



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 23, 2017 – 09:49 PM EST

PDB ID : 5FQ3  
Title : Crystal structure of the lipoprotein BT2262 from *Bacteroides thetaiotaomicron*  
Authors : Glenwright, A.J.; Pothula, K.R.; Chorev, D.S.; Basle, A.; Robinson, C.V.;  
Kleinekathoefer, U.; Bolam, D.N.; van den Berg, B.  
Deposited on : 2015-12-04  
Resolution : 3.10 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	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
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-20028442

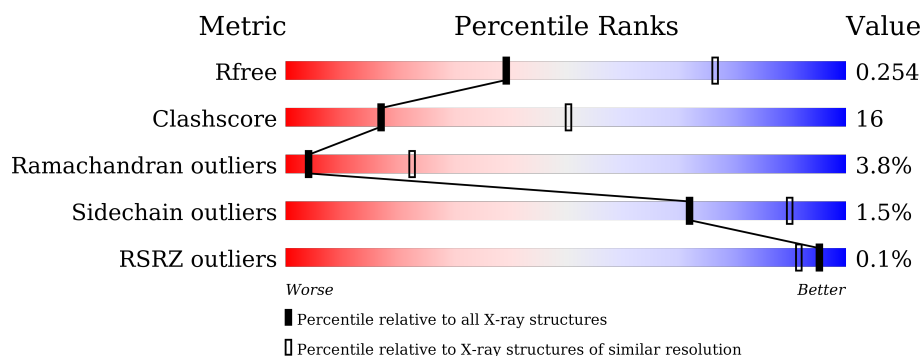
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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	213	<div> <div>69%</div> <div>21%</div> <div>• 7%</div> </div>
1	B	213	<div> <div>64%</div> <div>27%</div> <div>• 7%</div> </div>
1	C	213	<div> <div>62%</div> <div>29%</div> <div>• 7%</div> </div>
1	D	213	<div> <div>57%</div> <div>32%</div> <div>• 7%</div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 6180 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 BT\_2262.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	199	Total	C	N	O	S	0	0	0
			1545	993	238	309	5			
1	B	199	Total	C	N	O	S	0	0	0
			1545	993	238	309	5			
1	C	199	Total	C	N	O	S	0	0	0
			1545	993	238	309	5			
1	D	199	Total	C	N	O	S	0	0	0
			1545	993	238	309	5			

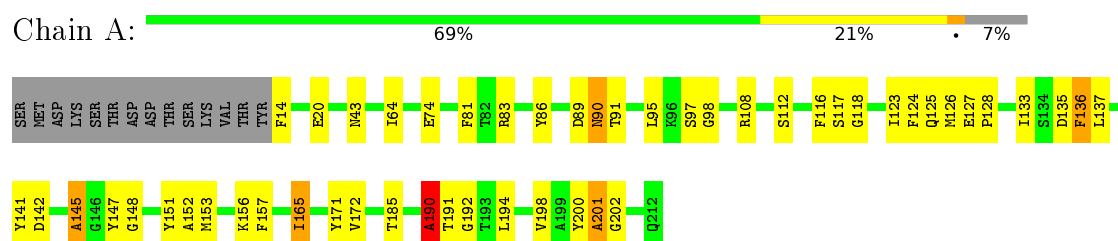
There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	EXPRESSION TAG	UNP Q8A5H7
A	1	MET	-	EXPRESSION TAG	UNP Q8A5H7
B	0	SER	-	EXPRESSION TAG	UNP Q8A5H7
B	1	MET	-	EXPRESSION TAG	UNP Q8A5H7
C	0	SER	-	EXPRESSION TAG	UNP Q8A5H7
C	1	MET	-	EXPRESSION TAG	UNP Q8A5H7
D	0	SER	-	EXPRESSION TAG	UNP Q8A5H7
D	1	MET	-	EXPRESSION TAG	UNP Q8A5H7

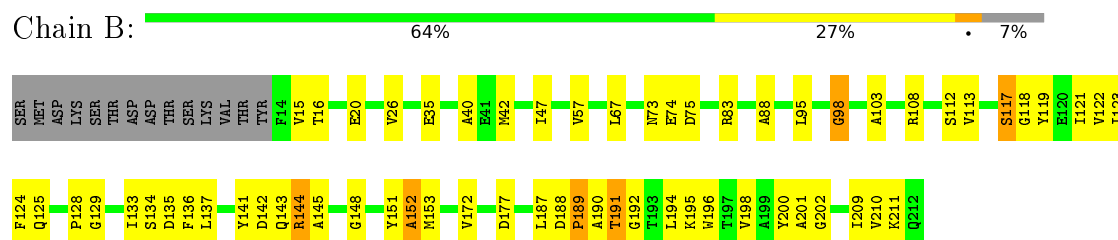
### 3 Residue-property plots

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.

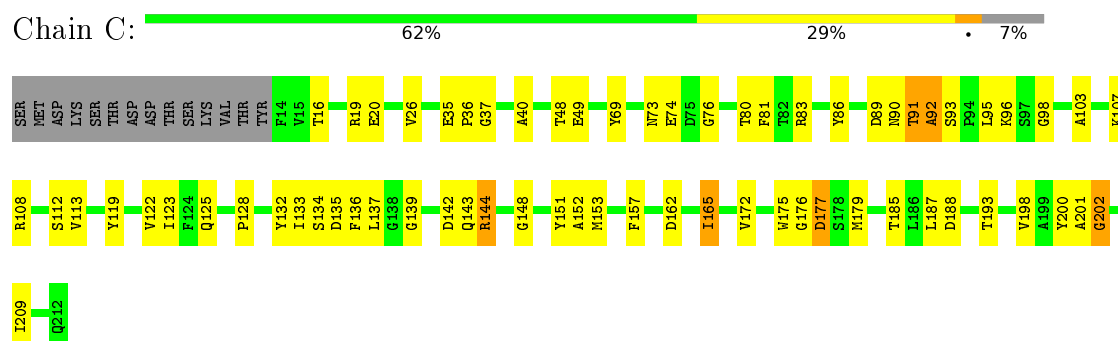
#### • Molecule 1: BT\_2262



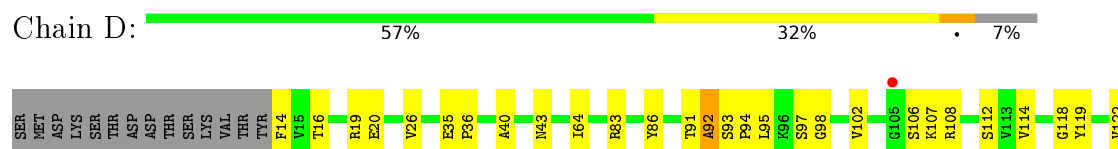
#### • Molecule 1: BT\_2262



#### • Molecule 1: BT\_2262



#### • Molecule 1: BT\_2262



I208	I209	V210	K211	Q212
I123	I124	I125	M126	
F131	I132	I133	S134	D135
F136	L137	G138	G139	W140
Y141	D142	Q143	R144	A145
G146	Y147			
Y151	A152	M153	K156	F157
M160				
M163	T164	I165		
V172				
W175	G176	D177		
L186	L187	D188	P189	A190
G192	T193	L194	K195	W196
Y200	A201	G202	Q203	L204

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	167.32Å 167.32Å 100.34Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.75 – 3.10 48.75 – 3.10	Depositor EDS
% Data completeness (in resolution range)	100.0 (48.75-3.10) 95.9 (48.75-3.10)	Depositor EDS
$R_{merge}$	0.18	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.32 (at 3.12Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.199 , 0.260 0.195 , 0.254	Depositor DCC
$R_{free}$ test set	1922 reflections (7.60%)	DCC
Wilson B-factor (Å <sup>2</sup> )	81.2	Xtriage
Anisotropy	0.177	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 47.4	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	0.94	EDS
Total number of atoms	6180	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.07% 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	0.52	0/1583	0.78	1/2156 (0.0%)
1	B	0.52	0/1583	0.78	0/2156
1	C	0.48	0/1583	0.75	1/2156 (0.0%)
1	D	0.45	0/1583	0.73	0/2156
All	All	0.49	0/6332	0.76	2/8624 (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	A	0	2
1	B	0	1
1	C	0	1
1	D	0	1
All	All	0	5

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	165	ILE	CG1-CB-CG2	-5.22	99.92	111.40
1	A	165	ILE	CG1-CB-CG2	-5.11	100.17	111.40

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	116	PHE	Peptide
1	A	190	ALA	Peptide
1	B	129	GLY	Peptide

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	C	107	LYS	Peptide
1	D	190	ALA	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	1545	0	1470	36	0
1	B	1545	0	1470	49	0
1	C	1545	0	1470	47	0
1	D	1545	0	1470	69	0
All	All	6180	0	5880	196	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 16.

The worst 5 of 196 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:D:20:GLU:O	1:D:83:ARG:NH2	2.12	0.83
1:D:188:ASP:HB2	1:D:193:THR:HG22	1.60	0.82
1:B:153:MET:HE1	1:B:198:VAL:HG21	1.64	0.79
1:A:153:MET:HE1	1:A:198:VAL:HG21	1.64	0.77
1:D:137:LEU:HD13	1:D:153:MET:HG3	1.66	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	197/213 (92%)	171 (87%)	20 (10%)	6 (3%)	5	28
1	B	197/213 (92%)	177 (90%)	12 (6%)	8 (4%)	3	20
1	C	197/213 (92%)	169 (86%)	19 (10%)	9 (5%)	3	17
1	D	197/213 (92%)	166 (84%)	24 (12%)	7 (4%)	4	24
All	All	788/852 (92%)	683 (87%)	75 (10%)	30 (4%)	4	22

5 of 30 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	90	ASN
1	A	145	ALA
1	A	190	ALA
1	C	90	ASN
1	C	91	THR

### 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	163/178 (92%)	160 (98%)	3 (2%)	66	88
1	B	163/178 (92%)	161 (99%)	2 (1%)	78	92
1	C	163/178 (92%)	161 (99%)	2 (1%)	78	92
1	D	163/178 (92%)	160 (98%)	3 (2%)	66	88
All	All	652/712 (92%)	642 (98%)	10 (2%)	72	90

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	144	ARG
1	C	95	LEU
1	D	136	PHE
1	B	95	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	162	ASP

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

Mol	Chain	Res	Type
1	C	143	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	199/213 (93%)	-0.14	0	100 100	52, 75, 107, 161	0
1	B	199/213 (93%)	-0.10	0	100 100	50, 75, 112, 149	0
1	C	199/213 (93%)	-0.13	0	100 100	58, 81, 112, 151	0
1	D	199/213 (93%)	0.02	1 (0%)	91 83	57, 103, 149, 176	0
All	All	796/852 (93%)	-0.09	1 (0%)	95 91	50, 80, 131, 176	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	105	GLY	2.8

### 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.