



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 3, 2024 – 03:18 PM JST

PDB ID : 8KAH  
Title : Crystal structure of SpyCas9-crRNA-tracrRNA complex bound to 18nt target DNA  
Authors : Chen, Y.; Chen, J.; Liu, L.  
Deposited on : 2023-08-03  
Resolution : 3.36 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.36.2  
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) : 2.36.2

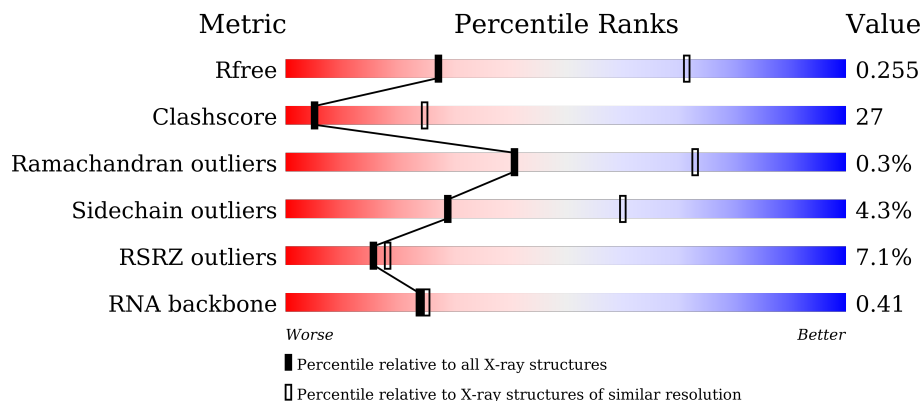
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.36 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	1558 (3.42-3.30)
Clashscore	141614	1627 (3.42-3.30)
Ramachandran outliers	138981	1599 (3.42-3.30)
Sidechain outliers	138945	1598 (3.42-3.30)
RSRZ outliers	127900	1507 (3.42-3.30)
RNA backbone	3102	1023 (3.80-2.92)

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	34	
1	E	34	
2	B	1368	
2	F	1368	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	C	26	
3	G	26	
4	D	11	
4	H	11	
5	I	65	
5	J	65	

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 27025 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 RNA (34-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	34	Total 725	C 325	N 127	O 239	P 34	0	0	0
1	E	32	Total 685	C 307	N 123	O 223	P 32	0	0	0

- Molecule 2 is a protein called CRISPR-associated endonuclease Cas9/Csn1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1322	Total 10732	C 6827	N 1864	O 2019	S 22	0	0	0
2	F	1327	Total 10695	C 6814	N 1845	O 2014	S 22	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	10	ALA	ASP	engineered mutation	UNP Q99ZW2
B	840	ALA	HIS	engineered mutation	UNP Q99ZW2
F	10	ALA	ASP	engineered mutation	UNP Q99ZW2
F	840	ALA	HIS	engineered mutation	UNP Q99ZW2

- Molecule 3 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	26	Total 521	C 254	N 85	O 157	P 25	0	0	0
3	G	26	Total 521	C 254	N 85	O 157	P 25	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(\*TP\*TP\*TP\*AP\*GP\*GP\*TP\*AP\*TP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	11	Total 225	110	37	68	10	0	0	0
4	H	11	Total 225	110	37	68	10	0	0	0

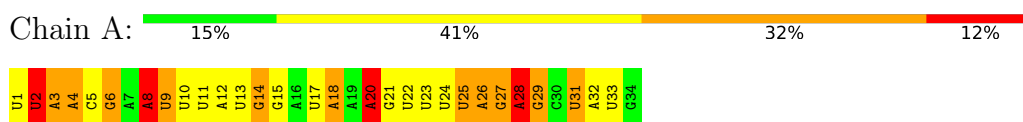
- Molecule 5 is a RNA chain called RNA (65-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	I	63	Total 1348	603	245	437	63	0	0	0
5	J	63	Total 1348	603	245	437	63	0	0	0

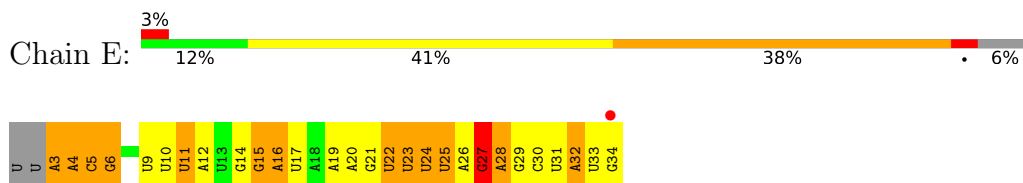
### 3 Residue-property plots [i](#)

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 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.

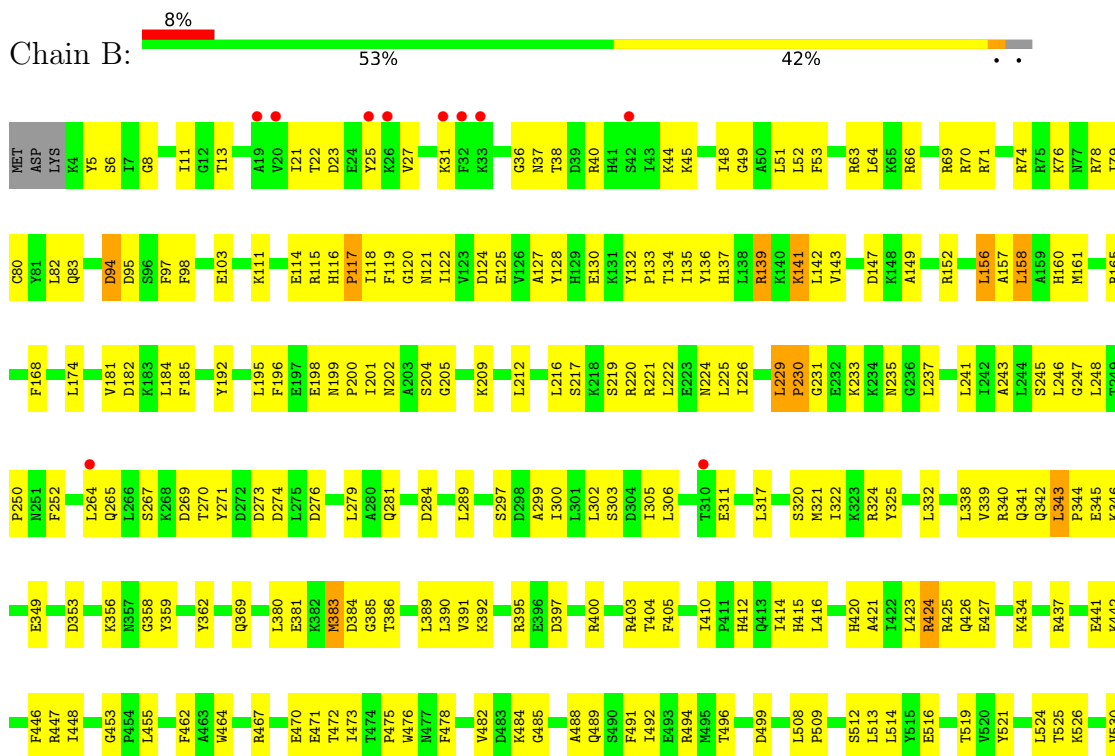
- Molecule 1: RNA (34-MER)

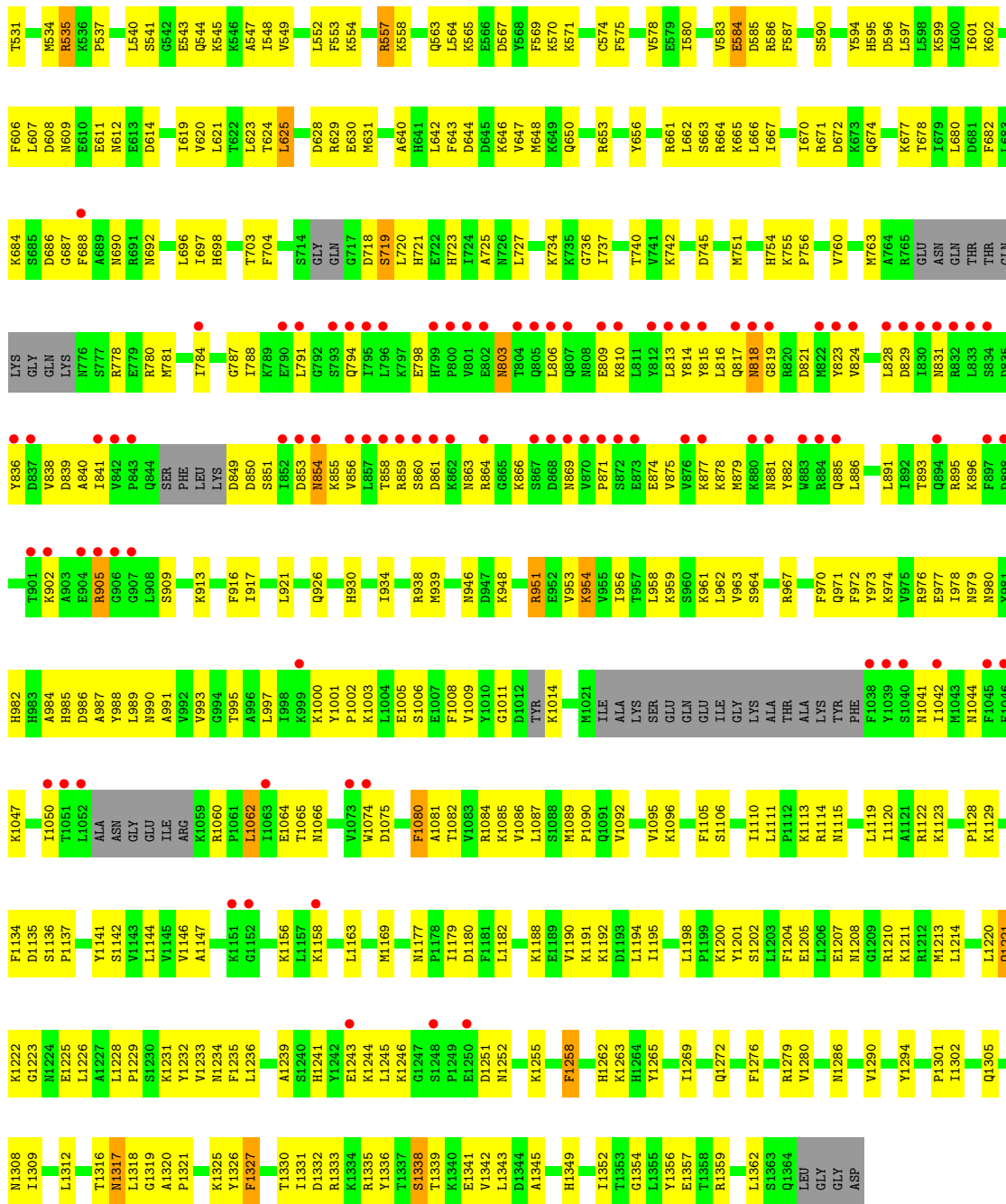


- Molecule 1: RNA (34-MER)

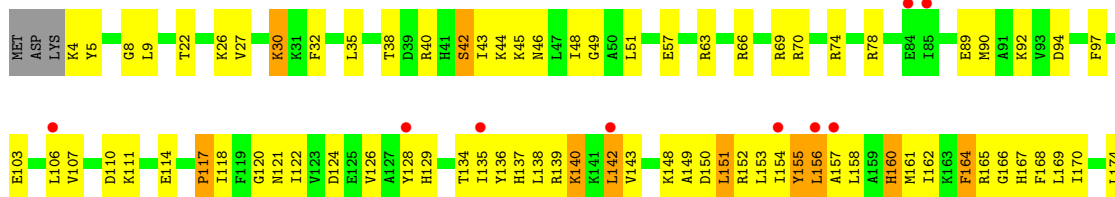


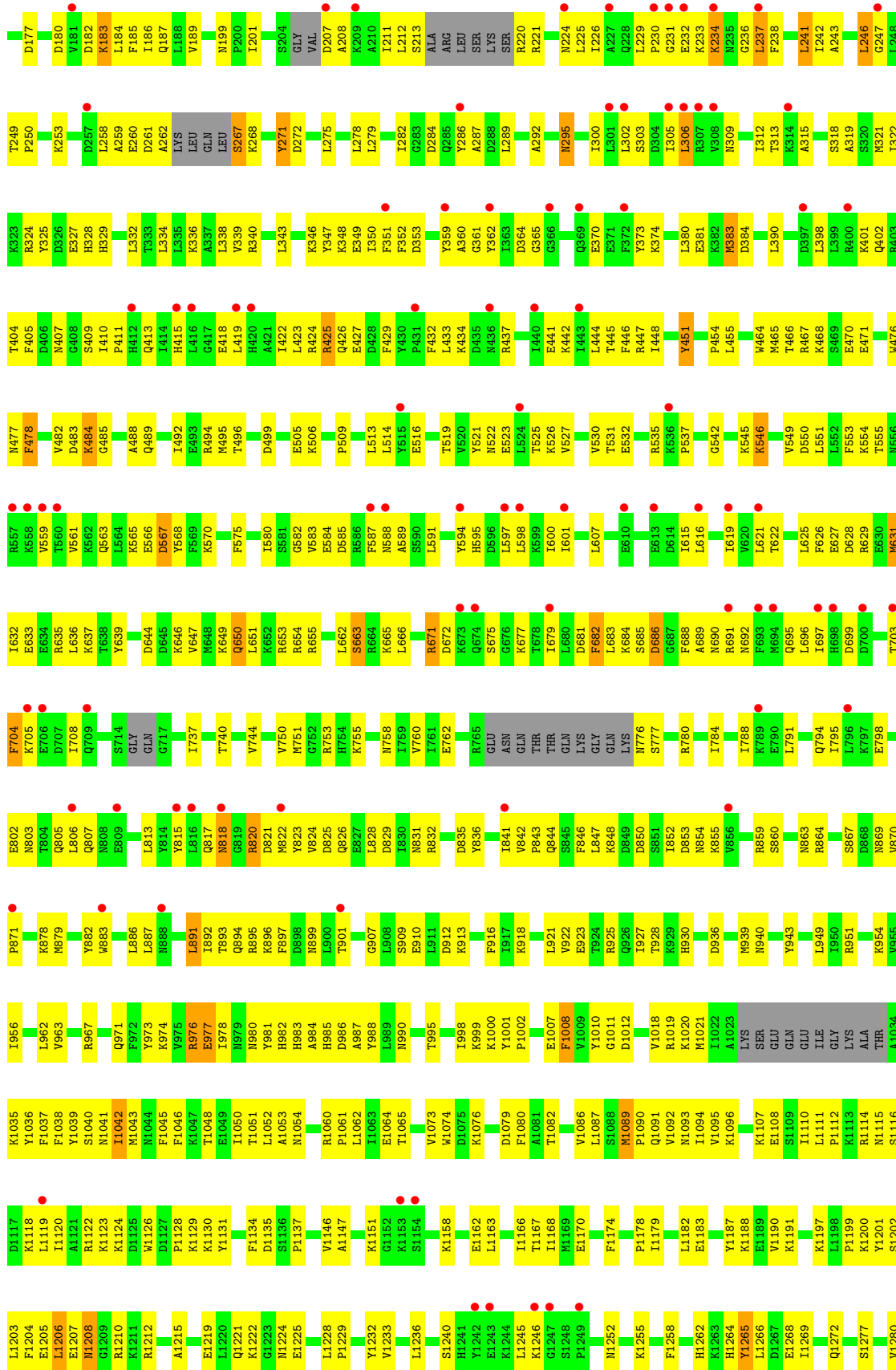
- Molecule 2: CRISPR-associated endonuclease Cas9/Csn1



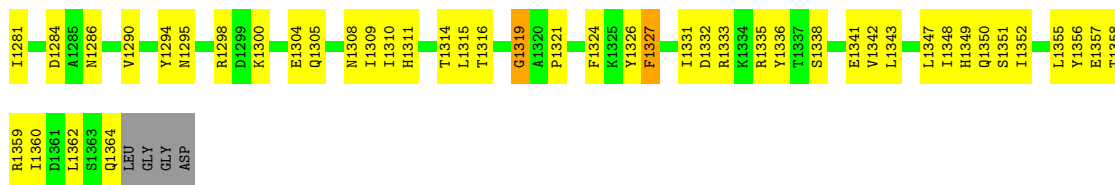


● Molecule 2: CRISPR-associated endonuclease Cas9/Csn1

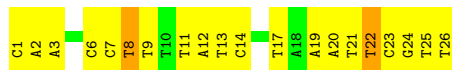
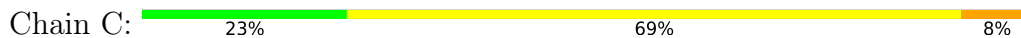




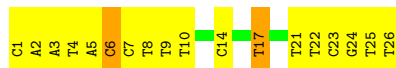




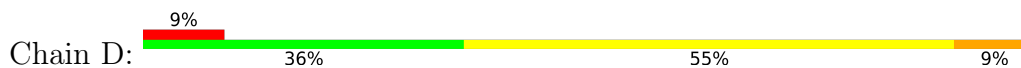
- Molecule 3: DNA (26-MER)



- Molecule 3: DNA (26-MER)



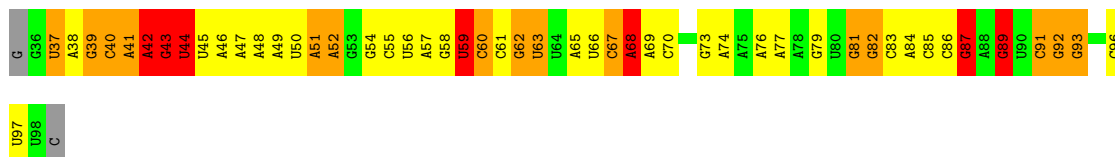
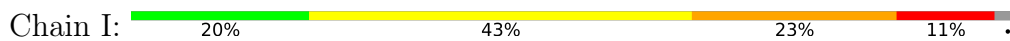
- Molecule 4: DNA (5'-D(\*TP\*TP\*TP\*AP\*GP\*GP\*TP\*AP\*TP\*TP\*G)-3')



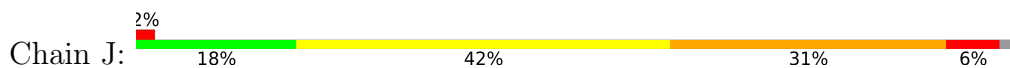
- Molecule 4: DNA (5'-D(\*TP\*TP\*TP\*AP\*GP\*GP\*TP\*AP\*TP\*TP\*G)-3')

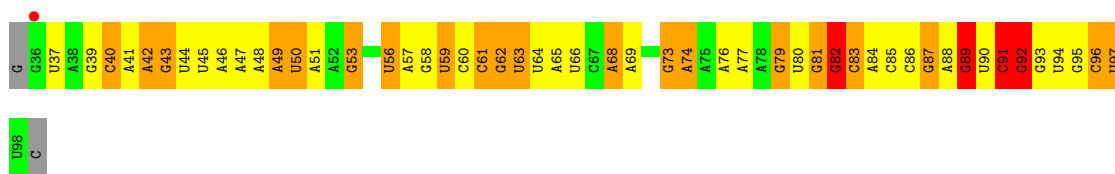


- Molecule 5: RNA (65-MER)



- Molecule 5: RNA (65-MER)





## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	145.21Å 130.60Å 148.73Å 90.00° 104.08° 90.00°	Depositor
Resolution (Å)	48.23 – 3.36 48.23 – 3.29	Depositor EDS
% Data completeness (in resolution range)	59.4 (48.23-3.36) 73.5 (48.23-3.29)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.34 (at 3.33Å)	Xtrriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, $R_{free}$	0.230 , 0.255 0.230 , 0.255	Depositor DCC
$R_{free}$ test set	1990 reflections (3.44%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	56.7	Xtrriage
Anisotropy	0.207	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 17.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtrriage
Estimated twinning fraction	0.069 for l,-k,h	Xtrriage
$F_o, F_c$ correlation	0.83	EDS
Total number of atoms	27025	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	81.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.53% 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](#)

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	1.01	1/811 (0.1%)	1.81	32/1261 (2.5%)
1	E	0.87	0/767	1.73	23/1193 (1.9%)
2	B	0.58	1/10915 (0.0%)	0.80	11/14673 (0.1%)
2	F	0.58	2/10879 (0.0%)	0.78	4/14636 (0.0%)
3	C	1.42	6/581 (1.0%)	1.36	6/893 (0.7%)
3	G	1.25	1/581 (0.2%)	1.34	4/893 (0.4%)
4	D	1.51	1/251 (0.4%)	1.41	3/387 (0.8%)
4	H	1.49	1/251 (0.4%)	1.36	2/387 (0.5%)
5	I	1.03	1/1509 (0.1%)	1.90	61/2350 (2.6%)
5	J	0.93	0/1509	1.80	55/2350 (2.3%)
All	All	0.73	14/28054 (0.0%)	1.10	201/39023 (0.5%)

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	5	DA	C3'-O3'	-10.11	1.30	1.44
4	H	3	DT	C1'-N1	6.85	1.58	1.49
3	C	8	DT	C3'-O3'	-6.36	1.35	1.44
5	I	41	A	N9-C4	-6.20	1.34	1.37
2	F	1319	GLY	C-N	-5.73	1.20	1.34

The worst 5 of 201 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	J	91	C	C4-C5-C6	-9.82	112.49	117.40
5	J	76	A	C8-N9-C4	-9.81	101.88	105.80
3	C	14	DC	O4'-C4'-C3'	-9.62	100.23	106.00
5	J	92	G	C5-C6-O6	-9.36	122.99	128.60
5	I	59	U	O5'-P-OP2	-9.32	97.31	105.70

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts

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	725	0	362	40	0
1	E	685	0	342	44	0
2	B	10732	0	10828	517	1
2	F	10695	0	10736	702	0
3	C	521	0	299	19	0
3	G	521	0	299	20	0
4	D	225	0	129	3	0
4	H	225	0	129	10	0
5	I	1348	0	678	38	0
5	J	1348	0	678	61	0
All	All	27025	0	24480	1373	1

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 27.

The worst 5 of 1373 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:F:142:LEU:CD1	2:F:154:ILE:HG12	1.37	1.54
2:F:142:LEU:HD11	2:F:154:ILE:CG1	1.39	1.51
2:F:142:LEU:CD2	2:F:154:ILE:HD11	1.41	1.48
2:F:142:LEU:CD2	2:F:154:ILE:CD1	1.94	1.45
2:B:1222:LYS:O	2:B:1318:LEU:CD2	1.66	1.42

All (1) 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)
2:B:202:ASN:OD1	2:B:541:SER:OG[2_545]	2.07	0.13

### 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	
2	B	1308/1368 (96%)	1268 (97%)	36 (3%)	4 (0%)	41	73
2	F	1313/1368 (96%)	1267 (96%)	41 (3%)	5 (0%)	34	68
All	All	2621/2736 (96%)	2535 (97%)	77 (3%)	9 (0%)	41	73

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	F	1042	ILE
2	F	585	ASP
2	F	1011	GLY
2	F	1020	LYS
2	B	1327	PHE

#### 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	
2	B	1170/1225 (96%)	1131 (97%)	39 (3%)	38	67
2	F	1155/1225 (94%)	1093 (95%)	62 (5%)	22	54
All	All	2325/2450 (95%)	2224 (96%)	101 (4%)	29	60

5 of 101 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	384	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	F	631	MET
2	F	1338	SER
2	F	425	ARG
2	F	532	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	920	GLN
2	F	807	GLN
2	F	178	ASN
2	F	794	GLN
2	F	46	ASN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	33/34 (97%)	11 (33%)	4 (12%)
1	E	32/34 (94%)	11 (34%)	2 (6%)
5	I	62/65 (95%)	19 (30%)	1 (1%)
5	J	62/65 (95%)	18 (29%)	1 (1%)
All	All	189/198 (95%)	59 (31%)	8 (4%)

5 of 59 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	2	U
1	A	3	A
1	A	4	A
1	A	6	G
1	A	9	U

5 of 8 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
5	J	42	A
5	I	42	A
1	E	3	A
1	A	28	A
1	E	27	G

#### 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](#)

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	34/34 (100%)	-0.28	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	18, 45, 134, 169	0
1	E	32/34 (94%)	-0.04	1 (3%) 49 52	38, 77, 181, 216	0
2	B	1322/1368 (96%)	0.24	103 (7%) <span style="border: 1px solid red; padding: 2px;">13</span> <span style="border: 1px solid red; padding: 2px;">15</span>	14, 71, 195, 218	0
2	F	1327/1368 (97%)	0.25	99 (7%) <span style="border: 1px solid red; padding: 2px;">14</span> <span style="border: 1px solid red; padding: 2px;">16</span>	13, 92, 149, 193	0
3	C	26/26 (100%)	-0.39	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	27, 50, 107, 115	0
3	G	26/26 (100%)	-0.54	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	45, 63, 112, 124	0
4	D	11/11 (100%)	0.05	1 (9%) <span style="border: 1px solid red; padding: 2px;">9</span> <span style="border: 1px solid red; padding: 2px;">11</span>	41, 53, 134, 168	0
4	H	11/11 (100%)	0.89	2 (18%) <span style="border: 1px solid red; padding: 2px;">1</span> <span style="border: 1px solid red; padding: 2px;">1</span>	41, 59, 110, 175	0
5	I	63/65 (96%)	-0.42	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	20, 74, 126, 157	0
5	J	63/65 (96%)	-0.38	1 (1%) <span style="border: 1px solid blue; padding: 2px;">72</span> <span style="border: 1px solid blue; padding: 2px;">74</span>	36, 56, 156, 192	0
All	All	2915/3008 (96%)	0.19	207 (7%) <span style="border: 1px solid red; padding: 2px;">16</span> <span style="border: 1px solid red; padding: 2px;">18</span>	13, 77, 168, 218	0

The worst 5 of 207 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	801	VAL	19.2
1	E	34	G	12.6
2	B	810	LYS	11.8
2	B	868	ASP	11.2
4	H	2	DT	9.8

### 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](#)

There are no ligands in this entry.

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

There are no such residues in this entry.