



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 02:42 pm BST

PDB ID : 3PST  
Title : Crystal structure of PUL and PFU(mutate) domain  
Authors : Liu, Y.; Sun, J.  
Deposited on : 2010-12-02  
Resolution : 2.00 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

<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
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

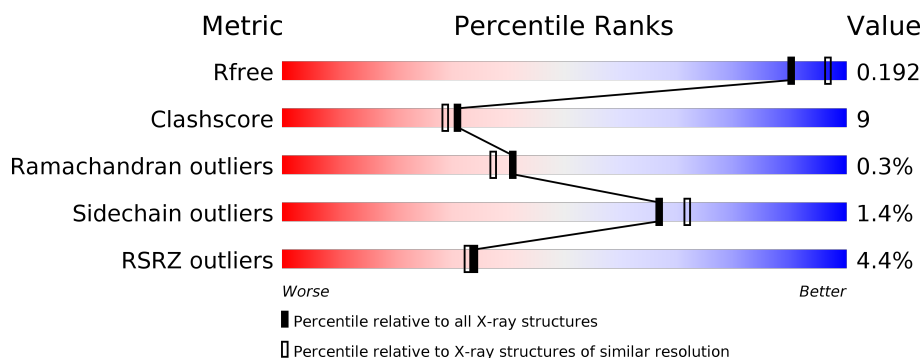
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 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  $\leq 5\%$ . 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	425	<div> <div>3%</div> <div> <div></div> <div>61%</div> <div>13%</div> <div>•</div> <div>25%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2875 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 Protein DOA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	317	Total	C	N	O	S	0	0	0
			2543	1633	408	496	6			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	291	MET	-	EXPRESSION TAG	UNP P36037
A	292	GLY	-	EXPRESSION TAG	UNP P36037
A	293	SER	-	EXPRESSION TAG	UNP P36037
A	294	SER	-	EXPRESSION TAG	UNP P36037
A	295	HIS	-	EXPRESSION TAG	UNP P36037
A	296	HIS	-	EXPRESSION TAG	UNP P36037
A	297	HIS	-	EXPRESSION TAG	UNP P36037
A	298	HIS	-	EXPRESSION TAG	UNP P36037
A	299	HIS	-	EXPRESSION TAG	UNP P36037
A	300	HIS	-	EXPRESSION TAG	UNP P36037
A	301	SER	-	EXPRESSION TAG	UNP P36037
A	302	SER	-	EXPRESSION TAG	UNP P36037
A	303	GLY	-	EXPRESSION TAG	UNP P36037
A	304	LEU	-	EXPRESSION TAG	UNP P36037
A	305	VAL	-	EXPRESSION TAG	UNP P36037
A	306	PRO	-	EXPRESSION TAG	UNP P36037
A	307	ARG	-	EXPRESSION TAG	UNP P36037
A	308	GLY	-	EXPRESSION TAG	UNP P36037
A	309	SER	-	EXPRESSION TAG	UNP P36037
A	310	HIS	-	EXPRESSION TAG	UNP P36037
A	311	MET	-	EXPRESSION TAG	UNP P36037
A	312	ALA	-	EXPRESSION TAG	UNP P36037
A	313	SER	-	EXPRESSION TAG	UNP P36037
A	314	MET	-	EXPRESSION TAG	UNP P36037
A	315	THR	-	EXPRESSION TAG	UNP P36037
A	316	GLY	-	EXPRESSION TAG	UNP P36037
A	317	GLY	-	EXPRESSION TAG	UNP P36037

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	318	GLN	-	EXPRESSION TAG	UNP P36037
A	319	GLN	-	EXPRESSION TAG	UNP P36037
A	320	MET	-	EXPRESSION TAG	UNP P36037
A	321	GLY	-	EXPRESSION TAG	UNP P36037
A	322	ARG	-	EXPRESSION TAG	UNP P36037
A	323	GLY	-	EXPRESSION TAG	UNP P36037
A	324	SER	-	EXPRESSION TAG	UNP P36037
A	434	LEU	PHE	ENGINEERED MUTATION	UNP P36037
A	484	LEU	PHE	ENGINEERED MUTATION	UNP P36037

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	332	Total O 332 332	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.56 Å 101.56 Å 72.02 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	37.53 – 2.00 37.53 – 2.00	Depositor EDS
% Data completeness (in resolution range)	93.5 (37.53-2.00) 99.0 (37.53-2.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.56 (at 2.00 Å)	Xtriage
Refinement program	PHENIX (phenix.refine)	Depositor
R, $R_{free}$	0.188 , 0.226 0.197 , 0.192	Depositor DCC
$R_{free}$ test set	1441 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.3	Xtriage
Anisotropy	0.007	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 50.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.075 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2875	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.09% 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	0.31	0/2592	0.48	0/3524

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	2543	0	2532	46	0
2	A	332	0	0	5	0
All	All	2875	0	2532	46	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 9.

The worst 5 of 46 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:595:VAL:HG12	1:A:599:ILE:HD11	1.61	0.83
1:A:469:VAL:H	1:A:676:ASN:HD21	1.30	0.78
1:A:661:GLN:HE22	1:A:696:ASN:HD22	1.31	0.78
1:A:595:VAL:O	1:A:599:ILE:HG13	1.87	0.74
1:A:378:ILE:HB	2:A:169:HOH:O	1.94	0.68

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	313/425 (74%)	299 (96%)	13 (4%)	1 (0%)	41 37

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	689	ASN

### 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	284/372 (76%)	280 (99%)	4 (1%)	67 72

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

Mol	Chain	Res	Type
1	A	384	THR
1	A	512	ILE
1	A	589	LEU
1	A	594	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:



Mol	Chain	Res	Type
1	A	616	GLN
1	A	689	ASN
1	A	661	GLN
1	A	510	HIS
1	A	676	ASN

### 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	317/425 (74%)	0.16	14 (4%) 34 33	18, 31, 54, 69	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	376	LYS	5.3
1	A	406	VAL	4.1
1	A	463	VAL	3.8
1	A	387	TYR	3.0
1	A	564	LYS	2.9

### 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.