



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 16, 2023 – 04:09 pm BST

PDB ID : 8PXH  
Title : Structure of TauA from E. coli, solved at wavelength 2.375 Å  
Authors : El Omari, K.; Duman, R.; Mykhaylyk, V.; Orr, C.; Qu, F.; Beis, K.; Wagner, A.  
Deposited on : 2023-07-23  
Resolution : 1.77 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 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  
Mogul : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
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

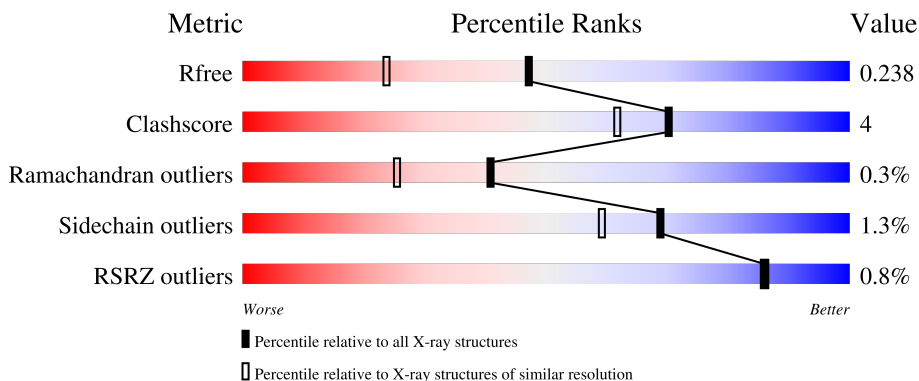
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.77 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	9185 (1.80-1.76)
Clashscore	141614	10184 (1.80-1.76)
Ramachandran outliers	138981	10051 (1.80-1.76)
Sidechain outliers	138945	10050 (1.80-1.76)
RSRZ outliers	127900	9032 (1.80-1.76)

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	298	 92% 7%
1	B	298	 92% 8%

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
3	IOD	A	406	-	-	-	X
3	IOD	B	405	-	-	X	-

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4848 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.

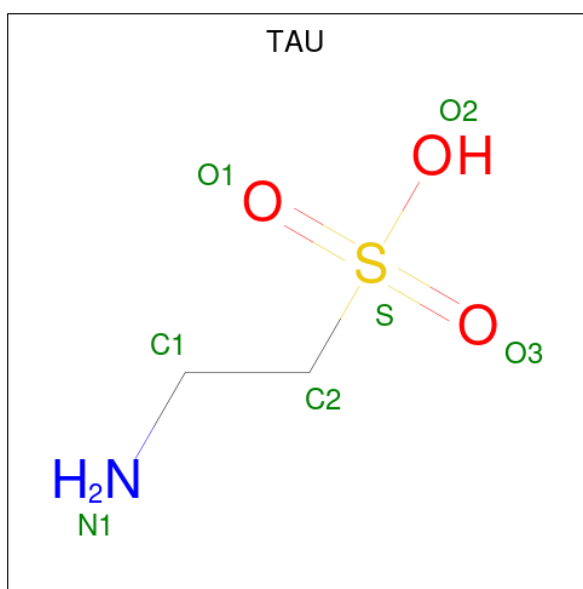
- Molecule 1 is a protein called Taurine ABC transporter substrate-binding protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	298	2267	1451	384	431	1	0	2	0
1	B	297	2262	1448	383	431		0	3	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	22	MET	-	initiating methionine	UNP C3TMR2
B	22	MET	-	initiating methionine	UNP C3TMR2

- Molecule 2 is 2-AMINOETHANESULFONIC ACID (three-letter code: TAU) (formula:  $C_2H_7NO_3S$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	7	2	1	3	1	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	B	1	7	2	1	3	1	0	0

- Molecule 3 is IODIDE ION (three-letter code: IOD) (formula: I).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	5	Total I 5 5	0	0
3	B	6	Total I 6 6	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	154	Total O 154 154	0	0
4	B	140	Total O 140 140	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: Taurine ABC transporter substrate-binding protein

Chain A:  92% 7%



- Molecule 1: Taurine ABC transporter substrate-binding protein

Chain B:  92% 8%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	45.01Å 71.32Å 165.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	82.51 – 1.77 82.52 – 1.77	Depositor EDS
% Data completeness (in resolution range)	85.7 (82.51-1.77) 85.7 (82.52-1.77)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.64 (at 1.77Å)	Xtrriage
Refinement program	REFMAC 5.8.0403	Depositor
R, $R_{free}$	0.202 , 0.231 0.215 , 0.238	Depositor DCC
$R_{free}$ test set	2251 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	27.8	Xtrriage
Anisotropy	0.106	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 36.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4848	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.07% 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](#)

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

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	0.53	1/2323 (0.0%)	0.82	7/3171 (0.2%)
1	B	0.41	0/2321	0.67	0/3169
All	All	0.48	1/4644 (0.0%)	0.75	7/6340 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	219	GLU	CD-OE1	-7.20	1.17	1.25

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	155	ASN	CB-CG-OD1	10.03	141.65	121.60
1	A	155	ASN	OD1-CG-ND2	-9.43	100.22	121.90
1	A	167	ARG	CG-CD-NE	-8.28	94.42	111.80
1	A	167	ARG	NE-CZ-NH1	-7.67	116.47	120.30
1	A	306	VAL	CA-CB-CG2	6.38	120.46	110.90

There are no chirality outliers.

All (2) planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	A	167	ARG	Sidechain
1	B	254	ARG	Sidechain

## 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	2267	0	2294	17	0
1	B	2262	0	2292	15	0
2	A	7	0	7	0	0
2	B	7	0	7	0	0
3	A	5	0	0	1	0
3	B	6	0	0	3	0
4	A	154	0	0	9	0
4	B	140	0	0	6	0
All	All	4848	0	4600	33	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 4.

The worst 5 of 33 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:97:LEU:HB2	4:A:556:HOH:O	1.59	1.00
3:A:404:IOD:I	4:A:583:HOH:O	2.48	0.99
1:A:55:TRP:CE2	4:A:624:HOH:O	2.35	0.80
1:B:157:GLN:NE2	3:B:406:IOD:I	2.86	0.78
4:A:571:HOH:O	3:B:405:IOD:I	2.72	0.77

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	298/298 (100%)	293 (98%)	4 (1%)	1 (0%)	41	25
1	B	298/298 (100%)	295 (99%)	2 (1%)	1 (0%)	41	25
All	All	596/596 (100%)	588 (99%)	6 (1%)	2 (0%)	41	25

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	130	ILE
1	B	130	ILE

### 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	242/240 (101%)	238 (98%)	4 (2%)	60	48
1	B	242/240 (101%)	240 (99%)	2 (1%)	81	76
All	All	484/480 (101%)	478 (99%)	6 (1%)	69	62

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	174	TYR
1	B	34	GLU
1	B	174	TYR
1	A	147	LYS
1	A	34	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	155	ASN
1	B	217	HIS

### 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](#)

Of 13 ligands modelled in this entry, 11 are monoatomic - leaving 2 for Mogul analysis.

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	TAU	B	402	-	6,6,6	1.11	1 (16%)	6,8,8	1.05	1 (16%)
2	TAU	A	401	-	6,6,6	0.53	0	6,8,8	1.51	2 (33%)

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	TAU	B	402	-	-	0/4/4/4	-
2	TAU	A	401	-	-	0/4/4/4	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	402	TAU	O2-S	2.44	1.56	1.47

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	402	TAU	O2-S-C2	2.35	109.58	105.77
2	A	401	TAU	O1-S-C2	-2.28	104.17	106.92
2	A	401	TAU	O3-S-O1	2.20	121.55	113.95

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	298/298 (100%)	0.21	4 (1%) 77 77	18, 26, 45, 64	0
1	B	297/298 (99%)	0.01	1 (0%) 94 93	19, 26, 40, 50	0
All	All	595/596 (99%)	0.11	5 (0%) 86 86	18, 26, 42, 64	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	305	ALA	3.7
1	A	306	VAL	3.6
1	A	155	ASN	2.8
1	B	24	ASN	2.4
1	A	113	THR	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<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	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	IOD	B	406	1/1	0.34	0.34	128,128,128,128	0
3	IOD	A	406	1/1	0.75	0.41	100,100,100,100	0
3	IOD	B	407	1/1	0.94	0.20	108,108,108,108	0
2	TAU	A	401	7/7	0.96	0.13	20,22,27,27	0
3	IOD	B	405	1/1	0.97	0.19	112,112,112,112	0
3	IOD	A	403	1/1	0.97	0.25	75,75,75,75	0
3	IOD	B	403	1/1	0.97	0.08	48,48,48,48	1
3	IOD	A	404	1/1	0.99	0.09	33,33,33,33	0
3	IOD	B	404	1/1	0.99	0.17	50,50,50,50	0
3	IOD	A	405	1/1	0.99	0.21	49,49,49,49	0
2	TAU	B	402	7/7	0.99	0.08	19,19,23,25	0
3	IOD	B	401	1/1	0.99	0.14	48,48,48,48	0
3	IOD	A	402	1/1	1.00	0.02	23,23,23,23	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.