



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 11:00 pm GMT

PDB ID : 3HO1
Title : Crystal structure of T. thermophilus Argonaute N546 mutant protein complexed with DNA guide strand and 12-nt RNA target strand
Authors : Wang, Y.; Li, H.; Sheng, G.; Patel, D.J.
Deposited on : 2009-06-01
Resolution : 2.60 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

<http://wwpdb.org/validation/2016/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)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

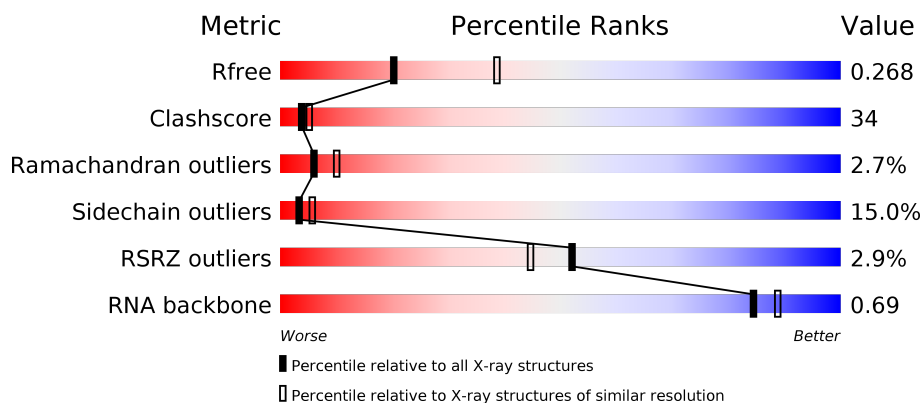
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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}	100719	2542 (2.60-2.60)
Clashscore	112137	2895 (2.60-2.60)
Ramachandran outliers	110173	2848 (2.60-2.60)
Sidechain outliers	110143	2848 (2.60-2.60)
RSRZ outliers	101464	2550 (2.60-2.60)
RNA backbone	2435	1140 (3.00-2.20)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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>3%</div> <div>49%</div> <div>41%</div> <div>8%</div> <div>•</div> </div>
2	X	21	<div> <div>10%</div> <div>5%</div> <div>62%</div> <div>5%</div> <div>29%</div> </div>
3	Y	12	<div> <div>25%</div> <div>67%</div> <div>8%</div> </div>

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
4	MG	X	22	-	-	-	X

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5767 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 argonaute.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	673	Total	C	N	O	S	0	0	0
			5145	3292	974	873	6			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	546	ASN	ASP	ENGINEERED	UNP Q746M7

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	X	15	Total	C	N	O	P	0	0	1
			299	140	58	87	14			

- Molecule 3 is a RNA chain called 5'-R(P*CP*CP*UP*AP*CP*UP*AP*CP*CP*UP*CP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	Y	12	Total	C	N	O	P	0	0	0
			229	101	34	82	12			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	X	1	Total	Mg	0	0
			1	1		

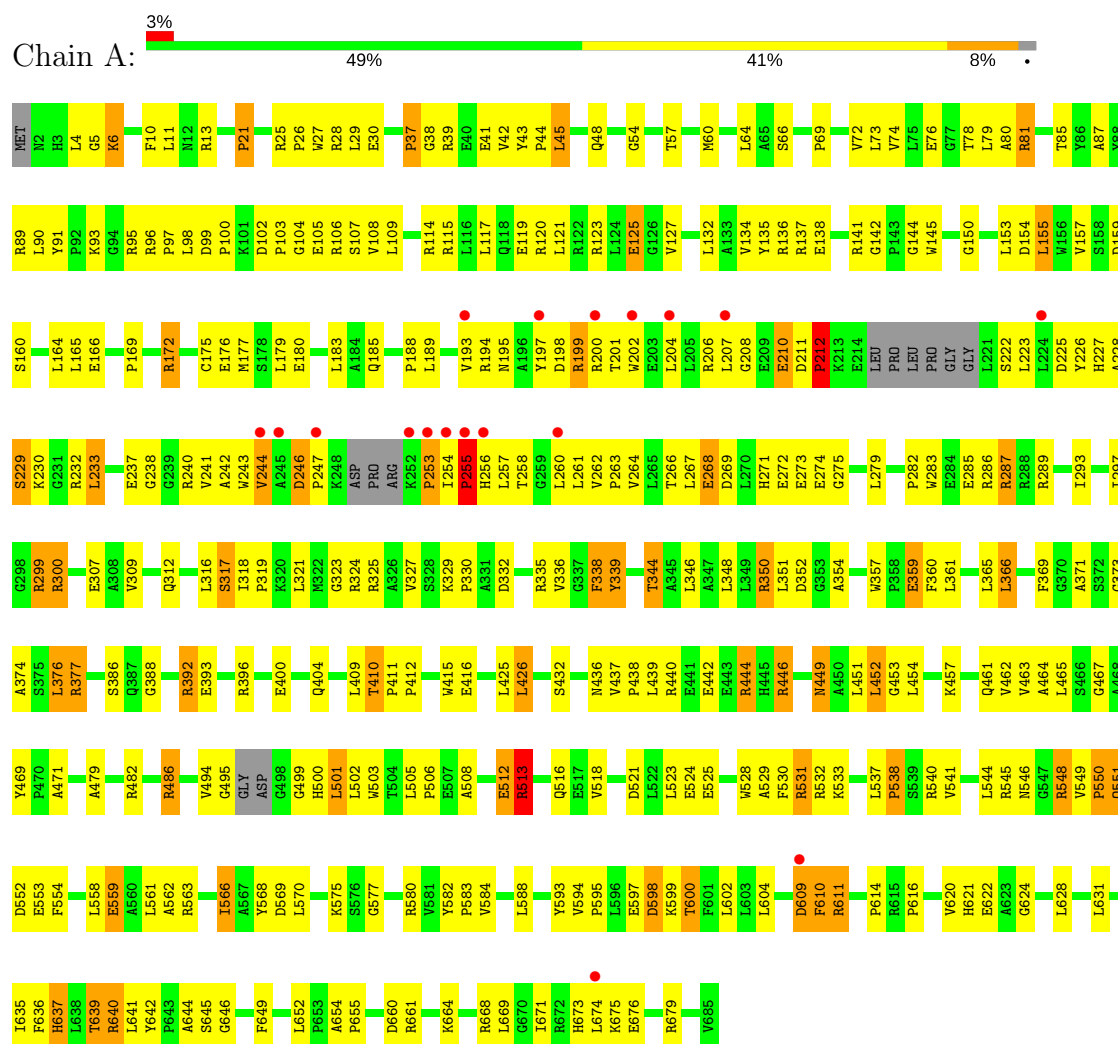
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	84	Total 84	O 84	0	0
5	X	4	Total 4	O 4	0	0
5	Y	5	Total 5	O 5	0	0

3 Residue-property plots

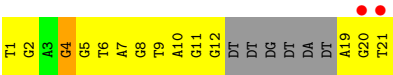
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are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

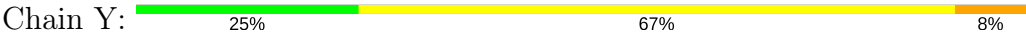


• Molecule 2: 5'-D(P*TP*GP*AP*GP*GP*TP*AP*GP*TP*AP*GP*GP*TP*TP*GP*TP*AP*T*AP*GP*T)-3'





● Molecule 3: 5'-R(P*CP*CP*UP*AP*CP*UP*AP*CP*CP*UP*CP*G)-3'



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	122.93Å 118.53Å 59.53Å 90.00° 114.91° 90.00°	Depositor
Resolution (Å)	30.00 – 2.60 40.61 – 2.61	Depositor EDS
% Data completeness (in resolution range)	95.0 (30.00-2.60) 95.7 (40.61-2.61)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.37 (at 2.61Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.220 , 0.258 0.238 , 0.268	Depositor DCC
R_{free} test set	1546 reflections (6.83%)	DCC
Wilson B-factor (Å ²)	52.4	Xtriage
Anisotropy	0.232	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 45.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	5767	wwPDB-VP
Average B, all atoms (Å ²)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.01% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.48	1/5265 (0.0%)	0.82	10/7159 (0.1%)
2	X	0.90	2/335 (0.6%)	0.98	2/515 (0.4%)
3	Y	0.67	1/252 (0.4%)	0.73	0/387
All	All	0.52	4/5852 (0.1%)	0.82	12/8061 (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
2	X	0	1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	X	1	DT	OP3-P	-10.21	1.48	1.61
2	X	19	DA	O3'-P	-7.02	1.52	1.61
3	Y	1	C	OP3-P	-6.36	1.53	1.61
1	A	513	ARG	C-O	-5.10	1.13	1.23

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	X	1	DT	OP1-P-OP2	-6.57	109.75	119.60
1	A	246	ASP	O-C-N	-6.53	108.70	121.10
1	A	253	PRO	N-CA-CB	6.10	110.62	103.30
1	A	189	LEU	N-CA-C	5.98	127.14	111.00
1	A	273	GLU	C-N-CA	5.73	136.02	121.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	X	4	DG	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	5145	0	5122	347	0
2	X	299	0	160	28	0
3	Y	229	0	118	11	0
4	X	1	0	0	0	0
5	A	84	0	0	9	0
5	X	4	0	0	0	0
5	Y	5	0	0	0	0
All	All	5767	0	5400	372	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 34.

The worst 5 of 372 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:210:GLU:HG3	1:A:243:TRP:HH2	1.09	1.08
1:A:210:GLU:HG3	1:A:243:TRP:CH2	1.90	1.07
1:A:153:LEU:HD13	1:A:297:ILE:HD12	1.37	1.01
1:A:671:ILE:HG21	1:A:674:LEU:HG	1.43	1.00
1:A:671:ILE:CG2	1:A:674:LEU:HG	1.90	1.00

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	665/685 (97%)	583 (88%)	64 (10%)	18 (3%)	6	10

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	5	GLY
1	A	188	PRO
1	A	246	ASP
1	A	247	PRO
1	A	339	TYR

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	488/549 (89%)	415 (85%)	73 (15%)	3	6

5 of 73 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	377	ARG
1	A	449	ASN
1	A	637	HIS
1	A	426	LEU
1	A	452	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such

sidechains are listed below:

Mol	Chain	Res	Type
1	A	404	GLN
1	A	500	HIS
1	A	551	GLN
1	A	342	GLN
1	A	546	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	Y	10/12 (83%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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 [i](#)

Of 1 ligands modelled in this entry, 1 is 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 [i](#)

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, 95th 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 ²)	Q<0.9
1	A	673/685 (98%)	-0.02	18 (2%) 55 48	28, 59, 99, 114	1 (0%)
2	X	15/21 (71%)	0.14	2 (13%) 4 2	35, 48, 126, 127	0
3	Y	12/12 (100%)	-0.67	0 100 100	57, 65, 79, 88	0
All	All	700/718 (97%)	-0.03	20 (2%) 52 45	28, 59, 99, 127	1 (0%)

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	197	TYR	4.2
1	A	253	PRO	4.2
1	A	245	ALA	4.1
1	A	247	PRO	3.9
1	A	244	VAL	3.6

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	LLDF	B-factors(\AA^2)	Q<0.9
4	MG	X	22	1/1	0.91	0.23	2.78	32,32,32,32	0

6.5 Other polymers [i](#)

There are no such residues in this entry.