wwPDB X-ray Structure Validation Summary Report (i)

Jun 16, 2024 – 05:39 AM EDT

PDB ID	:	2GJE
Title	:	Structure of a guideRNA-binding protein complex bound to a gRNA
Authors	:	Schumacher, M.A.; Karamooz, E.; Zikova, A.; Trantirek, L.; Lukes, J.
Deposited on	:	2006-03-30
Resolution	:	3.37 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

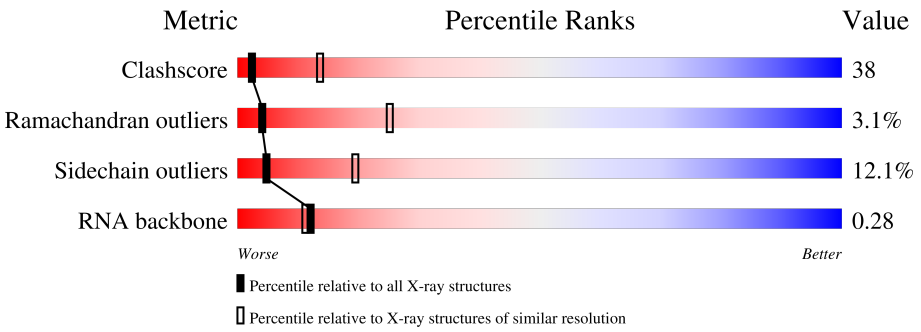
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.37 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))
Clashscore	141614	1762 (3.46-3.30)
Ramachandran outliers	138981	1732 (3.46-3.30)
Sidechain outliers	138945	1731 (3.46-3.30)
RNA backbone	3102	1001 (3.80-2.96)

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 >=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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	R	39	8% 15% 5% 5% 67%
2	S	5	20% 80%
3	A	195	27% 41% 7% • 26%
4	D	187	37% 35% 7% 21%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2769 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 RNA chain called guide RNA 40-mer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	R	13	272	124	49	87	12	0	0	0

- Molecule 2 is a RNA chain called RNA tetramer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	S	5	105	48	19	34	4	0	0	0

- Molecule 3 is a protein called mitochondrial RNA-binding protein 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
3	A	145	1187	752	216	215	1	3	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	MSE	MET	MODIFIED RESIDUE	UNP Q952G2
A	125	MSE	MET	MODIFIED RESIDUE	UNP Q952G2
A	204	MSE	MET	MODIFIED RESIDUE	UNP Q952G2

- Molecule 4 is a protein called mitochondrial RNA-binding protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
4	D	147	1204	759	222	219	1	3	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	43	GLU	LEU	CONFLICT	UNP P90629

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	46	MSE	MET	MODIFIED RESIDUE	UNP P90629
D	109	MSE	MET	MODIFIED RESIDUE	UNP P90629
D	139	MSE	MET	MODIFIED RESIDUE	UNP P90629

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	D	1	1	1	0	0

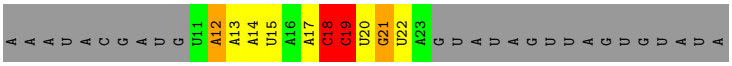
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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. 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. 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.

Note EDS was not executed.

• Molecule 1: guide RNA 40-mer

Chain R: 8% 15% 5% 5% 67%

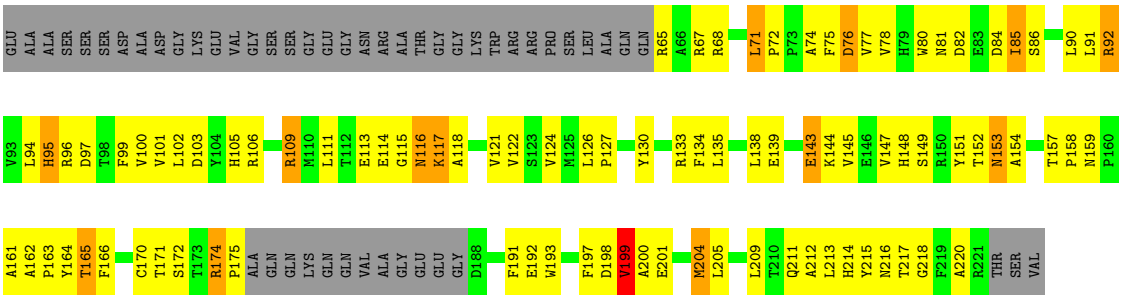


• Molecule 2: RNA tetramer

Chain S: 20% 80%

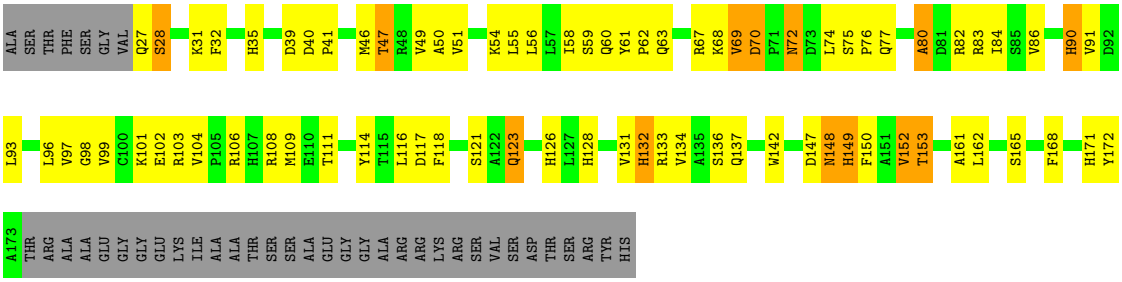
• Molecule 3: mitochondrial RNA-binding protein 2

Chain A: 27% 41% 7% • 26%



• Molecule 4: mitochondrial RNA-binding protein 1

Chain D: 37% 35% 7% 21%



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, α, β, γ	157.60Å 157.60Å 81.30Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	69.85 – 3.37	Depositor
% Data completeness (in resolution range)	91.8 (69.85-3.37)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.282 , 0.300	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2769	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

5 Model quality (i)

5.1 Standard geometry (i)

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	R	0.57	0/304	1.86	5/471 (1.1%)
2	S	0.75	0/117	1.20	1/181 (0.6%)
3	A	0.59	0/1213	0.88	2/1643 (0.1%)
4	D	0.54	0/1231	0.78	1/1657 (0.1%)
All	All	0.57	0/2865	1.03	9/3952 (0.2%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	R	19	C	O5'-P-OP2	-20.36	86.27	110.70
1	R	18	C	OP2-P-O3'	-18.18	65.21	105.20
1	R	18	C	OP1-P-O3'	-16.19	69.58	105.20
1	R	19	C	O5'-P-OP1	-15.56	91.69	105.70
2	S	40	U	O4'-C1'-N1	7.93	114.54	108.20

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	R	272	0	141	21	0
2	S	105	0	55	7	0
3	A	1187	0	1158	112	0
4	D	1204	0	1172	77	2

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	D	1	0	0	0	0
All	All	2769	0	2526	201	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 38.

The worst 5 of 201 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)
3:A:117:LYS:HD2	3:A:117:LYS:H	1.11	1.10
3:A:77:VAL:HG21	3:A:197:PHE:HE1	1.27	0.98
3:A:174:ARG:HH21	3:A:175:PRO:HG2	1.28	0.94
3:A:77:VAL:HG22	3:A:201:GLU:HG2	1.53	0.91
1:R:18:C:OP2	3:A:117:LYS:NZ	2.07	0.87

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)
4:D:67:ARG:NH2	4:D:67:ARG:NH2[8_556]	1.89	0.31
4:D:67:ARG:NH1	4:D:67:ARG:NH1[8_556]	1.91	0.29

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
3	A	141/195 (72%)	121 (86%)	16 (11%)	4 (3%)	5 27
4	D	145/187 (78%)	124 (86%)	16 (11%)	5 (3%)	3 23
All	All	286/382 (75%)	245 (86%)	32 (11%)	9 (3%)	4 25

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	161	ALA
4	D	28	SER
3	A	85	ILE
3	A	199	VAL
4	D	148	ASN

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
3	A	127/159 (80%)	112 (88%)	15 (12%)	5 20
4	D	130/155 (84%)	114 (88%)	16 (12%)	4 19
All	All	257/314 (82%)	226 (88%)	31 (12%)	5 19

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	204	MSE
4	D	132	HIS
4	D	47	THR
4	D	149	HIS
4	D	102	GLU

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

Mol	Chain	Res	Type
4	D	171	HIS
4	D	149	HIS
4	D	60	GLN
3	A	216	ASN
4	D	128	HIS

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	R	12/39 (30%)	6 (50%)	0
2	S	4/5 (80%)	0	0
All	All	16/44 (36%)	6 (37%)	0

5 of 6 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	R	12	A
1	R	15	U
1	R	17	A
1	R	18	C
1	R	19	C

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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)

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.