



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 08:01 pm BST

PDB ID : 6K7E  
Title : Crystal structure of MBPapo-Tim21 fusion protein with a 17-residue helical linker  
Authors : Bala, S.; Shimada, A.; Kohda, D.  
Deposited on : 2019-06-07  
Resolution : 1.53 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

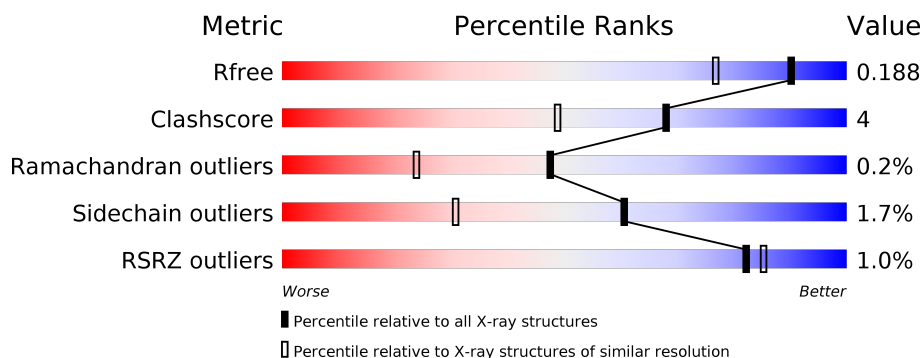
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.53 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	2556 (1.56-1.52)
Clashscore	141614	2634 (1.56-1.52)
Ramachandran outliers	138981	2580 (1.56-1.52)
Sidechain outliers	138945	2577 (1.56-1.52)
RSRZ outliers	127900	2524 (1.56-1.52)

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	507	<div> <div style="display: flex; align-items: center;"> <span style="font-size: small;">%</span> <div style="flex-grow: 1; position: relative;"> <div style="position: absolute; top: -5px; left: 0; width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> <div style="position: absolute; bottom: -5px; left: 0; width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> <div style="position: absolute; top: 5px; left: 0; width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> <div style="position: absolute; top: 5px; left: 0; width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> </div> <div style="margin-left: 5px;"> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>90%</span> <span>7% ..</span> </div> </div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4700 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 Maltose/maltodextrin-binding periplasmic protein, Mitochondrial import inner membrane translocase subunit TIM21.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	498	Total	C	N	O	S	0	0	0
			3913	2506	656	741	10			

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P0AEX9
A	313	VAL	ALA	engineered mutation	UNP P0AEX9
A	370	GLU	-	linker	UNP P0AEX9
A	371	ALA	-	linker	UNP P0AEX9
A	372	ALA	-	linker	UNP P0AEX9
A	373	ALA	-	linker	UNP P0AEX9
A	374	LYS	-	linker	UNP P0AEX9
A	375	GLU	-	linker	UNP P0AEX9
A	376	ALA	-	linker	UNP P0AEX9
A	377	ALA	-	linker	UNP P0AEX9
A	378	ALA	-	linker	UNP P0AEX9
A	379	LYS	-	linker	UNP P0AEX9
A	380	GLU	-	linker	UNP P0AEX9
A	381	ALA	-	linker	UNP P0AEX9
A	382	ALA	-	linker	UNP P0AEX9
A	383	ALA	-	linker	UNP P0AEX9
A	384	LYS	-	linker	UNP P0AEX9
A	385	ALA	-	linker	UNP P0AEX9
A	386	ALA	-	linker	UNP P0AEX9
A	500	LEU	-	expression tag	UNP P53220
A	501	GLU	-	expression tag	UNP P53220
A	502	HIS	-	expression tag	UNP P53220
A	503	HIS	-	expression tag	UNP P53220
A	504	HIS	-	expression tag	UNP P53220
A	505	HIS	-	expression tag	UNP P53220
A	506	HIS	-	expression tag	UNP P53220

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	507	HIS	-	expression tag	UNP P53220

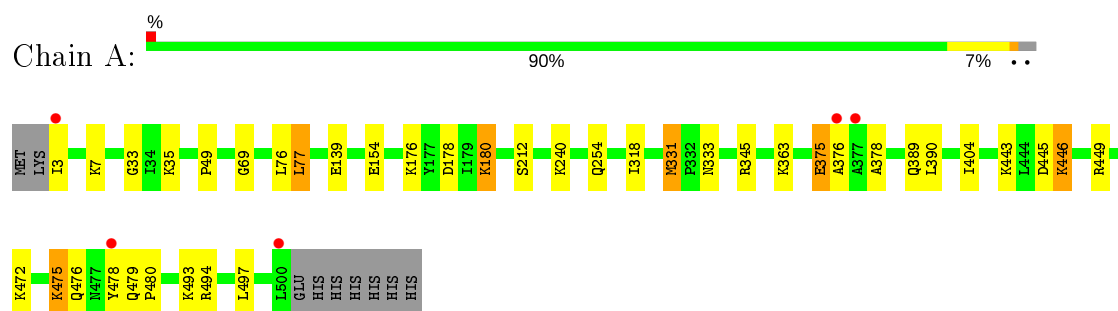
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	787	Total	O	0	0
			787	787		

### 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 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: Maltose/maltodextrin-binding periplasmic protein,Mitochondrial import inner membrane translocase subunit TIM21



## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	156.81Å 156.81Å 67.39Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	39.20 – 1.53 47.83 – 1.53	Depositor EDS
% Data completeness (in resolution range)	99.9 (39.20-1.53) 99.9 (47.83-1.53)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.28 (at 1.53Å)	Xtriage
Refinement program	PHENIX 1.8_1069	Depositor
R, $R_{free}$	0.166 , 0.189 0.164 , 0.188	Depositor DCC
$R_{free}$ test set	4494 reflections (4.87%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	24.1	Xtriage
Anisotropy	0.298	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 55.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.013 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	4700	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.58% 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 [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	0.86	1/4000 (0.0%)	0.84	3/5413 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	331	MET	CB-CG	9.02	1.80	1.51

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	331	MET	CB-CG-SD	-7.63	89.52	112.40
1	A	77	LEU	CB-CG-CD1	5.49	120.34	111.00
1	A	445	ASP	CB-CG-OD1	5.11	122.90	118.30

There are no chirality outliers.

There are no planarity outliers.

### 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	3913	0	3910	30	1
2	A	787	0	0	14	3
All	All	4700	0	3910	30	4

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 4.

The worst 5 of 30 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:331:MET:CB	1:A:331:MET:CG	1.80	1.52
1:A:331:MET:CB	1:A:331:MET:SD	2.54	0.95
1:A:139:GLU:OE2	2:A:601:HOH:O	1.86	0.94
1:A:443:LYS:NZ	2:A:603:HOH:O	2.04	0.89
1:A:212:SER:OG	2:A:602:HOH:O	1.92	0.86

All (4) 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:1158:HOH:O	2:A:1325:HOH:O[6_455]	1.98	0.22
2:A:783:HOH:O	2:A:1244:HOH:O[8_554]	1.99	0.21
2:A:607:HOH:O	2:A:656:HOH:O[6_455]	2.18	0.02
1:A:254:GLN:OE1	1:A:476:GLN:NE2[8_555]	2.19	0.01

## 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	496/507 (98%)	484 (98%)	11 (2%)	1 (0%)	47 24

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	375	GLU



### 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	406/415 (98%)	399 (98%)	7 (2%)	60 31

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	180	LYS
1	A	475	LYS
1	A	363	LYS
1	A	77	LEU
1	A	446	LYS

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

Mol	Chain	Res	Type
1	A	389	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	498/507 (98%)	-0.33	5 (1%) 82 85	8, 17, 36, 61	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	478	TYR	8.4
1	A	500	LEU	5.0
1	A	377	ALA	4.1
1	A	3	ILE	2.9
1	A	376	ALA	2.5

### 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.