



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 09:18 AM GMT

PDB ID : 3HXM  
Title : Structure of an argonaute complexed with guide DNA and target RNA duplex containing two mismatches.  
Authors : Wang, Y.; Li, H.; Sheng, G.; Patel, D.J.  
Deposited on : 2009-06-21  
Resolution : 3.10 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  
<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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

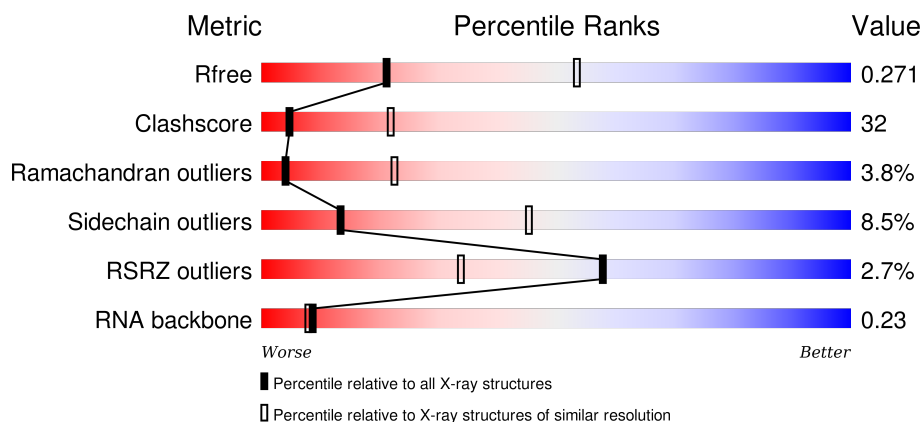
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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))
$R_{free}$	91344	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)
RNA backbone	2183	1010 (3.52-2.68)

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. 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>2%</div> <div>47%</div> <div>41%</div> <div>6%</div> <div>7%</div> </div>
2	C	21	<div> <div>10%</div> <div>14%</div> <div>33%</div> <div>10%</div> <div>5%</div> <div>38%</div> </div>
3	Y	20	<div> <div>5%</div> <div>10%</div> <div>15%</div> <div>20%</div> <div>55%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5225 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	637	Total	C	N	O	S	2	0	0
			4779	3053	887	833	6			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	13	Total	C	N	O	P	0	0	0
			258	120	48	77	13			

- Molecule 3 is a RNA chain called RNA (5'-R(\*UP\*AP\*UP\*AP\*CP\*AP\*AP\*CP\*UP\*CP\*AP\*CP\*UP\*AP\*CP\*CP\*UP\*CP\*GP\*U)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	Y	9	Total	C	N	O	P	0	0	0
			187	84	31	63	9			

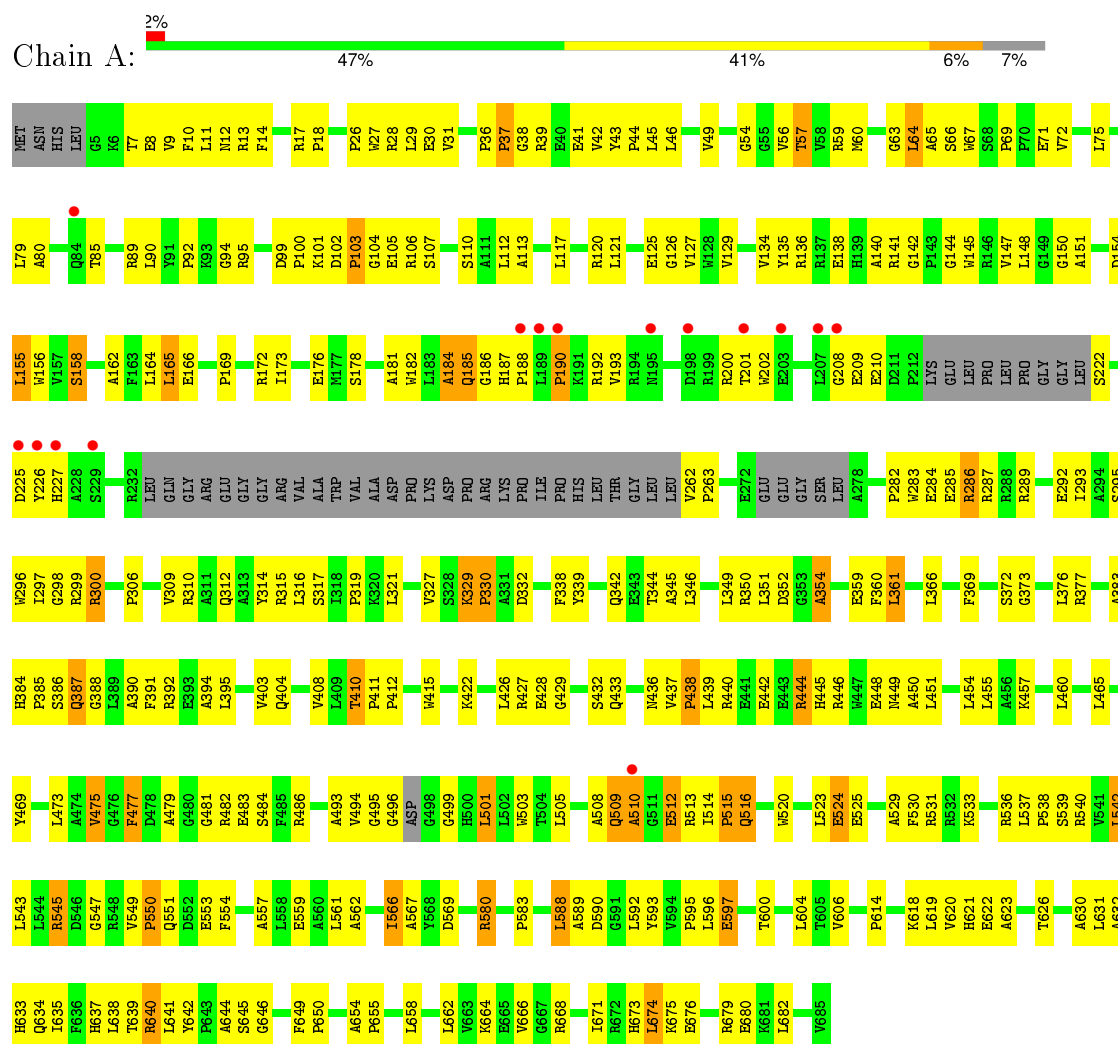
- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

### 3 Residue-property plots

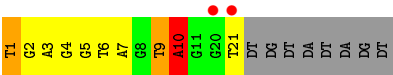
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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

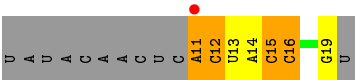


#### • Molecule 2: DNA (5'-D(P\*TP\*GP\*AP\*GP\*GP\*TP\*AP\*GP\*TP\*AP\*GP\*GP\*TP\*TP\*GP\*T P\*AP\*TP\*AP\*GP\*T)-3')





● Molecule 3: RNA (5'-R(\*UP\*AP\*UP\*AP\*CP\*AP\*AP\*CP\*UP\*CP\*AP\*CP\*UP\*AP\*CP\*CP\*UP\*CP\*GP\*U)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	121.84Å 118.87Å 59.19Å 90.00° 114.79° 90.00°	Depositor
Resolution (Å)	30.00 – 3.10 40.49 – 3.10	Depositor EDS
% Data completeness (in resolution range)	94.3 (30.00-3.10) 94.2 (40.49-3.10)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.37 (at 3.12Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.221 , 0.276 0.220 , 0.271	Depositor DCC
$R_{free}$ test set	980 reflections (8.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	62.8	Xtriage
Anisotropy	0.399	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 61.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 13951 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	5225	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.69% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.43	0/4888	0.75	4/6654 (0.1%)
2	C	1.02	1/288 (0.3%)	1.25	3/441 (0.7%)
3	Y	0.60	0/207	0.84	0/319
All	All	0.49	1/5383 (0.0%)	0.79	7/7414 (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	C	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	1	DT	OP3-P	-6.91	1.52	1.61

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	103	PRO	CA-N-CD	-8.77	99.23	111.50
2	C	10	DA	N9-C1'-C2'	-6.74	99.80	112.60
2	C	9	DT	O4'-C1'-N1	6.34	112.44	108.00
1	A	481	GLY	N-CA-C	-5.85	98.48	113.10
1	A	190	PRO	N-CA-CB	5.85	110.31	103.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	10	DA	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	4779	0	4658	301	0
2	C	258	0	137	24	0
3	Y	187	0	98	10	0
4	C	1	0	0	0	0
All	All	5225	0	4893	320	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 32.

The worst 5 of 320 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:103:PRO:HD2	1:A:104:GLY:H	1.10	1.11
1:A:327:VAL:HG23	1:A:332:ASP:HB2	1.30	1.10
1:A:531:ARG:HH11	1:A:537:LEU:HG	1.09	1.08
1:A:531:ARG:NH1	1:A:537:LEU:HG	1.77	0.98
1:A:121:LEU:HD22	1:A:134:VAL:HG21	1.47	0.96

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	627/685 (92%)	521 (83%)	82 (13%)	24 (4%)	4	22

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	39	ARG
1	A	184	ALA
1	A	190	PRO
1	A	508	ALA
1	A	509	GLN

### 5.3.2 Protein sidechains ⓘ

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	445/549 (81%)	407 (92%)	38 (8%)	13	45

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	346	LEU
1	A	444	ARG
1	A	580	ARG
1	A	366	LEU
1	A	451	LEU

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

Mol	Chain	Res	Type
1	A	449	ASN
1	A	673	HIS
1	A	634	GLN
1	A	404	GLN
1	A	607	HIS

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	Y	9/20 (45%)	3 (33%)	1 (11%)

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	Y	12	C
3	Y	15	C
3	Y	16	C

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
3	Y	11	A

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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 ⓘ

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	637/685 (92%)	-0.10	15 (2%) 62 39	38, 68, 113, 127	2 (0%)
2	C	13/21 (61%)	0.82	2 (15%) 3 1	48, 59, 130, 135	2 (15%)
3	Y	9/20 (45%)	0.31	1 (11%) 7 2	78, 90, 126, 132	0
All	All	659/726 (90%)	-0.08	18 (2%) 58 34	38, 69, 114, 135	4 (0%)

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	20	DG	5.9
2	C	21	DT	5.6
1	A	189	LEU	4.0
1	A	208	GLY	3.8
1	A	510	ALA	3.3

### 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 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, 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	MG	C	30	1/1	0.84	0.28	1.62	15,15,15,15	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.