



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 08:15 am BST

PDB ID : 5JS2  
Title : Human Argonaute-2 Bound to a Modified siRNA  
Authors : Schirle, N.T.; MacRae, I.J.  
Deposited on : 2016-05-07  
Resolution : 2.95 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
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

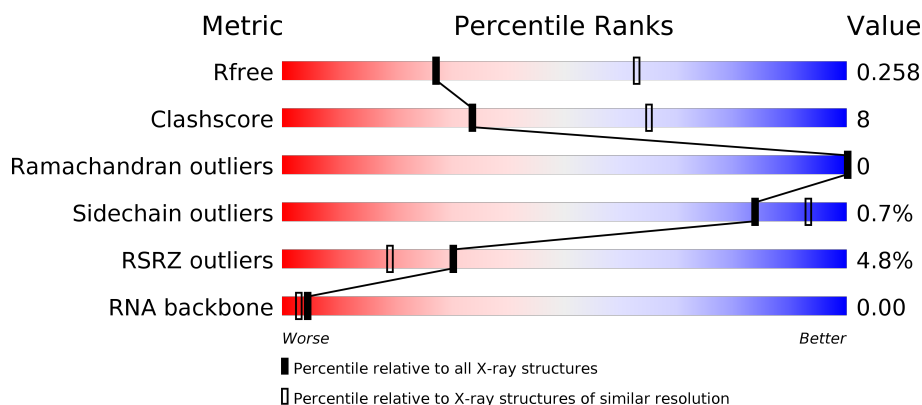
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.95 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)
RNA backbone	3102	1065 (3.22-2.70)

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	859	<div> <div>4%</div> <div>78%</div> <div>15%</div> <div>7%</div> </div>
2	B	6	<div> <div>17%</div> <div>83%</div> </div>

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 6593 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 Protein argonaute-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	800	Total	C	N	O	S	0	0	0
			6432	4095	1156	1142	39			

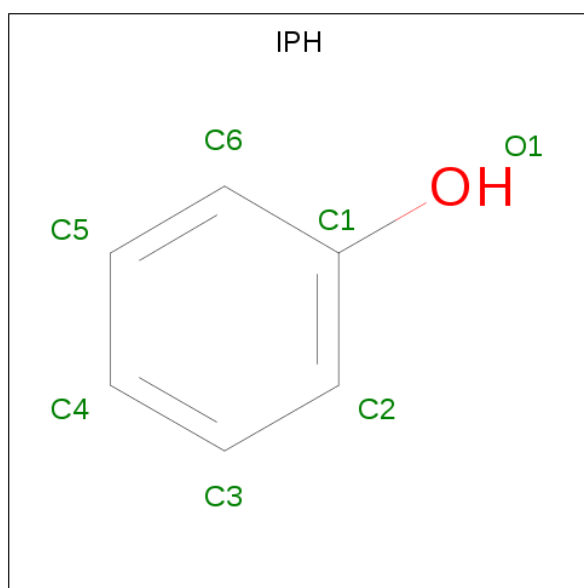
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	387	ASP	SER	engineered mutation	UNP Q9UKV8

- Molecule 2 is a RNA chain called modified siRNA.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
2	B	6	Total	C	F	N	O	P	S	0	0	0
			130	62	3	16	40	6	3			

- Molecule 3 is PHENOL (three-letter code: IPH) (formula: C<sub>6</sub>H<sub>6</sub>O).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			7	6	1		
3	A	1	Total	C	O	0	0
			7	6	1		

- Molecule 4 is PHOSPHATE ION (three-letter code: PO4) (formula: O<sub>4</sub>P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	O	P	0	0
			5	4	1		

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Mg	0	0
			1	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	11	Total	O	0	0
			11	11		



- Molecule 1: Protein argonaute-2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	63.10Å 107.55Å 68.89Å 90.00° 107.39° 90.00°	Depositor
Resolution (Å)	38.98 – 2.95 38.98 – 2.95	Depositor EDS
% Data completeness (in resolution range)	98.1 (38.98-2.95) 98.1 (38.98-2.95)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.77 (at 2.95Å)	Xtriage
Refinement program	PHENIX (1.10.1 _2155: ???)	Depositor
R, $R_{free}$	0.222 , 0.261 0.220 , 0.258	Depositor DCC
$R_{free}$ test set	916 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	63.3	Xtriage
Anisotropy	0.169	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 43.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	6593	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.88% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: MG, IPH, 6OO, 6NW, 6OP, F2T, UFT, PO4

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	0.34	0/6582	0.46	1/8907 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	111	ASP	CB-CG-OD2	5.22	123.00	118.30

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	6432	0	6487	104	0
2	B	130	0	18	9	0
3	A	14	0	12	0	0
4	A	5	0	0	0	0
5	A	1	0	0	0	0
6	A	11	0	0	1	0
All	All	6593	0	6517	111	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 8.

The worst 5 of 111 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)
2:B:902:F2T:SP1	2:B:902:F2T:P	1.52	1.49
1:A:251:THR:HB	1:A:254:GLN:HG3	1.21	1.15
1:A:251:THR:HG22	1:A:253:SER:H	1.19	1.04
1:A:389:ASN:ND2	6:A:1001:HOH:O	1.91	1.03
1:A:251:THR:HB	1:A:254:GLN:CG	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	786/859 (92%)	757 (96%)	29 (4%)	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	712/752 (95%)	707 (99%)	5 (1%)	84	93

All (5) residues with a non-rotameric sidechain are listed below:



Mol	Chain	Res	Type
1	A	252	ASP
1	A	254	GLN
1	A	297	GLN
1	A	581	GLN
1	A	804	TYR

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

Mol	Chain	Res	Type
1	A	246	GLN
1	A	297	GLN
1	A	334	GLN
1	A	581	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	0/6	-	-

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 ⓘ

6 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	UFT	B	906	2	14,21,22	6.55	11 (78%)	14,30,33	2.02	3 (21%)
2	6NW	B	903	2	18,25,26	4.36	6 (33%)	18,36,39	1.98	4 (22%)
2	6OO	B	905	2	15,22,23	4.60	9 (60%)	17,31,34	1.68	2 (11%)
2	6OP	B	901	2	21,27,27	4.69	7 (33%)	23,39,39	2.37	6 (26%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	UFT	B	904	2	14,21,22	4.78	9 (64%)	14,30,33	1.51	2 (14%)
2	F2T	B	902	2	14,21,22	7.64	6 (42%)	14,30,33	1.37	2 (14%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	UFT	B	906	2	-	4/5/25/26	0/2/2/2
2	6NW	B	903	2	-	0/4/27/28	0/3/3/3
2	6OO	B	905	2	-	2/6/27/28	0/2/2/2
2	6OP	B	901	2	-	7/10/31/31	0/2/2/2
2	UFT	B	904	2	-	0/5/25/26	0/2/2/2
2	F2T	B	902	2	-	2/4/25/26	0/2/2/2

The worst 5 of 48 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	902	F2T	C2'-C1'	-21.97	1.25	1.53
2	B	906	UFT	C2'-C3'	-18.04	1.27	1.52
2	B	901	6OP	C3'-C4'	-17.26	1.23	1.53
2	B	902	F2T	O4'-C1'	14.00	1.60	1.41
2	B	903	6NW	O4'-C1'	13.73	1.60	1.41

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	901	6OP	C4'-C5'-C6'	-7.74	109.05	124.02
2	B	903	6NW	N3-C2-N1	-6.21	118.97	128.68
2	B	906	UFT	O4'-C1'-C2'	5.39	111.35	105.79
2	B	901	6OP	C4-N3-C2	5.38	119.69	115.14
2	B	905	6OO	C2-N3-C4	4.58	120.98	116.34

There are no chirality outliers.

5 of 15 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	906	UFT	C2'-C1'-N1-C6
2	B	906	UFT	O4'-C1'-N1-C6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms
2	B	901	6OP	C3'-C4'-C5'-C6'
2	B	901	6OP	O4'-C1'-N1-C6
2	B	901	6OP	C2'-C1'-N1-C6

There are no ring outliers.

5 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	906	UFT	1	0
2	B	903	6NW	2	0
2	B	905	6OO	2	0
2	B	901	6OP	1	0
2	B	902	F2T	4	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	IPH	A	901	-	7,7,7	0.36	0	8,8,8	0.24	0
3	IPH	A	902	-	7,7,7	0.37	0	8,8,8	0.25	0
4	PO4	A	903	-	4,4,4	0.91	0	6,6,6	0.42	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	IPH	A	901	-	-	-	0/1/1/1
3	IPH	A	902	-	-	-	0/1/1/1

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 [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	800/859 (93%)	0.23	38 (4%) 30 19	30, 58, 97, 140	0
2	B	0/6	-	-	-	-
All	All	800/865 (92%)	0.23	38 (4%) 30 19	30, 58, 97, 140	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	449	LYS	4.5
1	A	448	ILE	3.9
1	A	513	ALA	3.9
1	A	455	CYS	3.8
1	A	485	ILE	3.7

### 6.2 Non-standard residues in protein, DNA, RNA chains [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
2	UFT	B	906	20/21	0.83	0.23	74,96,102,103	0
2	6OO	B	905	21/22	0.92	0.17	48,80,84,87	0
2	6OP	B	901	26/26	0.94	0.15	55,75,81,84	0
2	F2T	B	902	20/21	0.94	0.15	62,69,85,100	0
2	UFT	B	904	20/21	0.96	0.20	51,66,72,75	0
2	6NW	B	903	23/24	0.97	0.18	52,63,69,71	0

### 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( $\text{\AA}^2$ )	Q<0.9
5	MG	A	904	1/1	0.78	0.18	68,68,68,68	0
4	PO4	A	903	5/5	0.93	0.18	70,73,85,95	0
3	IPH	A	901	7/7	0.94	0.28	52,55,59,64	0
3	IPH	A	902	7/7	0.97	0.14	48,52,55,64	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.