

# wwPDB X-ray Structure Validation Summary Report (i)

#### Mar 9, 2024 – 02:06 PM EST

PDB ID	:	3OCI
Title	:	Crystal structure of TBP (TATA box binding protein)
Authors	:	Cui, S.; Wollmann, P.; Moldt, M.; Hopfner, KP.
Deposited on		
Resolution	:	1.90 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 A user guide is available at https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) 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 (1)) were used in the production of this report:

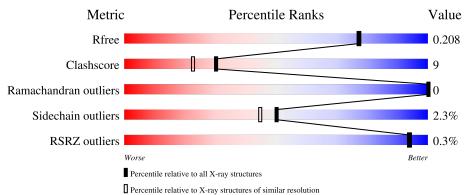
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
$\mathrm{EDS}$	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:  $X\text{-}RAY \, DIFFRACTION$ 

The reported resolution of this entry is 1.90 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\textbf{Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
$R_{free}$	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 >=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 <=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	А	218	64%	17%	18%			
1	В	218	66%	16%	• 17%			

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:



Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	EDO	А	199	-	-	Х	-



# 2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3367 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	۸	179	Total	С	Ν	0	$\mathbf{S}$	17	10	0
	I A	179	1501	969	267	256	9	11	10	0
1	р	180	Total	С	Ν	0	S	-91	1.4	0
	I B	100	1537	990	272	266	9		14	U

• Molecule 1 is a protein called TRANSCRIPTION INITIATION FACTOR TFIID (TFIID-1).

Chain	Residue	Modelled	Actual	Comment	Reference
А	-19	MET	-	expression tag	UNP Q8ST28
А	-18	GLY	-	expression tag	UNP Q8ST28
А	-17	SER	-	expression tag	UNP Q8ST28
А	-16	SER	-	expression tag	UNP Q8ST28
А	-15	HIS	-	expression tag	UNP Q8ST28
А	-14	HIS	-	expression tag	UNP Q8ST28
А	-13	HIS	-	expression tag	UNP Q8ST28
А	-12	HIS	-	expression tag	UNP Q8ST28
А	-11	HIS	-	expression tag	UNP Q8ST28
A	-10	HIS	-	expression tag	UNP Q8ST28
A	-9	SER	-	expression tag	UNP Q8ST28
А	-8	SER	-	expression tag	UNP Q8ST28
А	-7	GLY	-	expression tag	UNP Q8ST28
А	-6	LEU	-	expression tag	UNP Q8ST28
А	-5	VAL	-	expression tag	UNP Q8ST28
А	-4	PRO	-	expression tag	UNP Q8ST28
А	-3	ARG	-	expression tag	UNP Q8ST28
А	-2	GLY	-	expression tag	UNP Q8ST28
А	-1	SER	-	expression tag	UNP Q8ST28
А	0	HIS	-	expression tag	UNP Q8ST28
В	-19	MET	-	expression tag	UNP Q8ST28
В	-18	GLY	-	expression tag	UNP Q8ST28
В	-17	SER	-	expression tag	UNP Q8ST28
В	-16	SER	-	expression tag	UNP Q8ST28
В	-15	HIS	_	expression tag	UNP Q8ST28

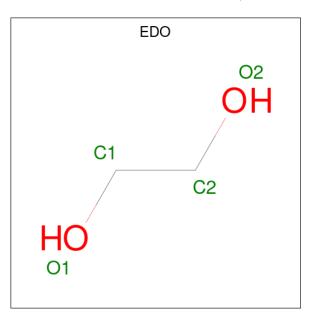
There are 40 discrepancies between the modelled and reference sequences:



Chain	Residue	Modelled	Actual	Comment	Reference
В	-14	HIS	-	expression tag	UNP Q8ST28
В	-13	HIS	-	expression tag	UNP Q8ST28
В	-12	HIS	-	expression tag	UNP Q8ST28
В	-11	HIS	-	expression tag	UNP Q8ST28
В	-10	HIS	-	expression tag	UNP Q8ST28
В	-9	SER	-	expression tag	UNP Q8ST28
В	-8	SER	-	expression tag	UNP Q8ST28
В	-7	GLY	-	expression tag	UNP Q8ST28
В	-6	LEU	-	expression tag	UNP Q8ST28
В	-5	VAL	-	expression tag	UNP Q8ST28
В	-4	PRO	-	expression tag	UNP Q8ST28
В	-3	ARG	-	expression tag	UNP Q8ST28
В	-2	GLY	-	expression tag	UNP Q8ST28
В	-1	SER	-	expression tag	UNP Q8ST28
В	0	HIS	-	expression tag	UNP Q8ST28

Continued from previous page...

• Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0



Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	В	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0
2	В	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0

• Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	144	Total O 144 144	0	0
3	В	161	Total O 161 161	0	0



## 3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: TRANSCRIPTION INITIATION FACTOR TFIID (TFIID-1)





## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
$\begin{array}{c} \text{Cell constants} \\ \text{a, b, c, } \alpha, \beta, \gamma \end{array}$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Depositor
Resolution (Å)	$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	Depositor EDS
% Data completeness	94.0 (24.68-1.90)	Depositor
(in resolution range)	94.0 (26.10-1.90)	EDS
R <sub>merge</sub>	(Not available)	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	2.34 (at 1.91 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.1_357)	Depositor
D D	0.179 , $0.215$	Depositor
$R, R_{free}$	0.172 , $0.208$	DCC
$R_{free}$ test set	1950 reflections $(5.01\%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	25.3	Xtriage
Anisotropy	0.152	Xtriage
Bulk solvent $k_{sol}(e/A^3)$ , $B_{sol}(A^2)$	0.35 , $46.6$	EDS
L-test for twinning <sup>2</sup>	$<  L  > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	$\begin{array}{c} 0.036 \; {\rm for} \; -2/3 ^{\rm *h-1}/3 ^{\rm *k} + 2/3 ^{\rm *l}, -1/3 ^{\rm *h-2}/3 ^{\rm *k} + 2/3 ^{\rm *l}, 2/3 ^{\rm *h-2}/3 ^{\rm *k} + 1/3 ^{\rm *l} \\ 0.022 \; {\rm for} \; -{\rm h}, 1/3 ^{\rm *h-1}/3 ^{\rm *k} + 2/3 ^{\rm *l}, 2/3 ^{\rm *h} + 4/3 ^{\rm *k} \\ {\rm k} + 1/3 ^{\rm *l} \\ 0.025 \; {\rm for} \; -1/3 ^{\rm *h+1}/3 ^{\rm *k} - 2/3 ^{\rm *l}, -{\rm k}, -4/3 ^{\rm *h-2}/3 \\ {\rm *k} + 1/3 ^{\rm *l} \\ 0.023 \; {\rm for} \; -{\rm h}, 2/3 ^{\rm *h} + 1/3 ^{\rm *k} - 2/3 ^{\rm *l}, -2/3 ^{\rm *h-4}/3 ^{\rm *k} \\ {\rm k} - 1/3 ^{\rm *l} \\ 0.026 \; {\rm for} \; 1/3 ^{\rm *h+2}/3 ^{\rm *k} + 2/3 ^{\rm *l}, -{\rm k}, 4/3 ^{\rm *h+2}/3 \\ {\rm *k} - 1/3 ^{\rm *l} \\ 0.037 \; {\rm for} \; -1/3 ^{\rm *h-2}/3 ^{\rm *k} - 2/3 ^{\rm *l}, -2/3 ^{\rm *h-1}/3 ^{\rm *k} + 2/3 ^{\rm *l}, -2/3 ^{\rm *h-2}/3 ^{\rm *l}, -2/3 ^{\rm *h-1}/3 ^{\rm *l} \\ 0.487 \; {\rm for} \; {\rm h}, {\rm h}, {\rm h}, {\rm k}, {\rm l} \end{array}$	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3367	wwPDB-VP
Average B, all atoms $(Å^2)$	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 4.90% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 5 Model quality (i)

### 5.1 Standard geometry (i)

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

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		
	Unam	RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.40	0/1532	0.65	3/2062~(0.1%)	
1	В	0.40	0/1568	0.65	3/2110~(0.1%)	
All	All	0.40	0/3100	0.65	6/4172~(0.1%)	

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$\mathbf{Observed}(^{o})$	$Ideal(^{o})$
1	А	178	ARG	NE-CZ-NH2	-12.46	114.07	120.30
1	А	178	ARG	NE-CZ-NH1	12.09	126.34	120.30
1	В	178	ARG	NE-CZ-NH1	-12.00	114.30	120.30
1	В	178	ARG	NE-CZ-NH2	11.46	126.03	120.30
1	А	178	ARG	CD-NE-CZ	5.84	131.77	123.60

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	А	1501	0	1577	33	0
1	В	1537	0	1605	29	0
2	А	16	0	24	7	0
2	В	8	0	12	2	0



Continued from previous page...

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 58 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:B:34:LEU:HD12	1:B:38:LEU:HD11	1.76	0.67
1:A:126:PHE:CE1	1:A:196[B]:ARG:HG2	2.29	0.67
1:A:34:LEU:HD12	1:A:38:LEU:HD11	1.76	0.67
1:B:32:VAL:HG12	1:B:113:PHE:HA	1.78	0.66
1:A:32:VAL:HG12	1:A:113:PHE:HA	1.80	0.64

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	Perce	entiles
1	А	187/218~(86%)	184 (98%)	3~(2%)	0	100	100
1	В	192/218~(88%)	188 (98%)	4 (2%)	0	100	100
All	All	379/436~(87%)	372~(98%)	7 (2%)	0	100	100

There are no Ramachandran outliers to report.



Chain Non-H H(model) H(added) Clashes Symm-Clashes Mol 3 14450 А 0 0 3 В 0 5 0 1610 All All 3367 0 3218 580

#### 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	А	164/188~(87%)	161~(98%)	3~(2%)	59 55		
1	В	168/188~(89%)	163~(97%)	5(3%)	41 33		
All	All	332/376~(88%)	324~(98%)	8 (2%)	50 43		

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	В	111	ASP
1	В	104	LEU
1	В	92[A]	MET
1	В	65	ARG
1	В	92[B]	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	А	26	GLN
1	А	42	ASN
1	А	137	HIS
1	В	26	GLN
1	В	42	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 monosaccharides in this entry.

### 5.6 Ligand geometry (i)

6 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	Turne	Chain	Res	Link	В	ond leng	$\operatorname{gths}$	B	Bond ang	gles
NIOI	Type	Chain	nes	LIIIK	Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2
2	EDO	В	199	-	$3,\!3,\!3$	0.41	0	$2,\!2,\!2$	0.57	0
2	EDO	А	200	-	3,3,3	0.66	0	$2,\!2,\!2$	0.12	0
2	EDO	А	201	-	3,3,3	0.43	0	2,2,2	0.51	0
2	EDO	А	199	-	3,3,3	0.52	0	2,2,2	0.20	0
2	EDO	А	202	-	3,3,3	0.53	0	2,2,2	0.60	0
2	EDO	В	200	-	3,3,3	0.70	0	2,2,2	0.25	0

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	EDO	В	199	-	-	0/1/1/1	-
2	EDO	А	200	-	-	0/1/1/1	-
2	EDO	А	201	-	-	0/1/1/1	-
2	EDO	А	199	-	-	1/1/1/1	-
2	EDO	А	202	-	-	1/1/1/1	-
2	EDO	В	200	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:



Mol	Chain	Res	Type	Atoms
2	А	199	EDO	O1-C1-C2-O2
2	А	202	EDO	O1-C1-C2-O2

There are no ring outliers.

3 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	В	199	EDO	2	0
2	А	201	EDO	1	0
2	А	199	EDO	6	0

#### 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,  $95^{th}$  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	$\langle RSRZ \rangle$	# RSRZ > 2	$OWAB(Å^2)$	Q<0.9
1	А	179/218~(82%)	-0.45	0 100 100	15, 27, 51, 77	6 (3%)
1	В	180/218~(82%)	-0.45	1 (0%) 89 90	16, 27, 56, 77	8 (4%)
All	All	359/436~(82%)	-0.45	1 (0%) 94 94	15, 27, 52, 77	14 (3%)

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	В	19	SER	2.3

### 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)

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. The B-factors column lists the minimum, median,  $95^{th}$  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	RSCC	RSR	$B-factors(Å^2)$	$Q{<}0.9$
2	EDO	А	200	4/4	0.92	0.14	$26,\!26,\!33,\!35$	0
2	EDO	А	202	4/4	0.92	0.14	24,34,43,43	0
2	EDO	В	200	4/4	0.93	0.10	22,24,29,31	0
2	EDO	А	201	4/4	0.96	0.08	$19,\!25,\!25,\!37$	0



Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$B-factors(Å^2)$	Q < 0.9
2	EDO	В	199	4/4	0.97	0.06	21,25,29,34	0
2	EDO	А	199	4/4	0.97	0.14	$19,\!33,\!43,\!46$	0

### 6.5 Other polymers (i)

There are no such residues in this entry.

