



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 30, 2021 – 12:40 PM EDT

PDB ID : 7KI3  
Title : Human Argonaute2:miR-122 bound to the HCV genotype 1a site-1 RNA  
Authors : Gebert, L.F.R.; MacRae, I.J.  
Deposited on : 2020-10-22  
Resolution : 3.00 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 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	:	2.23.2
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)	:	2.23.2

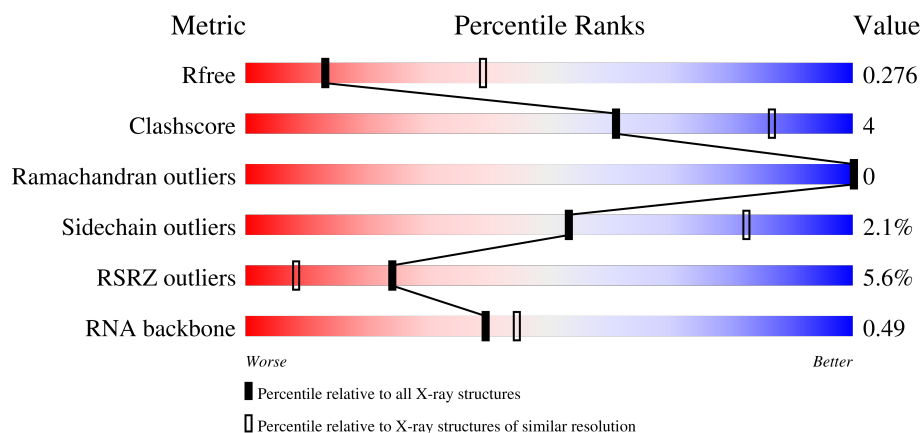
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)
RNA backbone	3102	1173 (3.30-2.70)

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 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  $\leq 5\%$ . 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>2%</div> <div>84%</div> <div>9%</div> <div>6%</div> </div>
1	D	859	<div> <div>8%</div> <div>79%</div> <div>12%</div> <div>9%</div> </div>
2	B	23	<div> <div>9%</div> <div>43%</div> <div>26%</div> <div>13%</div> <div>17%</div> </div>
2	E	23	<div> <div>26%</div> <div>17%</div> <div>•</div> <div>52%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	C	29	<div><div></div><div>48%</div><div>41%</div><div>10%</div></div>
3	F	29	<div><div></div><div>38%</div><div>21%</div><div>14%</div><div>28%</div></div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 14417 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	804	Total	C	N	O	S	0	0	0
			6433	4091	1159	1142	41			
1	D	783	Total	C	N	O	S	0	0	0
			6266	3987	1125	1114	40			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	387	ASP	SER	engineered mutation	UNP Q9UKV8
A	824	ALA	SER	engineered mutation	UNP Q9UKV8
A	828	ASP	SER	engineered mutation	UNP Q9UKV8
A	831	ASP	SER	engineered mutation	UNP Q9UKV8
A	834	ALA	SER	engineered mutation	UNP Q9UKV8
D	387	ASP	SER	engineered mutation	UNP Q9UKV8
D	824	ALA	SER	engineered mutation	UNP Q9UKV8
D	828	ASP	SER	engineered mutation	UNP Q9UKV8
D	831	ASP	SER	engineered mutation	UNP Q9UKV8
D	834	ALA	SER	engineered mutation	UNP Q9UKV8

- Molecule 2 is a RNA chain called miR-122.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	19	Total	C	N	O	P	0	0	0
			413	183	75	136	19			
2	E	11	Total	C	N	O	P	0	0	0
			241	106	43	81	11			

- Molecule 3 is a RNA chain called HCV genotype 1a miR-122 site-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	29	Total	C	N	O	P	0	0	0
			614	275	112	199	28			

*Continued on next page...*

*Continued from previous page...*

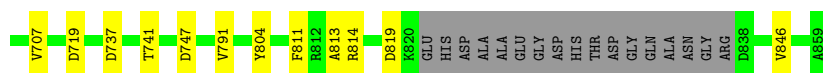
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	21	Total	C	N	O	P	0	0	0
			447	199	82	145	21			

- Molecule 4 is BARIUM ION (three-letter code: BA) (formula: Ba).

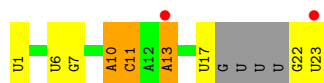
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Ba	0	0
			1	1		
4	C	1	Total	Ba	0	0
			1	1		
4	D	1	Total	Ba	0	0
			1	1		

- Molecule 1: Protein argonaute-2





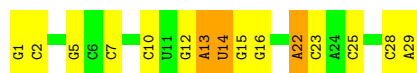
- Molecule 2: miR-122



- Molecule 2: miR-122



- Molecule 3: HCV genotype 1a miR-122 site-1



- Molecule 3: HCV genotype 1a miR-122 site-1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	87.31Å 112.96Å 207.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	99.23 – 3.00 103.84 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.9 (99.23-3.00) 89.3 (103.84-2.70)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.74 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.14rc1_3161	Depositor
R, $R_{free}$	0.233 , 0.276 0.233 , 0.276	Depositor DCC
$R_{free}$ test set	2847 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	45.1	Xtriage
Anisotropy	0.267	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 41.9	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	0.88	EDS
Total number of atoms	14417	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.16% 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: BA

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	0.31	0/6585	0.50	0/8911
1	D	0.32	0/6406	0.50	1/8656 (0.0%)
2	B	0.55	1/461 (0.2%)	0.79	0/714
2	E	0.73	1/268 (0.4%)	0.83	0/413
3	C	0.39	0/685	0.89	2/1066 (0.2%)
3	F	0.25	0/497	0.85	0/770
All	All	0.34	2/14902 (0.0%)	0.57	3/20530 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	1	U	OP3-P	-10.63	1.48	1.61
2	B	1	U	OP3-P	-10.36	1.48	1.61

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	10	C	C6-N1-C2	7.10	123.14	120.30
1	D	420	TYR	CB-CA-C	-5.27	99.86	110.40
3	C	16	G	C8-N9-C4	5.20	108.48	106.40

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	6433	0	6492	52	2
1	D	6266	0	6310	60	2
2	B	413	0	205	5	0
2	E	241	0	119	3	0
3	C	614	0	318	11	0
3	F	447	0	232	4	0
4	A	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
All	All	14417	0	13676	125	2

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 4.

The worst 5 of 125 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:D:289:ALA:HA	1:D:308:VAL:HG23	1.49	0.94
1:A:64:GLU:O	3:C:1:G:C6	2.37	0.76
2:B:11:C:N4	3:C:5:G:O2'	2.22	0.73
1:A:179:ARG:NH1	2:B:13:A:N3	2.39	0.70
1:A:630:ARG:NH1	1:A:641:ASP:OD2	2.25	0.69

All (2) 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:A:188:CYS:SG	1:D:188:CYS:SG[2_654]	1.64	0.56
1:A:854:ARG:NE	1:D:148:SER:O[1_455]	2.02	0.18

## 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	796/859 (93%)	752 (94%)	44 (6%)	0	100	100
1	D	761/859 (89%)	715 (94%)	46 (6%)	0	100	100
All	All	1557/1718 (91%)	1467 (94%)	90 (6%)	0	100	100

There are no Ramachandran outliers to report.

### 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	709/750 (94%)	699 (99%)	10 (1%)	67	88
1	D	691/750 (92%)	671 (97%)	20 (3%)	42	76
All	All	1400/1500 (93%)	1370 (98%)	30 (2%)	53	82

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	274	GLN
1	D	707	VAL
1	D	329	GLN
1	D	819	ASP
1	D	462	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	18/23 (78%)	3 (16%)	2 (11%)
2	E	10/23 (43%)	1 (10%)	1 (10%)
3	C	28/29 (96%)	5 (17%)	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	F	19/29 (65%)	7 (36%)	2 (10%)
All	All	75/104 (72%)	16 (21%)	5 (6%)

5 of 16 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	11	C
2	B	13	A
2	B	23	U
3	C	12	G
3	C	13	A

All (5) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	10	A
2	B	22	G
2	E	22	G
3	F	16	G
3	F	17	C

## 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](#)

Of 3 ligands modelled in this entry, 3 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 [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	804/859 (93%)	-0.11	19 (2%) 59 30	18, 45, 107, 141	0
1	D	783/859 (91%)	0.17	73 (9%) 8 3	18, 49, 150, 172	0
2	B	19/23 (82%)	0.46	2 (10%) 6 2	31, 72, 153, 155	0
2	E	11/23 (47%)	-0.19	0 100 100	29, 33, 142, 148	0
3	C	29/29 (100%)	0.04	0 100 100	44, 91, 151, 169	0
3	F	21/29 (72%)	0.09	0 100 100	43, 86, 130, 144	0
All	All	1667/1822 (91%)	0.04	94 (5%) 24 8	18, 47, 143, 172	0

The worst 5 of 94 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	240	PHE	7.0
1	D	328	LEU	6.5
1	D	258	PHE	6.4
1	D	130	VAL	6.3
1	D	279	TYR	6.3

### 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 monosaccharides 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	BA	C	101	1/1	0.89	0.07	128,128,128,128	0
4	BA	A	901	1/1	0.99	0.15	54,54,54,54	0
4	BA	D	901	1/1	0.99	0.14	69,69,69,69	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.