



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 23, 2024 – 08:45 AM EDT

PDB ID : 6E16  
Title : Ternary structure of c-Myc-TBP-TAF1  
Authors : Wei, Y.; Dong, A.; Sunnerhagen, M.; Penn, L.; Tong, Y.; Edwards, A.M.;  
Arrowsmith, C.H.; Structural Genomics Consortium (SGC)  
Deposited on : 2018-07-09  
Resolution : 2.40 Å(reported)

This is a Full wwPDB X-ray Structure 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
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)	:	2.37.1

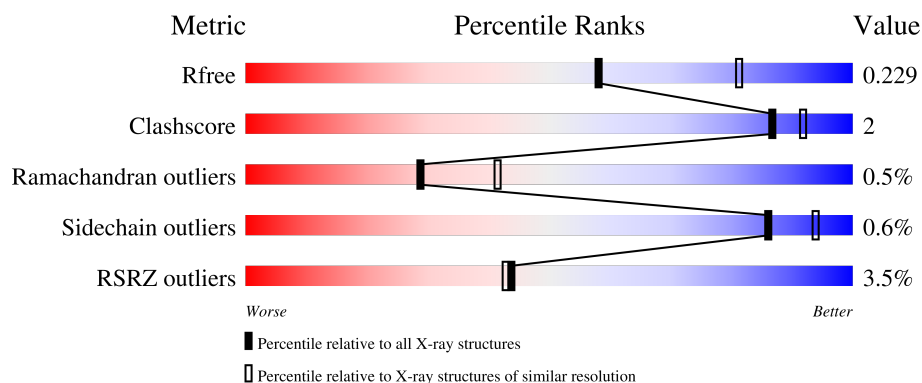
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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 of 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	267	<div> <div>3%</div> <div>79%</div> <div>6%</div> <div>15%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1693 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 Transcription initiation factor TFIID subunit 1, Myc proto-oncogene protein, TATA-box-binding protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	226	Total	C	N	O	S	0	0	1
			1678	1093	271	306	8			

There are 25 discrepancies between the modelled and reference sequences:

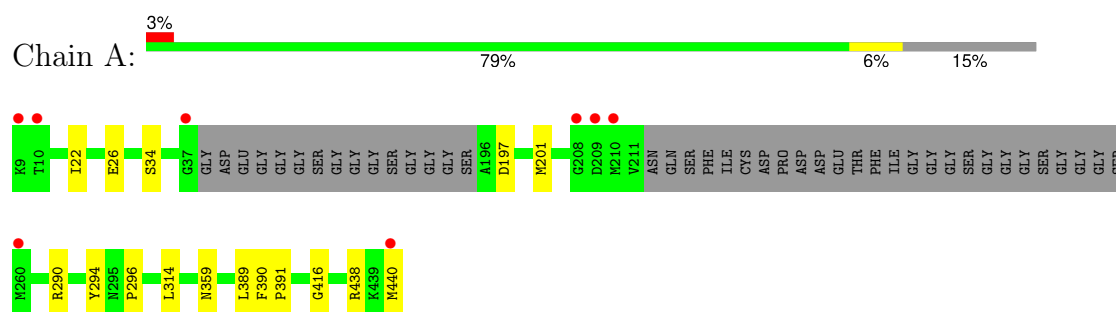
Chain	Residue	Modelled	Actual	Comment	Reference
A	184	GLY	-	linker	UNP P46677
A	185	GLY	-	linker	UNP P46677
A	186	GLY	-	linker	UNP P46677
A	187	SER	-	linker	UNP P46677
A	188	GLY	-	linker	UNP P46677
A	189	GLY	-	linker	UNP P46677
A	190	GLY	-	linker	UNP P46677
A	191	SER	-	linker	UNP P46677
A	192	GLY	-	linker	UNP P46677
A	193	GLY	-	linker	UNP P46677
A	194	GLY	-	linker	UNP P46677
A	195	SER	-	linker	UNP P46677
A	248	GLY	-	linker	UNP P01106
A	249	GLY	-	linker	UNP P01106
A	250	GLY	-	linker	UNP P01106
A	251	SER	-	linker	UNP P01106
A	252	GLY	-	linker	UNP P01106
A	253	GLY	-	linker	UNP P01106
A	254	GLY	-	linker	UNP P01106
A	255	SER	-	linker	UNP P01106
A	256	GLY	-	linker	UNP P01106
A	257	GLY	-	linker	UNP P01106
A	258	GLY	-	linker	UNP P01106
A	259	SER	-	linker	UNP P01106
A	260	MET	-	linker	UNP P01106

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	15	Total	O	0	0
			15	15		



- Molecule 1: Transcription initiation factor TFIID subunit 1, Myc proto-oncogene protein, TATA-box-binding protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.82Å 72.82Å 78.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.41 – 2.40 36.41 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.4 (36.41-2.40) 99.4 (36.41-2.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.28 (at 2.39Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, $R_{free}$	0.177 , 0.220 0.185 , 0.229	Depositor DCC
$R_{free}$ test set	732 reflections (4.55%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	54.4	Xtriage
Anisotropy	0.050	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 48.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.045 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	1693	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.01% 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	0.42	0/1707	0.56	0/2310

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 [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	1678	0	1656	7	0
2	A	15	0	0	0	0
All	All	1693	0	1656	7	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 2.

All (7) 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:359:ASN:HD21	1:A:416:GLY:H	1.34	0.73
1:A:26:GLU:HG2	1:A:389:LEU:HD13	1.88	0.56
1:A:438:ARG:HG2	1:A:440:MET:HG3	1.87	0.55
1:A:390:PHE:CD1	1:A:391:PRO:HD2	2.42	0.55

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:294:TYR:CE2	1:A:296:PRO:HG3	2.42	0.55
1:A:22:ILE:HD11	1:A:314:LEU:HD22	1.95	0.47
1:A:34:SER:OG	1:A:201:MET:SD	2.75	0.42

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	220/267 (82%)	208 (94%)	11 (5%)	1 (0%)	29 41

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	290	ARG

### 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	165/209 (79%)	164 (99%)	1 (1%)	86 94

All (1) residues with a non-rotameric sidechain are listed below:



Mol	Chain	Res	Type
1	A	197	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	11	ASN
1	A	198	GLN
1	A	288	HIS
1	A	344	GLN
1	A	359	ASN
1	A	419	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 monosaccharides 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	226/267 (84%)	-0.16	8 (3%)	44 43	34, 53, 86, 122	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	440	MET	5.0
1	A	209	ASP	4.7
1	A	208	GLY	3.4
1	A	9	LYS	3.0
1	A	260	MET	2.9
1	A	210	MET	2.2
1	A	10	THR	2.2
1	A	37	GLY	2.1

### 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 monosaccharides 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.