Full wwPDB X-ray Structure Validation Report (i)

May 23, 2020 – 06:06 am BST

PDB ID	:	6RA4
Title	:	Human ARGONAUTE-2 PAZ DOMAIN (214-347) IN COMPLEX WITH CGUGACUCU
Authors	:	Rondeau, J.-M.; Bourgier, E.
Deposited on	:	2019-04-05
Resolution	:	1.90 Å(reported)

The following versions of software and data (see references (i)) 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 (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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.
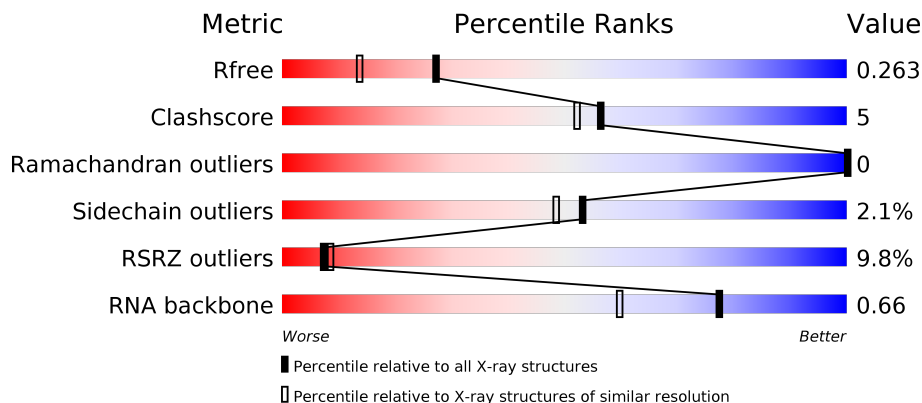
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)
RNA backbone	3102	1013 (2.42-1.38)

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 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 <=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	136	10% / 77% / 16% / 7%
1	B	136	10% / 83% / 12% / ··
2	L	9	11% / 56% / 11% / 33%
2	M	9	67% / 33%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2590 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	127	Total	C	N	O	S	0	0	0
			1039	664	186	181	8			
1	B	130	Total	C	N	O	S	0	0	0
			1061	677	191	185	8			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	212	GLY	-	expression tag	UNP Q9UKV8
A	213	PRO	-	expression tag	UNP Q9UKV8
B	212	GLY	-	expression tag	UNP Q9UKV8
B	213	PRO	-	expression tag	UNP Q9UKV8

- Molecule 2 is a RNA chain called RNA (5'-R(*CP*GP*UP*GP*AP*CP*UP*CP*U)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	L	9	Total	C	N	O	P	0	0	0
			185	84	30	63	8			
2	M	9	Total	C	N	O	P	0	0	0
			185	84	30	63	8			

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	L	1	Total 6	C 3	O 3	0	0
3	M	1	Total 6	C 3	O 3	0	0

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	39	Total 39	O 39	0	0
4	B	27	Total 27	O 27	0	0
4	L	16	Total 16	O 16	0	0
4	M	26	Total 26	O 26	0	0

3 Residue-property plots (i)

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.

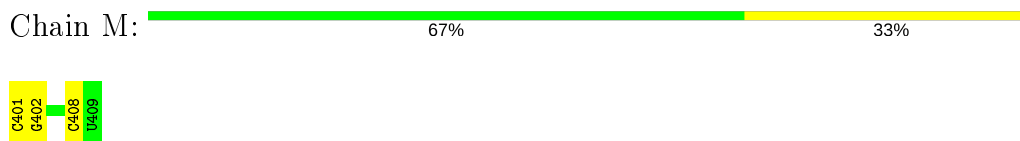- Molecule 1: Protein argonaute-2
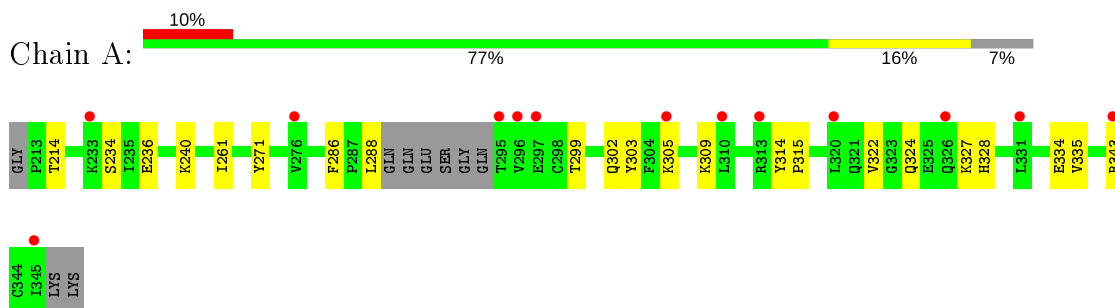
- Molecule 1: Protein argonaute-2
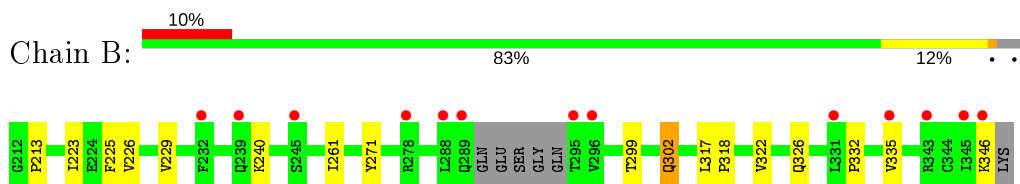
- Molecule 2: RNA (5'-R(*CP*GP*UP*GP*AP*CP*UP*CP*U)-3')

- Molecule 2: RNA (5'-R(*CP*GP*UP*GP*AP*CP*UP*CP*U)-3')

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	72.62Å 87.67Å 56.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.89 – 1.90 28.89 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.0 (28.89-1.90) 99.1 (28.89-1.90)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.23 (at 1.91Å)	Xtriage
Refinement program	CNS 2002	Depositor
R, R_{free}	0.239 , 0.268 0.232 , 0.263	Depositor DCC
R_{free} test set	1425 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	37.2	Xtriage
Anisotropy	0.252	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 47.2	EDS
L-test for twinning[2]	$< |L| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2590	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.14% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: GOL

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	0.59	0/1061	0.66	0/1430
1	B	0.54	0/1083	0.61	0/1459
2	L	0.77	0/205	0.92	1/317 (0.3%)
2	M	0.79	0/205	0.81	0/317
All	All	0.60	0/2554	0.69	1/3523 (0.0%)

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
2	L	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	L	404	G	O4'-C1'-N9	5.17	112.34	108.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	L	404	G	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	1039	0	1067	14	0
1	B	1061	0	1090	8	0
2	L	185	0	98	2	0
2	M	185	0	98	0	0
3	L	6	0	8	1	0
3	M	6	0	8	0	0
4	A	39	0	0	1	0
4	B	27	0	0	0	0
4	L	16	0	0	1	0
4	M	26	0	0	0	0
All	All	2590	0	2369	24	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 5.

All (24) 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:234:SER:HB2	1:A:236:GLU:OE1	1.87	0.75
1:B:299:THR:OG1	1:B:302:GLN:HB2	1.89	0.71
1:A:324:GLN:HG3	1:A:327:LYS:HE3	1.75	0.68
1:A:314:TYR:HD1	1:A:334:GLU:HG3	1.65	0.60
1:A:271:TYR:HB3	1:A:322:VAL:HB	1.84	0.59
1:B:223:ILE:O	1:B:226:VAL:HG12	2.10	0.51
1:B:261:ILE:HA	1:B:335:VAL:O	2.10	0.51
1:A:324:GLN:HG2	4:L:607:HOH:O	2.12	0.49
1:A:328:HIS:CD2	2:L:409:U:O2	2.66	0.49
1:B:225:PHE:O	1:B:229:VAL:HG23	2.13	0.48
1:B:213:PRO:HD2	1:B:346:LYS:NZ	2.28	0.48
1:A:286:PHE:HZ	1:A:303:TYR:CD1	2.33	0.46
4:A:403:HOH:O	3:L:501:GOL:H31	2.15	0.46
1:A:299:THR:OG1	1:A:302:GLN:HB2	2.16	0.46
1:B:332:PRO:O	1:B:335:VAL:HG22	2.17	0.44
1:B:271:TYR:HB3	1:B:322:VAL:HB	1.99	0.44
1:A:261:ILE:HA	1:A:335:VAL:O	2.18	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:302:GLN:O	1:A:305:LYS:HB3	2.20	0.42
1:A:343:ARG:HE	1:A:343:ARG:HB2	1.69	0.41
1:B:317:LEU:HB3	1:B:318:PRO:HD2	2.02	0.41
1:A:214:THR:HG23	1:A:343:ARG:HG2	2.03	0.41
1:A:286:PHE:HE2	1:A:288:LEU:HD13	1.86	0.41
1:A:314:TYR:N	1:A:315:PRO:HD3	2.36	0.40
2:L:406:C:O3'	2:L:407:U:H4'	2.21	0.40

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	123/136 (90%)	117 (95%)	6 (5%)	0	100	100
1	B	126/136 (93%)	125 (99%)	1 (1%)	0	100	100
All	All	249/272 (92%)	242 (97%)	7 (3%)	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	118/125 (94%)	116 (98%)	2 (2%)	60	57

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	120/125 (96%)	117 (98%)	3 (2%)	47	41
All	All	238/250 (95%)	233 (98%)	5 (2%)	53	48

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

Mol	Chain	Res	Type
1	A	240	LYS
1	A	309	LYS
1	B	240	LYS
1	B	302	GLN
1	B	326	GLN

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	328	HIS
1	B	283	HIS
1	B	324	GLN
1	B	326	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	L	8/9 (88%)	1 (12%)	1 (12%)
2	M	9/9 (100%)	2 (22%)	1 (11%)
All	All	17/18 (94%)	3 (17%)	2 (11%)

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	L	407	U
2	M	402	G
2	M	408	C

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	L	406	C
2	M	401	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

2 ligands 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	#$\|Z\| > 2$
3	GOL	L	501	-	5,5,5	0.76	0	5,5,5	0.68	0
3	GOL	M	501	-	5,5,5	0.76	0	5,5,5	0.38	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	GOL	L	501	-	-	3/4/4/4	-
3	GOL	M	501	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	L	501	GOL	O1-C1-C2-C3
3	L	501	GOL	O1-C1-C2-O2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	L	501	GOL	O2-C2-C3-O3

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	L	501	GOL	1	0

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, 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	127/136 (93%)	0.86	13 (10%)	6	8	28, 41, 66, 84	0
1	B	130/136 (95%)	0.81	13 (10%)	7	8	28, 44, 65, 74	0
2	L	9/9 (100%)	0.32	1 (11%)	5	6	29, 37, 71, 91	0
2	M	9/9 (100%)	-0.01	0	100	100	30, 37, 49, 74	0
All	All	275/290 (94%)	0.79	27 (9%)	7	8	28, 42, 67, 91	0

All (27) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	295	THR	9.7
1	A	296	VAL	8.3
1	A	233	LYS	6.2
1	B	346	LYS	4.9
1	B	296	VAL	3.9
1	B	289	GLN	3.8
1	B	295	THR	3.6
1	B	331	LEU	3.5
1	A	326	GLN	3.3
1	A	305	LYS	3.2
1	B	245	SER	3.0
2	L	407	U	2.9
1	A	345	ILE	2.6
1	B	345	ILE	2.6
1	A	343	ARG	2.6
1	B	288	LEU	2.6
1	B	343	ARG	2.5
1	B	239	GLN	2.4
1	A	331	LEU	2.4
1	A	313	ARG	2.4
1	A	297	GLU	2.3

Continued on next page...

Full wwPDB X-ray Structure Validation Report

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	310	LEU	2.3
1	A	320	LEU	2.1
1	B	335	VAL	2.1
1	A	276	VAL	2.1
1	B	232	PHE	2.0
1	B	278	ARG	2.0

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, 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	B-factors(Å^2)	Q<0.9
3	GOL	L	501	6/6	0.83	0.26	52,58,60,63	0
3	GOL	M	501	6/6	0.89	0.22	48,52,52,53	0

6.5 Other polymers (i)

There are no such residues in this entry.