
ARG	10
CYS	10
THR	10
PHE	10
N92	10
L98	10
L106	10
R122	10
P123	10
ARG	10

MSE	HIS	HIS	HIS	HIS	HIS	HIS	H0	M1	T7	G8	H21	R35	E36	P37	G38	G39	V40	P41	W50	E58	R59	L60	D68	G69	H79	ARG	ALA	VAL	PRO	ALA	ALA	GLY	GLU	GLU	C89	T90	F91	N92	T97	L98	L99	I100	L106	D114	D119	R122	P123	N130
-----	-----	-----	-----	-----	-----	-----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.78Å 88.05Å 79.38Å 90.00° 108.34° 90.00°	Depositor
Resolution (Å)	28.64 – 2.75 28.64 – 2.75	Depositor EDS
% Data completeness (in resolution range)	98.6 (28.64-2.75) 98.7 (28.64-2.75)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.09 (at 2.76Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.224 , 0.255 0.232 , 0.253	Depositor DCC
$R_{free}$ test set	526 reflections (5.07%)	DCC
Wilson B-factor (Å <sup>2</sup> )	48.7	Xtriage
Anisotropy	0.351	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 23.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.52$ , $\langle L^2 \rangle = 0.36$	Xtriage
Outliers	0 of 10905 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	2580	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.38% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	A	0.67	0/849	0.99	7/1154 (0.6%)
1	B	0.65	0/860	0.90	1/1169 (0.1%)
1	C	0.68	0/874	0.88	1/1188 (0.1%)
All	All	0.67	0/2583	0.92	9/3511 (0.3%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	92	ASN	N-CA-C	6.51	128.58	111.00
1	A	1	MSE	N-CA-CB	6.50	122.30	110.60
1	A	119	ASP	CB-CG-OD2	-6.20	112.72	118.30
1	A	119	ASP	N-CA-C	-5.95	94.94	111.00
1	A	1	MSE	CB-CA-C	-5.89	98.62	110.40

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	830	0	786	9	0
1	B	840	0	797	11	0
1	C	854	0	811	13	0
2	A	24	0	0	1	0
2	B	16	0	0	4	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	16	0	0	2	0
All	All	2580	0	2394	31	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 6.

The worst 5 of 31 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:B:68:ASP:HB2	2:B:213:HOH:O	1.83	0.78
1:A:78:LEU:O	1:A:90:THR:HA	1.89	0.72
1:C:68:ASP:HB2	2:C:216:HOH:O	1.98	0.63
1:B:27:ASP:HB2	1:B:122:ARG:HH22	1.63	0.63
1:A:93:GLN:HG2	1:A:104:ARG:HB3	1.82	0.61

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	108/131 (82%)	104 (96%)	4 (4%)	0	100	100
1	B	109/131 (83%)	106 (97%)	2 (2%)	1 (1%)	21	52
1	C	111/131 (85%)	107 (96%)	4 (4%)	0	100	100
All	All	328/393 (84%)	317 (97%)	10 (3%)	1 (0%)	46	77

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	68	ASP



### 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	84/95 (88%)	81 (96%)	3 (4%)	42	75
1	B	85/95 (90%)	84 (99%)	1 (1%)	78	94
1	C	87/95 (92%)	85 (98%)	2 (2%)	58	87
All	All	256/285 (90%)	250 (98%)	6 (2%)	58	87

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

Mol	Chain	Res	Type
1	A	119	ASP
1	C	97	THR
1	B	1	MSE
1	A	2	GLN
1	C	1	MSE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	2	GLN
1	B	2	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

There are no ligands in this entry.

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	109/131 (83%)	-0.12	3 (2%) 56 50	30, 55, 92, 114	0
1	B	110/131 (83%)	0.02	3 (2%) 58 51	32, 50, 88, 100	0
1	C	112/131 (85%)	0.07	10 (8%) 12 8	33, 51, 96, 134	0
All	All	331/393 (84%)	-0.01	16 (4%) 34 27	30, 52, 92, 134	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	89	CYS	4.8
1	C	90	THR	4.5
1	C	38	GLY	4.0
1	C	91	PHE	3.9
1	A	91	PHE	3.5

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.