



# Full wwPDB X-ray Structure Validation Report

Nov 25, 2014 – 11:24 AM EST

PDB ID : 4WUI  
Title : Crystal structure of TrpF from *Jonesia denitrificans*  
Authors : Michalska, K.; Verduzco-Castro, E.A.; Endres, M.; Barona-Gomez, F.;  
Joachimciak, A.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : 2014-10-31  
Resolution : 1.09 Å(reported)

This is a full wwPDB 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/ValidationPDFNotes.html>

---

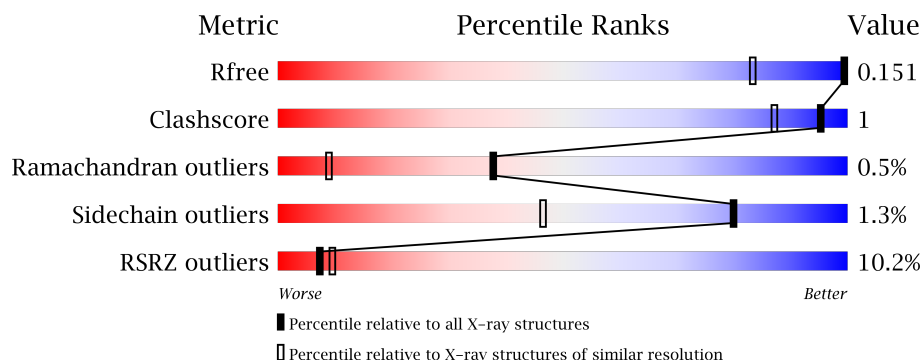
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : stable24195  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.1.3  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable24195

# 1 Overall quality at a glance


The reported resolution of this entry is 1.09 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))
$R_{free}$	66092	1400 (1.20-1.00)
Clashscore	79885	1559 (1.20-1.00)
Ramachandran outliers	78287	1474 (1.20-1.00)
Sidechain outliers	78261	1472 (1.20-1.00)
RSRZ outliers	66119	1400 (1.20-1.00)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	207	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	CIT	A	302	-	X
2	CIT	A	303	-	X

## 2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 1896 atoms, of which 0 are hydrogen and 0 are deuterium.

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.

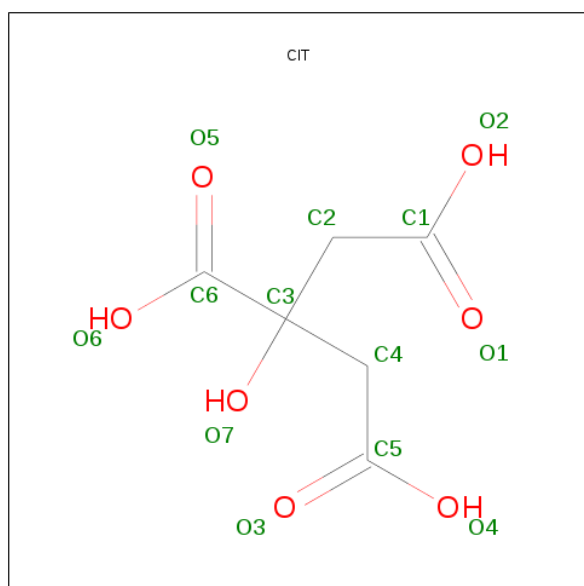
- Molecule 1 is a protein called N-(5'-phosphoribosyl)anthranilate isomerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	205	1591	998	278	310	5	0	12	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP C7R2I5
A	-1	ASN	-	expression tag	UNP C7R2I5
A	0	ALA	-	expression tag	UNP C7R2I5

- Molecule 2 is CITRIC ACID (three-letter code: CIT) (formula: C<sub>6</sub>H<sub>8</sub>O<sub>7</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			13	6	7		
2	A	1	Total	C	O	0	0
			13	6	7		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			13	6	7		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	260	Total	O	0	10
			266	266		

i

- Molecule 1: N-(5'-phosphoribosyl)anthranilate isomerase

Figure 1: A schematic diagram of the 20 amino acids used in the study. The amino acids are arranged in a grid. The first column contains SER, ASN, and AO. The second column contains V5. The third column contains L8. The fourth column contains Q18. The fifth column contains R35. The sixth column contains TI04, SI05, LI06, SI07, AI08, SI09, and PI10. The seventh column contains P127, Q128, AI29, C130, SI31, G132, HI33, TI34, and W135. The eighth column contains G153. The ninth column contains D162. The tenth column contains R198. The eleventh column contains C199, V200, S201, C202, P203, and F204. Red dots are placed above the following amino acids: SER, ASN, AO, V5, L8, Q18, R35, TI04, SI05, LI06, SI07, AI08, SI09, PI10, P127, Q128, AI29, C130, SI31, G132, HI33, TI34, W135, R198, C199, V200, S201, C202, P203, and F204.

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	49.42Å 64.29Å 66.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.71 – 1.09 24.71 – 1.09	Depositor EDS
% Data completeness (in resolution range)	99.3 (24.71-1.09) 99.1 (24.71-1.09)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.49 (at 1.09Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.128 , 0.138 0.141 , 0.151	Depositor DCC
$R_{free}$ test set	1776 reflections (2.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	10.7	Xtriage
Anisotropy	0.437	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.43 , 43.8	EDS
Estimated twinning fraction	0.012 for -h,l,k	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 88164 reflections	Xtriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	1896	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.40% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CIT

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	0/1630	0.76	2/2242 (0.1%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	198	ARG	NE-CZ-NH2	-6.63	116.98	120.30
1	A	198	ARG	NE-CZ-NH1	5.34	122.97	120.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1591	0	1549	4	0
2	A	39	0	15	0	0
3	A	266	0	0	0	0
All	All	1896	0	1564	4	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 1.

All (4) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:5[B]:VAL:HG12	1:A:8:LEU:HD11	1.91	0.53
1:A:135:TRP:HB3	1:A:162[B]:ASP:OD2	2.10	0.50
1:A:5[B]:VAL:CG1	1:A:8:LEU:HD11	2.47	0.44
1:A:104:THR:OG1	1:A:110:PRO:HB3	2.18	0.44

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	215/207 (104%)	208 (97%)	6 (3%)	1 (0%)	38 9

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	153	GLY

### 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	167/157 (106%)	163 (98%)	4 (2%)	61 19

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	18[A]	GLN

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	18[B]	GLN
1	A	35[A]	ARG
1	A	35[B]	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

3 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	CIT	A	301	-	12,12,12	0.97	0	17,17,17	2.05	5 (29%)
2	CIT	A	302	-	12,12,12	0.93	0	17,17,17	1.41	1 (5%)
2	CIT	A	303	-	12,12,12	0.99	0	17,17,17	1.54	4 (23%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means

no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	CIT	A	301	-	-	0/16/16/16	0/0/0/0
2	CIT	A	302	-	-	0/16/16/16	0/0/0/0
2	CIT	A	303	-	-	0/16/16/16	0/0/0/0

There are no bond length outliers.

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	301	CIT	O6-C6-C3	5.14	120.36	112.89
2	A	301	CIT	O5-C6-C3	-4.11	116.54	122.20
2	A	302	CIT	O6-C6-C3	4.02	118.73	112.89
2	A	303	CIT	O6-C6-C3	3.76	118.36	112.89
2	A	303	CIT	C4-C3-C2	-2.68	103.41	109.63
2	A	301	CIT	O2-C1-C2	2.65	123.78	114.63
2	A	301	CIT	O3-C5-C4	-2.44	115.05	122.75
2	A	301	CIT	O2-C1-O1	-2.36	117.34	123.31
2	A	303	CIT	O4-C5-O3	-2.27	117.58	123.31
2	A	303	CIT	O4-C5-C4	2.09	121.86	114.63

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	205/207 (99%)	0.45	21 (10%) 7 9	8, 11, 28, 43	0

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	202	CYS	11.0
1	A	203	PRO	8.2
1	A	204	ARG	7.5
1	A	131	SER	7.4
1	A	133	HIS	6.8
1	A	132	GLY	6.3
1	A	201	SER	5.7
1	A	109	SER	5.4
1	A	107	SER	5.3
1	A	130	GLY	5.2
1	A	108	ALA	4.7
1	A	110	PRO	4.3
1	A	134	THR	3.9
1	A	128	GLN	3.5
1	A	0	ALA	3.0
1	A	106	LEU	2.9
1	A	104	THR	2.7
1	A	127	PRO	2.5
1	A	200	VAL	2.4
1	A	5[A]	VAL	2.1
1	A	105	SER	2.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	CIT	A	302	13/13	0.23	17.82	16,22,31,31	13
2	CIT	A	303	13/13	0.15	8.09	15,18,21,24	0
2	CIT	A	301	13/13	0.13	1.49	13,15,24,25	0

### 6.5 Other polymers ⓘ

There are no such residues in this entry.