



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 22, 2024 – 09:06 PM EDT

PDB ID : 6G6N  
Title : Crystal structure of the computationally designed Tako8 protein in C2  
Authors : Noguchi, H.; Addy, C.; Simoncini, D.; Van Meervelt, L.; Schiex, T.; Zhang, K.Y.J.; Tame, J.R.H.; Voet, A.R.D.  
Deposited on : 2018-04-01  
Resolution : 2.00 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
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

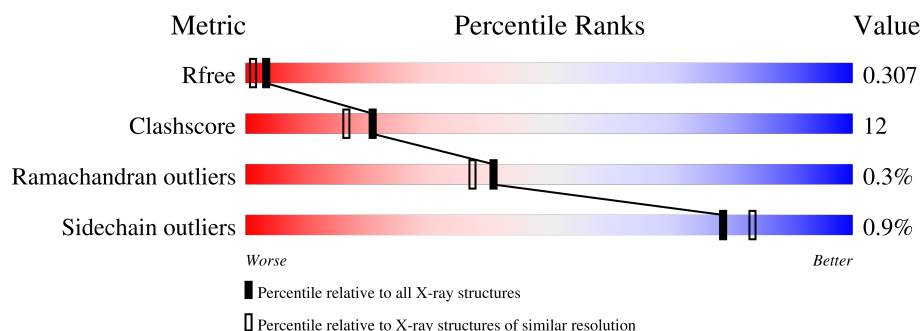
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*




The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	325	 77% 21% .
1	B	325	 75% 22% .
1	C	325	 71% 26% ..

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7475 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 Tako8.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	319	Total	C	N	O	0	0	0
			2425	1484	431	510			
1	B	318	Total	C	N	O	0	0	0
			2416	1479	429	508			
1	C	318	Total	C	N	O	0	0	0
			2416	1479	429	508			

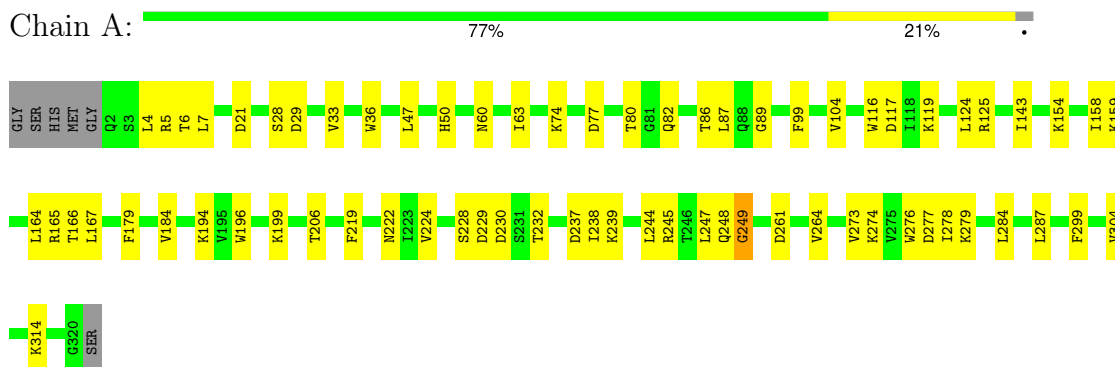
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	70	Total	O	0	0
			70	70		
2	B	76	Total	O	0	0
			76	76		
2	C	72	Total	O	0	0
			72	72		

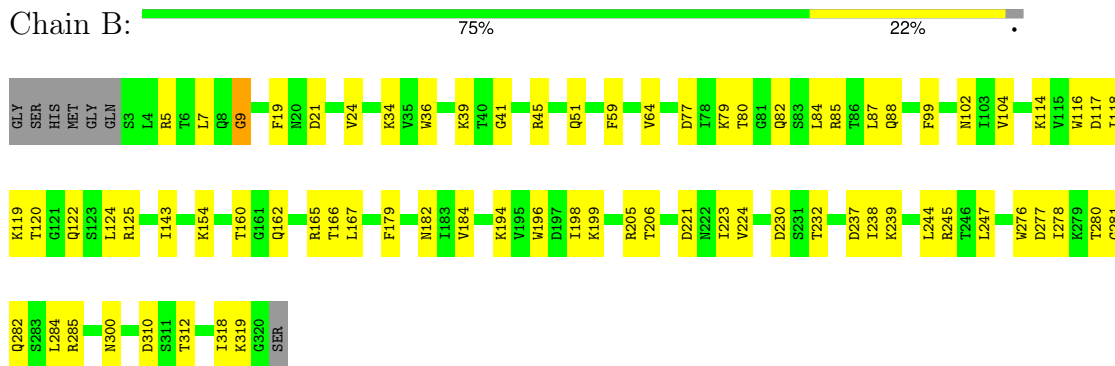
### 3 Residue-property plots

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. 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. 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.

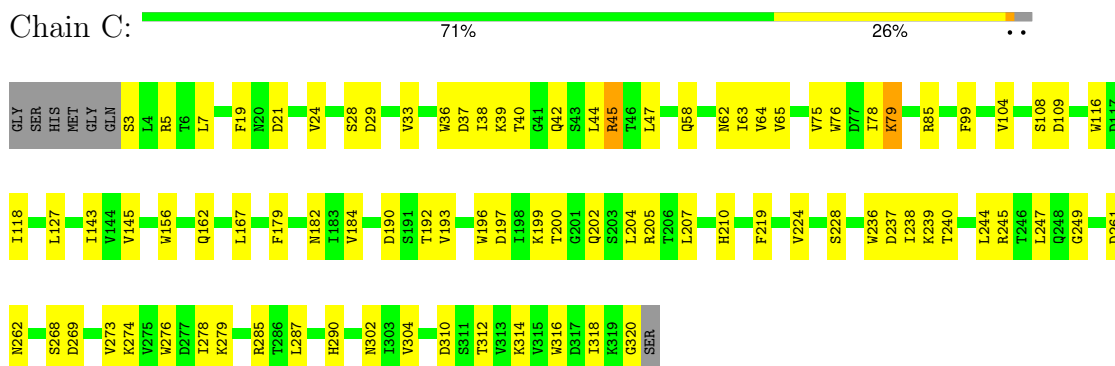
#### • Molecule 1: Tako8



#### • Molecule 1: Tako8



#### • Molecule 1: Tako8



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.55Å 132.30Å 81.02Å 90.00° 125.26° 90.00°	Depositor
Resolution (Å)	46.77 – 2.00 46.78 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.9 (46.77-2.00) 99.9 (46.78-2.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	111.59 (at 2.00Å)	Xtriage
Refinement program	PHENIX (1.13_2998: ???)	Depositor
R, $R_{free}$	0.243 , 0.266 0.284 , 0.307	Depositor DCC
$R_{free}$ test set	2485 reflections (4.59%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	45.4	Xtriage
Anisotropy	0.027	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 100.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.52$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	0.469 for -h,-h-2*k,1/2*h-1/2*k 0.468 for -h,h+2*k,1/2*h+1/2*k 0.468 for -1/2*h+1/2*k-l,3/2*h+1/2*k+l,1/2*h-1/2*k 0.469 for 1/2*h-1/2*k+l,-1/2*h+1/2*k+l,-h-1/2*h+1/2*k+l,1/2*h+1/2*k-l,-h-1/2*h-1/2*k-l,-3/2*h+1/2*k-l,1/2*h+1/2*k 0.469 for 1/2*h+1/2*k+l,3/2*h-1/2*k+l,-l 0.469 for 1/2*h-1/2*k+l,-3/2*h-1/2*k-l,-l 0.470 for -1/2*h+1/2*k-l,1/2*h-1/2*k-l,-1/2*h-1/2*k 0.469 for -1/2*h-1/2*k-l,-1/2*h-1/2*k+l,-1/2*h+1/2*k 0.469 for h,-k,-h-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	7475	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	57.0	wwPDB-VP

<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.

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.80% of the height of the origin peak. No significant pseudotranslation is detected.*

## 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	0.25	0/2456	0.51	0/3339
1	B	0.24	0/2447	0.51	0/3327
1	C	0.26	0/2447	0.51	0/3327
All	All	0.25	0/7350	0.51	0/9993

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	2425	0	2368	52	0
1	B	2416	0	2360	56	0
1	C	2416	0	2360	77	0
2	A	70	0	0	0	0
2	B	76	0	0	3	0
2	C	72	0	0	3	0
All	All	7475	0	7088	175	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 12.

The worst 5 of 175 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:119:LYS:HE3	1:C:62:ASN:HB3	1.56	0.86
1:A:224:VAL:HG23	1:A:238:ILE:HG22	1.60	0.82
1:B:88:GLN:O	1:B:116:TRP:HH2	1.64	0.80
1:B:277:ASP:HB2	1:B:284:LEU:HD21	1.65	0.77
1:C:78:ILE:HG22	1:C:79:LYS:HE3	1.64	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	317/325 (98%)	295 (93%)	20 (6%)	2 (1%)	25	19
1	B	316/325 (97%)	292 (92%)	23 (7%)	1 (0%)	41	37
1	C	316/325 (97%)	295 (93%)	21 (7%)	0	100	100
All	All	949/975 (97%)	882 (93%)	64 (7%)	3 (0%)	41	37

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	89	GLY
1	A	249	GLY
1	B	9	GLY

### 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	287/291 (99%)	285 (99%)	2 (1%)	84	88
1	B	286/291 (98%)	283 (99%)	3 (1%)	76	81
1	C	286/291 (98%)	283 (99%)	3 (1%)	76	81
All	All	859/873 (98%)	851 (99%)	8 (1%)	78	83

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

Mol	Chain	Res	Type
1	C	261	ASP
1	C	79	LYS
1	B	221	ASP
1	B	125	ARG
1	C	45	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	202	GLN
1	B	162	GLN
1	C	42	GLN
1	C	298	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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.