



# Full wwPDB X-ray Structure Validation Report ⓘ

May 15, 2020 – 12:59 am BST

PDB ID : 2FG1  
Title : Structure of a Protein of Unknown Function from *Bacteroides thetaiotaomicron*.  
Authors : Cuff, M.E.; Bigelow, L.; Abdullah, J.; Collart, F.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : 2005-12-20  
Resolution : 1.25 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

<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

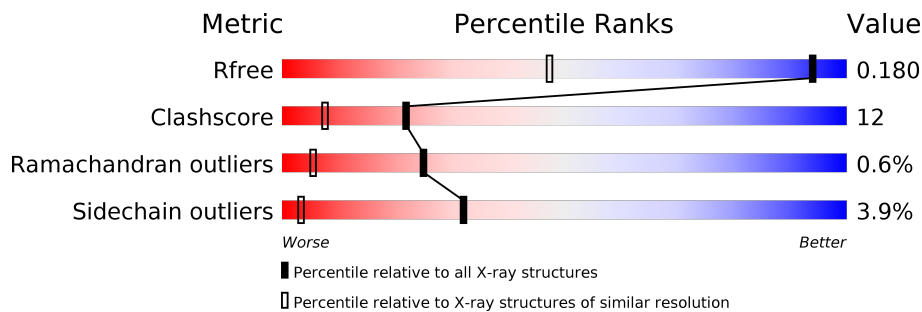
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.25 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	1023 (1.28-1.24)
Clashscore	141614	1060 (1.28-1.24)
Ramachandran outliers	138981	1029 (1.28-1.24)
Sidechain outliers	138945	1028 (1.28-1.24)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	158	 85% 13% ..

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1704 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 conserved hypothetical protein BT1257.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	N	O	S				Se
1	A	157	1432	930	240	251	4	7	0	25	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	CLONING ARTIFACT	UNP Q8A8B0
A	-1	ASN	-	CLONING ARTIFACT	UNP Q8A8B0
A	0	ALA	-	CLONING ARTIFACT	UNP Q8A8B0
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A8B0
A	45	MSE	MET	MODIFIED RESIDUE	UNP Q8A8B0
A	79	MSE	MET	MODIFIED RESIDUE	UNP Q8A8B0
A	120	MSE	MET	MODIFIED RESIDUE	UNP Q8A8B0
A	135	MSE	MET	MODIFIED RESIDUE	UNP Q8A8B0


- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	272	Total	O	0	0
			272	272		

### 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. 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. 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: conserved hypothetical protein BT1257

Chain A:  85% 13% ..



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	62.17Å 62.17Å 36.51Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 1.25 31.48 – 1.22	Depositor EDS
% Data completeness (in resolution range)	99.4 (50.00-1.25) 98.3 (31.48-1.22)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.21 (at 1.22Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.168 , 0.181 0.167 , 0.180	Depositor DCC
$R_{free}$ test set	2038 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	12.3	Xtrriage
Anisotropy	0.476	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 33.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.036 for h,-k,-l	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	1704	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.80% 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	0.41	0/1474	0.61	0/1969

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	1432	0	1501	35	0
2	A	272	0	0	3	1
All	All	1704	0	1501	35	1

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 12.

All (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:A:131[B]:LYS:HD2	1:A:134[B]:LEU:CD1	1.68	1.23
1:A:45[A]:MSE:HG3	1:A:68[A]:VAL:HG21	1.23	1.10
1:A:131[B]:LYS:HD2	1:A:134[B]:LEU:HD13	1.28	1.08
1:A:1[A]:MSE:HA	1:A:1[A]:MSE:HE3	1.43	1.00

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:45[A]:MSE:CG	1:A:68[A]:VAL:HG21	1.99	0.91
1:A:131[B]:LYS:CD	1:A:134[B]:LEU:CD1	2.49	0.91
1:A:131[B]:LYS:CD	1:A:134[B]:LEU:HD13	2.00	0.90
1:A:131[B]:LYS:HD2	1:A:134[B]:LEU:HD12	1.55	0.88
1:A:45[A]:MSE:HG3	1:A:68[A]:VAL:CG2	2.06	0.85
1:A:1[A]:MSE:CE	1:A:1[A]:MSE:HA	2.08	0.84
1:A:45[A]:MSE:CG	1:A:68[A]:VAL:CG2	2.63	0.77
1:A:24:ILE:HG13	1:A:123[B]:ILE:HG12	1.75	0.67
1:A:131[B]:LYS:CE	1:A:134[B]:LEU:CD1	2.74	0.66
1:A:1[A]:MSE:CE	1:A:1[A]:MSE:CA	2.78	0.57
1:A:131[B]:LYS:CE	1:A:134[B]:LEU:HD13	2.36	0.55
1:A:45[B]:MSE:HG3	2:A:179:HOH:O	2.08	0.54
1:A:24:ILE:HD11	1:A:123[B]:ILE:HG23	1.89	0.53
1:A:105:LYS:HE3	2:A:410:HOH:O	2.10	0.52
1:A:45[A]:MSE:HG2	1:A:68[A]:VAL:CG2	2.40	0.51
1:A:131[B]:LYS:CE	1:A:134[B]:LEU:HD11	2.41	0.51
1:A:23:HIS:HA	1:A:123[B]:ILE:HD11	1.94	0.50
1:A:19[B]:LYS:NZ	2:A:232:HOH:O	2.45	0.49
1:A:1[A]:MSE:HG3	1:A:147:GLU:HA	1.96	0.47
1:A:71[B]:GLU:HB2	1:A:74:LEU:HB3	1.95	0.47
1:A:131[B]:LYS:HE3	1:A:134[B]:LEU:CD1	2.45	0.46
1:A:31:TRP:CE2	1:A:36:VAL:HB	2.51	0.46
1:A:24:ILE:CD1	1:A:123[B]:ILE:HG23	2.47	0.45
1:A:45[A]:MSE:HG2	1:A:68[A]:VAL:HG22	1.99	0.45
1:A:131[B]:LYS:HE3	1:A:134[B]:LEU:HD13	2.00	0.43
1:A:1[A]:MSE:HE2	1:A:2:GLU:N	2.27	0.42
1:A:111[B]:THR:HG23	1:A:148[B]:ILE:CD1	2.49	0.41
1:A:142:GLU:O	1:A:146[B]:LYS:HE2	2.20	0.41
1:A:131[B]:LYS:HB3	1:A:134[B]:LEU:HD13	2.03	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:334:HOH:O	2:A:410:HOH:O[3_654]	2.09	0.11

## 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	181/158 (115%)	174 (96%)	6 (3%)	1 (1%)	25 5

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	32	GLY

### 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	155/125 (124%)	146 (94%)	9 (6%)	20 1

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1[A]	MSE
1	A	1[B]	MSE
1	A	44[A]	LYS
1	A	44[B]	LYS
1	A	90	LYS
1	A	131[A]	LYS
1	A	131[B]	LYS
1	A	134[A]	LEU
1	A	134[B]	LEU



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](#)

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

### 6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.