



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 11:34 AM GMT

PDB ID : 3PE5  
Title : Three-dimensional Structure of protein A7VV38\_9CLOT from Clostridium leptum DSM 753, Northeast Structural Genomics Consortium Target QIR103  
Authors : Kuzin, A.; Su, M.; Seetharaman, J.; Sahdev, S.; Xiao, R.; Ciccocanti, C.; Wang, D.; Everett, J.K.; Nair, R.; Acton, T.B.; Rost, B.; Montelione, G.T.; Hunt, J.F.; Tong, L.; Northeast Structural Genomics Consortium (NESG)  
Deposited on : 2010-10-25  
Resolution : 2.38 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

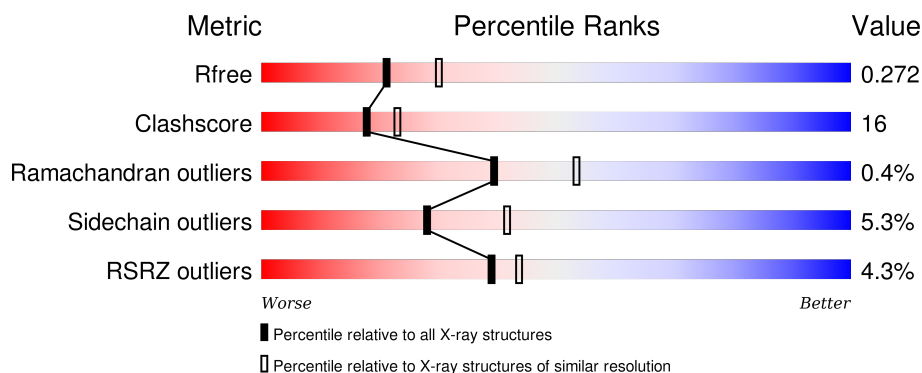
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.38 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	4019 (2.40-2.36)
Clashscore	102246	4595 (2.40-2.36)
Ramachandran outliers	100387	4520 (2.40-2.36)
Sidechain outliers	100360	4522 (2.40-2.36)
RSRZ outliers	91569	4034 (2.40-2.36)

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	403	
1	B	403	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4402 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	278	Total	C	N	O	Se	0	0	0
			2169	1370	357	430	12			
1	B	278	Total	C	N	O	Se	0	0	0
			2169	1370	357	430	12			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	396	LEU	-	EXPRESSION TAG	UNP A7VV38
A	397	GLU	-	EXPRESSION TAG	UNP A7VV38
A	398	HIS	-	EXPRESSION TAG	UNP A7VV38
A	399	HIS	-	EXPRESSION TAG	UNP A7VV38
A	400	HIS	-	EXPRESSION TAG	UNP A7VV38
A	401	HIS	-	EXPRESSION TAG	UNP A7VV38
A	402	HIS	-	EXPRESSION TAG	UNP A7VV38
A	403	HIS	-	EXPRESSION TAG	UNP A7VV38
B	396	LEU	-	EXPRESSION TAG	UNP A7VV38
B	397	GLU	-	EXPRESSION TAG	UNP A7VV38
B	398	HIS	-	EXPRESSION TAG	UNP A7VV38
B	399	HIS	-	EXPRESSION TAG	UNP A7VV38
B	400	HIS	-	EXPRESSION TAG	UNP A7VV38
B	401	HIS	-	EXPRESSION TAG	UNP A7VV38
B	402	HIS	-	EXPRESSION TAG	UNP A7VV38
B	403	HIS	-	EXPRESSION TAG	UNP A7VV38

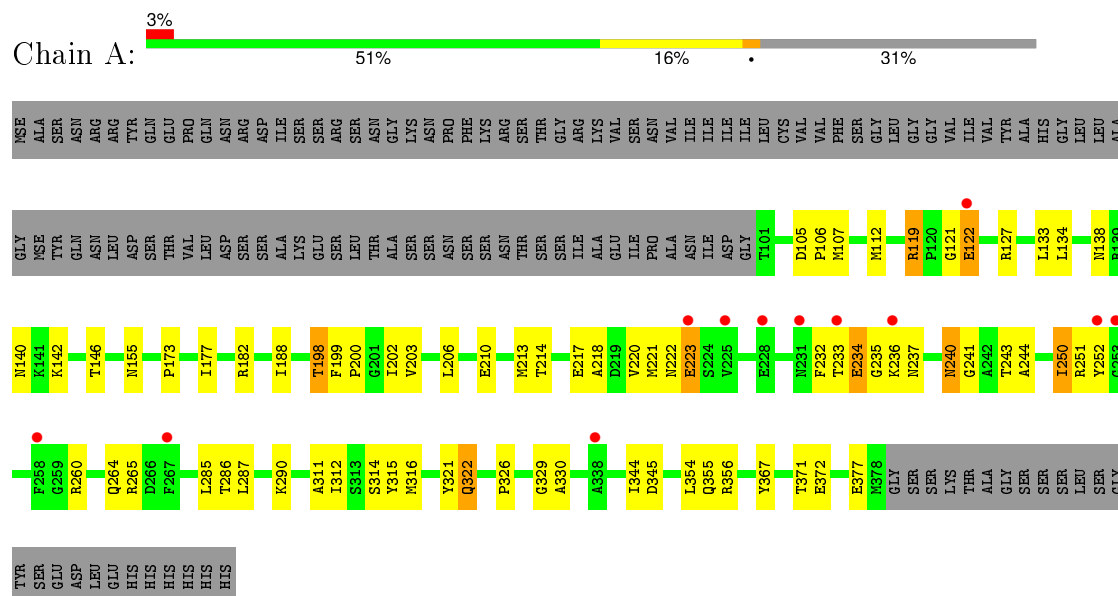
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	36	Total	O	0	0
			36	36		
2	B	28	Total	O	0	0
			28	28		

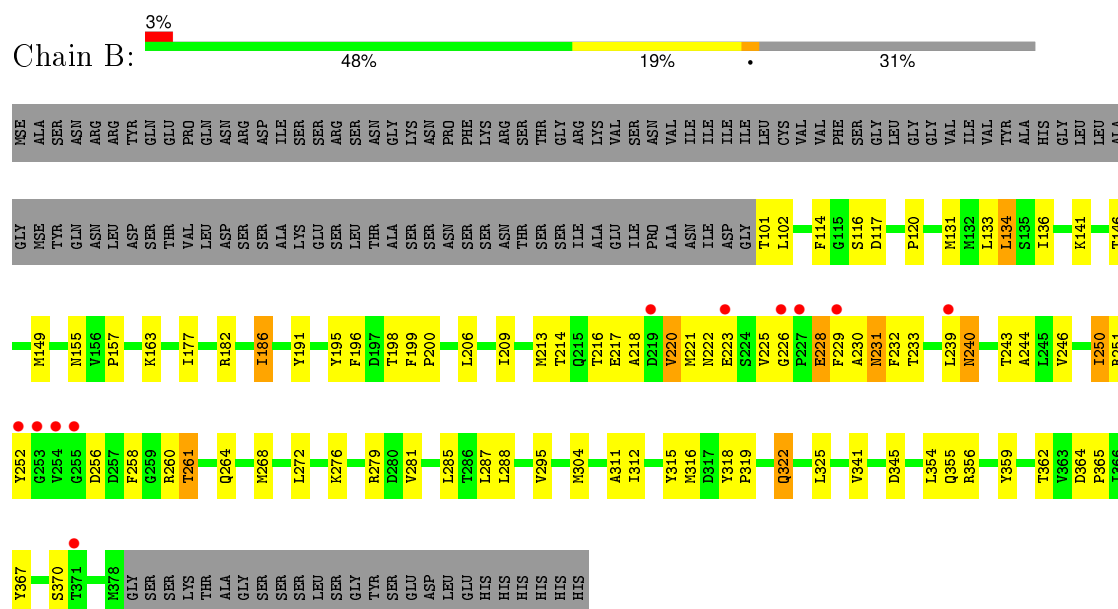
### 3 Residue-property plots

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: Uncharacterized protein



#### • Molecule 1: Uncharacterized protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.78Å 49.62Å 105.28Å 90.00° 106.35° 90.00°	Depositor
Resolution (Å)	35.40 – 2.38 44.54 – 2.38	Depositor EDS
% Data completeness (in resolution range)	95.8 (35.40-2.38) 95.8 (44.54-2.38)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.09 (at 2.39Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.4_486)	Depositor
R, $R_{free}$	0.211 , 0.279 0.202 , 0.272	Depositor DCC
$R_{free}$ test set	1366 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	35.2	Xtriage
Anisotropy	0.715	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 52.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 27299 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4402	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.27% 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	0.43	0/2199	0.57	0/2964
1	B	0.42	0/2199	0.57	0/2964
All	All	0.42	0/4398	0.57	0/5928

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

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	122	GLU	Peptide
1	A	329	GLY	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	2169	0	2107	62	0
1	B	2169	0	2107	74	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	36	0	0	1	0
2	B	28	0	0	0	0
All	All	4402	0	4214	135	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 135 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:233:THR:HG22	1:A:234:GLU:H	1.20	1.01
1:A:233:THR:HG22	1:A:234:GLU:N	1.86	0.90
1:A:233:THR:H	1:A:237:ASN:HD21	1.16	0.90
1:B:258:PHE:O	1:B:261:THR:HG22	1.71	0.90
1:B:356:ARG:HH11	1:B:356:ARG:HG3	1.39	0.86

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	276/403 (68%)	262 (95%)	13 (5%)	1 (0%)	39	53
1	B	276/403 (68%)	262 (95%)	13 (5%)	1 (0%)	39	53
All	All	552/806 (68%)	524 (95%)	26 (5%)	2 (0%)	39	53

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	214	THR
1	B	120	PRO

### 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	234/326 (72%)	222 (95%)	12 (5%)	29	44
1	B	234/326 (72%)	221 (94%)	13 (6%)	26	38
All	All	468/652 (72%)	443 (95%)	25 (5%)	28	42

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	371	THR
1	B	134	LEU
1	B	322	GLN
1	B	133	LEU
1	B	186	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	322	GLN
1	A	355	GLN
1	B	240	ASN
1	A	264	GLN
1	B	264	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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	266/403 (66%)	0.47	12 (4%) 37 42	24, 47, 88, 146	0
1	B	266/403 (66%)	0.45	11 (4%) 41 46	27, 48, 93, 132	0
All	All	532/806 (66%)	0.46	23 (4%) 39 44	24, 48, 91, 146	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	252	TYR	8.8
1	B	227	PRO	5.8
1	A	253	GLY	5.1
1	A	225	VAL	4.1
1	B	252	TYR	4.0

### 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 [i](#)

There are no such residues in this entry.