



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 08:54 am BST

PDB ID : 5XOU  
Title : Crystal structure of T. thermophilus Argonaute protein complexed with a bulge 7T8 on the guide strand  
Authors : Sheng, G.; Wang, J.; Zhao, H.; Wang, Y.  
Deposited on : 2017-05-31  
Resolution : 2.63 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

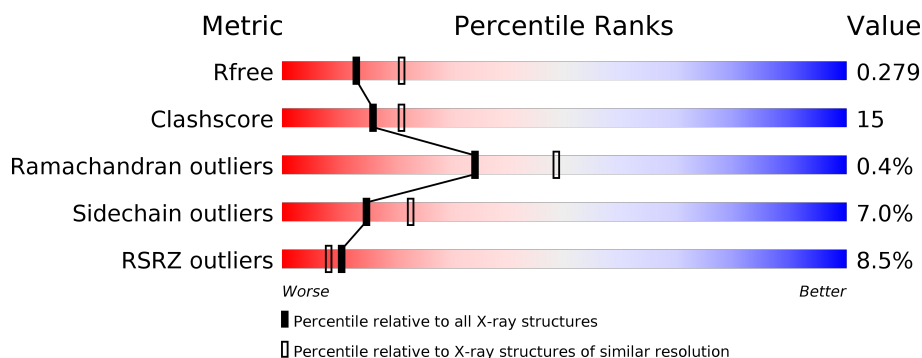
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.63 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	1426 (2.66-2.62)
Clashscore	141614	1472 (2.66-2.62)
Ramachandran outliers	138981	1446 (2.66-2.62)
Sidechain outliers	138945	1446 (2.66-2.62)
RSRZ outliers	127900	1408 (2.66-2.62)

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 on 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	685	<div> <div>8%</div> <div> <div></div> <div>57%</div> <div>20%</div> <div>• •</div> <div>18%</div> </div> </div>
1	B	685	<div> <div>7%</div> <div> <div></div> <div>59%</div> <div>18%</div> <div>5%</div> <div>•</div> <div>16%</div> </div> </div>
2	C	22	<div> <div>32%</div> <div>23%</div> <div>9%</div> <div>36%</div> </div>
2	E	22	<div> <div>23%</div> <div>36%</div> <div>14%</div> <div>27%</div> </div>
3	D	19	<div> <div>53%</div> <div>32%</div> <div>16%</div> </div>
3	F	19	<div> <div>37%</div> <div>47%</div> <div>16%</div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10096 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 TtAgo (D546N).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	564	Total	C	N	O	S	0	0	0
			4373	2799	819	751	4			
1	B	574	Total	C	N	O	S	0	0	0
			4435	2837	832	762	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	546	ASN	ASP	engineered mutation	UNP Q746M7
B	546	ASN	ASP	engineered mutation	UNP Q746M7

- Molecule 2 is a DNA chain called DNA (5'-D(P\*TP\*GP\*AP\*GP\*GP\*TP\*AP\*TP\*GP\*GP\*TP\*TP\*GP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	14	Total	C	N	O	P	0	0	0
			295	140	52	89	14			
2	E	16	Total	C	N	O	P	0	0	0
			338	160	62	100	16			

- Molecule 3 is a DNA chain called DNA (5'-D(\*AP\*CP\*AP\*AP\*CP\*CP\*TP\*AP\*CP\*TP\*AP\*CP\*CP\*TP\*CP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	16	Total	C	N	O	P	0	0	0
			317	153	57	92	15			
3	F	16	Total	C	N	O	P	0	0	0
			317	153	57	92	15			

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Mg 1 1	0	0
4	A	1	Total Mg 1 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	5	Total O 5 5	0	0
5	B	11	Total O 11 11	0	0
5	E	2	Total O 2 2	0	0
5	F	1	Total O 1 1	0	0

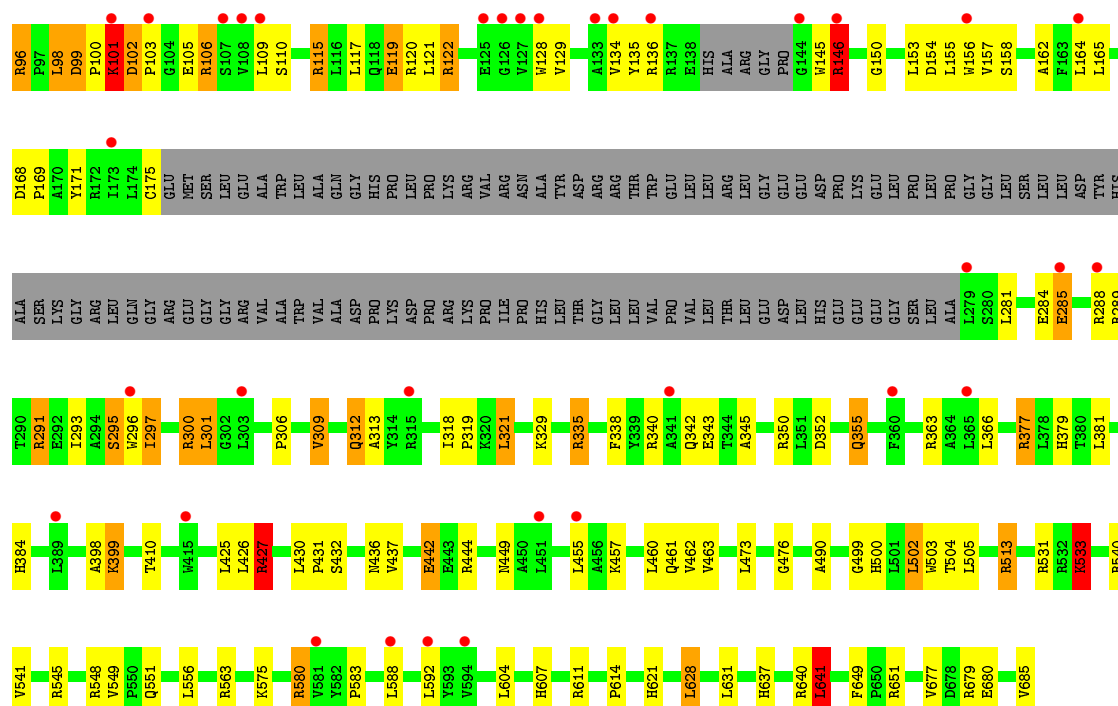
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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 ( $\text{RSRZ} > 2$ ). 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.

Chain A:

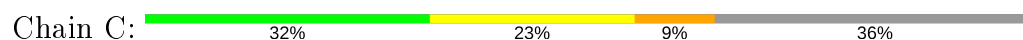
Chain B:

7% 59% 18% 5% 16%

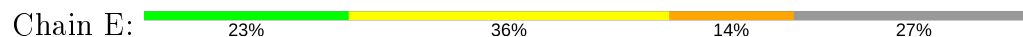
RET ASN HIS L4 L13 P18 L19 L20 L21 P21 E22 E23 L24 L25 P26 L27 K28 R28 L29 E30 V31 P35 E41 V42 Y43 P44 L45 V49 E52 A53 A55 L56 M60 A65 L73 W74 L75 L78 L79 A80 R81 R82 Y86 R89 L90 Y91 P92 K93 G94



- Molecule 2: DNA (5'-D(P\*TP\*GP\*AP\*GP\*GP\*TP\*AP\*TP\*GP\*GP\*TP\*TP\*GP\*T)-3')



- Molecule 2: DNA (5'-D(P\*TP\*GP\*AP\*GP\*GP\*TP\*AP\*TP\*GP\*GP\*TP\*TP\*GP\*T)-3')



- Molecule 3: DNA (5'-D(\*AP\*CP\*AP\*AP\*CP\*CP\*TP\*AP\*CP\*TP\*AP\*CP\*CP\*TP\*CP\*G)-3')



- Molecule 3: DNA (5'-D(\*AP\*CP\*AP\*AP\*CP\*CP\*TP\*AP\*CP\*TP\*AP\*CP\*CP\*TP\*CP\*G)-3')



DT	DA	DT	A4	C5	C9	T10	A11	C12	T13	A14	C15	C16	G19
----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	112.75Å 114.98Å 160.44Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.17 – 2.63 43.17 – 2.63	Depositor EDS
% Data completeness (in resolution range)	97.3 (43.17-2.63) 92.0 (43.17-2.63)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.47 (at 2.61Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.235 , 0.276 0.240 , 0.279	Depositor DCC
$R_{free}$ test set	3091 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	58.5	Xtriage
Anisotropy	0.156	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 25.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.038 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	10096	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.50% 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

### 5.1 Standard geometry

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

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	0.95	23/4474 (0.5%)	1.48	80/6073 (1.3%)
1	B	0.91	19/4537 (0.4%)	1.40	63/6158 (1.0%)
2	C	1.18	1/329 (0.3%)	1.25	3/504 (0.6%)
2	E	1.13	2/379 (0.5%)	1.14	3/584 (0.5%)
3	D	0.89	0/354	1.02	1/542 (0.2%)
3	F	0.84	0/354	1.03	0/542
All	All	0.95	45/10427 (0.4%)	1.40	150/14403 (1.0%)

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	3
1	B	0	4
All	All	0	7

The worst 5 of 45 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	309	VAL	CB-CG1	-14.16	1.23	1.52
1	A	309	VAL	CB-CG2	-10.81	1.30	1.52
1	B	309	VAL	CB-CG2	-9.87	1.32	1.52
1	A	300	ARG	CZ-NH1	-9.60	1.20	1.33
2	C	1	DT	OP3-P	-8.94	1.50	1.61

The worst 5 of 150 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	300	ARG	NE-CZ-NH1	-30.68	104.96	120.30
1	A	300	ARG	NE-CZ-NH2	22.17	131.39	120.30
1	B	300	ARG	NE-CZ-NH1	-19.97	110.32	120.30
1	B	335	ARG	NE-CZ-NH1	-18.67	110.97	120.30
1	A	95	ARG	NE-CZ-NH2	-18.21	111.19	120.30

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	312	GLN	Sidechain
1	A	39	ARG	Mainchain
1	A	99	ASP	Sidechain
1	B	285	GLU	Sidechain
1	B	312	GLN	Sidechain

## 5.2 Too-close contacts

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	4373	0	4421	137	0
1	B	4435	0	4461	132	1
2	C	295	0	162	5	0
2	E	338	0	183	11	0
3	D	317	0	181	7	0
3	F	317	0	181	6	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	5	0	0	0	0
5	B	11	0	0	2	0
5	E	2	0	0	0	0
5	F	1	0	0	0	0
All	All	10096	0	9589	284	1

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 15.

The worst 5 of 284 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:119:GLU:OE2	1:A:122:ARG:NH2	1.87	1.06
1:A:124:LEU:HD23	1:A:300:ARG:HD3	1.35	1.06
1:A:25:ARG:CZ	1:A:95:ARG:HH22	1.72	1.03
1:A:513:ARG:NH2	1:A:551:GLN:O	1.94	1.00
1:B:119:GLU:OE2	1:B:122:ARG:NH2	1.96	0.97

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:122:ARG:NH1	1:B:563:ARG:O[4_455]	1.93	0.27

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	558/685 (82%)	541 (97%)	14 (2%)	3 (0%)	29	43
1	B	568/685 (83%)	554 (98%)	12 (2%)	2 (0%)	34	48
All	All	1126/1370 (82%)	1095 (97%)	26 (2%)	5 (0%)	34	48

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	355	GLN
1	B	355	GLN
1	A	83	GLY
1	B	146	ARG
1	A	597	GLU

### 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	427/549 (78%)	396 (93%)	31 (7%)	14	21
1	B	429/549 (78%)	400 (93%)	29 (7%)	16	24
All	All	856/1098 (78%)	796 (93%)	60 (7%)	15	23

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	570	LEU
1	B	42	VAL
1	B	504	THR
1	A	641	LEU
1	B	78	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	312	GLN
1	A	500	HIS
1	B	500	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 carbohydrates in this entry.

## 5.6 Ligand geometry

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	564/685 (82%)	0.59	56 (9%)	7 5	36, 65, 118, 151	0
1	B	574/685 (83%)	0.54	46 (8%)	12 9	37, 67, 116, 137	0
2	C	14/22 (63%)	-0.35	0	100 100	58, 67, 104, 116	0
2	E	16/22 (72%)	-0.32	0	100 100	58, 73, 124, 131	0
3	D	16/19 (84%)	-0.58	0	100 100	66, 90, 110, 117	0
3	F	16/19 (84%)	-0.39	0	100 100	61, 93, 111, 116	0
All	All	1200/1452 (82%)	0.52	102 (8%)	10 8	36, 67, 117, 151	0

The worst 5 of 102 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	82	MET	8.8
1	B	81	ARG	7.4
1	B	103	PRO	6.3
1	B	136	ARG	6.2
1	A	86	TYR	5.4

### 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 carbohydrates 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(Å <sup>2</sup> )	Q<0.9
4	MG	A	701	1/1	0.97	0.25	63,63,63,63	0
4	MG	B	701	1/1	0.97	0.18	57,57,57,57	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.