



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 30, 2025 – 04:07 PM JST

PDB ID : 9WVI / pdb\_00009wvi  
Title : Crystal structure of TONSL UBL mutant - R934W  
Authors : Karmakar, A.; Roy, S.  
Deposited on : 2025-09-20  
Resolution : 2.19 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 : 4-5-2 with Phenix2.0  
Xtriage (Phenix) : 2.0  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.010 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.46

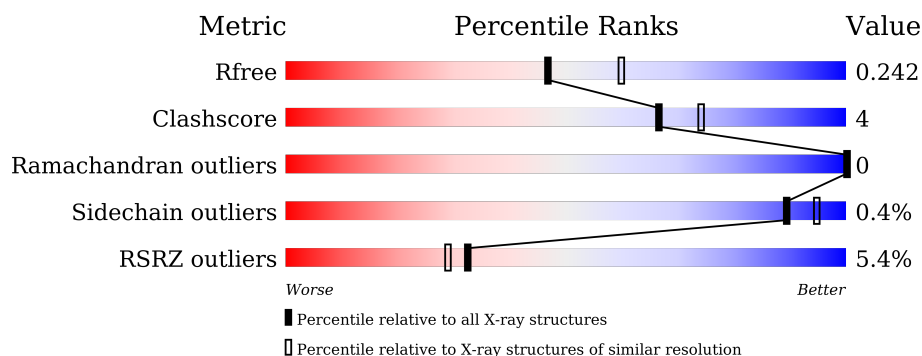
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.19 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))
$R_{free}$	164625	5791 (2.20-2.20)
Clashscore	180529	6634 (2.20-2.20)
Ramachandran outliers	177936	6560 (2.20-2.20)
Sidechain outliers	177891	6561 (2.20-2.20)
RSRZ outliers	164620	5791 (2.20-2.20)

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	85	<div> <div>2%</div> <div> <div></div> <div>86%</div> <div>7%</div> <div>7%</div> </div> </div>
1	B	85	<div> <div>6%</div> <div> <div></div> <div>82%</div> <div>13%</div> <div>5%</div> </div> </div>
1	C	85	<div> <div>8%</div> <div> <div></div> <div>76%</div> <div>13%</div> <div>11%</div> </div> </div>
1	D	85	<div> <div>4%</div> <div> <div></div> <div>89%</div> <div>5%</div> <div>6%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2656 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 Tonsoku-like protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	79	Total	C	N	O	S	0	0	0
			638	408	111	118	1			
1	B	81	Total	C	N	O	S	0	0	0
			654	417	115	121	1			
1	C	76	Total	C	N	O	S	0	0	0
			613	395	106	111	1			
1	D	80	Total	C	N	O	S	0	0	0
			644	411	112	120	1			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	927	HIS	-	expression tag	UNP Q96HA7
A	928	HIS	-	expression tag	UNP Q96HA7
A	929	HIS	-	expression tag	UNP Q96HA7
A	930	HIS	-	expression tag	UNP Q96HA7
A	931	HIS	-	expression tag	UNP Q96HA7
A	932	HIS	-	expression tag	UNP Q96HA7
A	934	TRP	ARG	variant	UNP Q96HA7
B	927	HIS	-	expression tag	UNP Q96HA7
B	928	HIS	-	expression tag	UNP Q96HA7
B	929	HIS	-	expression tag	UNP Q96HA7
B	930	HIS	-	expression tag	UNP Q96HA7
B	931	HIS	-	expression tag	UNP Q96HA7
B	932	HIS	-	expression tag	UNP Q96HA7
B	934	TRP	ARG	variant	UNP Q96HA7
C	927	HIS	-	expression tag	UNP Q96HA7
C	928	HIS	-	expression tag	UNP Q96HA7
C	929	HIS	-	expression tag	UNP Q96HA7
C	930	HIS	-	expression tag	UNP Q96HA7
C	931	HIS	-	expression tag	UNP Q96HA7
C	932	HIS	-	expression tag	UNP Q96HA7
C	934	TRP	ARG	variant	UNP Q96HA7

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	927	HIS	-	expression tag	UNP Q96HA7
D	928	HIS	-	expression tag	UNP Q96HA7
D	929	HIS	-	expression tag	UNP Q96HA7
D	930	HIS	-	expression tag	UNP Q96HA7
D	931	HIS	-	expression tag	UNP Q96HA7
D	932	HIS	-	expression tag	UNP Q96HA7
D	934	TRP	ARG	variant	UNP Q96HA7

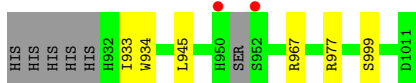
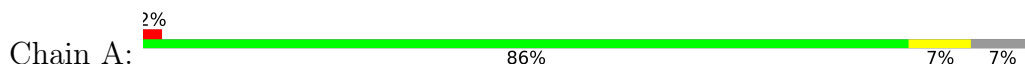
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	37	Total O 37 37	0	0
2	B	34	Total O 34 34	0	0
2	C	22	Total O 22 22	0	0
2	D	14	Total O 14 14	0	0

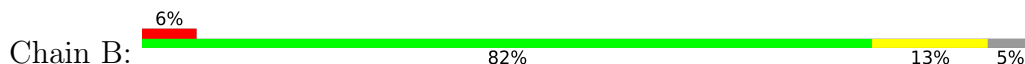
### 3 Residue-property plots [i](#)

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.

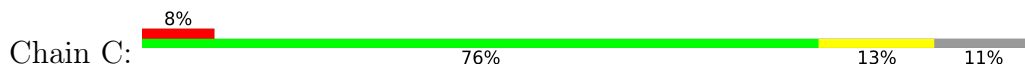
- Molecule 1: Tonsoku-like protein



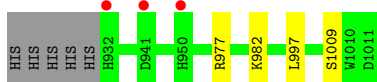
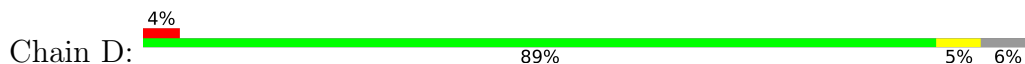
- Molecule 1: Tonsoku-like protein



- Molecule 1: Tonsoku-like protein



- Molecule 1: Tonsoku-like protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	39.88Å 81.11Å 102.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.55 – 2.19 40.55 – 2.19	Depositor EDS
% Data completeness (in resolution range)	99.9 (40.55-2.19) 99.9 (40.55-2.19)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.91 (at 2.18Å)	Xtriage
Refinement program	PHENIX 1.21.2_5419	Depositor
R, $R_{free}$	0.204 , 0.243 0.205 , 0.242	Depositor DCC
$R_{free}$ test set	890 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	41.6	Xtriage
Anisotropy	0.215	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 44.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2656	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.41% 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	0.35	0/654	0.54	0/895
1	B	0.33	0/672	0.55	0/921
1	C	0.35	0/628	0.55	0/861
1	D	0.30	0/661	0.52	0/906
All	All	0.33	0/2615	0.54	0/3583

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	638	0	624	5	0
1	B	654	0	637	7	0
1	C	613	0	608	6	0
1	D	644	0	630	2	0
2	A	37	0	0	0	0
2	B	34	0	0	1	0
2	C	22	0	0	0	0
2	D	14	0	0	0	0
All	All	2656	0	2499	18	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 18 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:977:ARG:CZ	1:A:977:ARG:HA	2.27	0.64
1:A:933:ILE:HD12	1:A:999:SER:HB2	1.83	0.60
1:B:977:ARG:HH21	1:C:966:GLN:NE2	2.04	0.55
1:D:977:ARG:HD2	1:D:1009:SER:OG	2.07	0.54
1:C:934:TRP:HB3	1:C:945:LEU:HD11	1.93	0.50

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	75/85 (88%)	74 (99%)	1 (1%)	0	100	100
1	B	79/85 (93%)	79 (100%)	0	0	100	100
1	C	72/85 (85%)	71 (99%)	1 (1%)	0	100	100
1	D	78/85 (92%)	77 (99%)	1 (1%)	0	100	100
All	All	304/340 (89%)	301 (99%)	3 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	70/76 (92%)	70 (100%)	0	100	100
1	B	72/76 (95%)	71 (99%)	1 (1%)	62	77
1	C	67/76 (88%)	67 (100%)	0	100	100
1	D	71/76 (93%)	71 (100%)	0	100	100
All	All	280/304 (92%)	279 (100%)	1 (0%)	89	95

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

Mol	Chain	Res	Type
1	B	1004	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	955	HIS
1	D	942	HIS
1	C	966	GLN
1	B	955	HIS
1	D	938	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 oligosaccharides 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	79/85 (92%)	0.04	2 (2%) 58 55	31, 43, 67, 103	0
1	B	81/85 (95%)	0.27	5 (6%) 28 25	34, 48, 68, 95	0
1	C	76/85 (89%)	0.36	7 (9%) 16 14	34, 47, 75, 115	0
1	D	80/85 (94%)	0.40	3 (3%) 44 41	39, 56, 95, 105	0
All	All	316/340 (92%)	0.27	17 (5%) 32 29	31, 48, 77, 115	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	932	HIS	5.0
1	B	931	HIS	4.0
1	C	949	PRO	3.1
1	C	953	ASP	3.0
1	C	933	ILE	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 oligosaccharides 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.