



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 7, 2017 – 02:49 PM EDT

PDB ID : 5UDZ  
Title : Human LIN28A in complex with let-7f-1 microRNA pre-element  
Authors : Nam, Y.; Wang, L.; Sliz, P.  
Deposited on : unknown  
Resolution : 2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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

<http://wwpdb.org/validation/2016/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) : 1.9-1692  
EDS : rb-20029824  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20029824

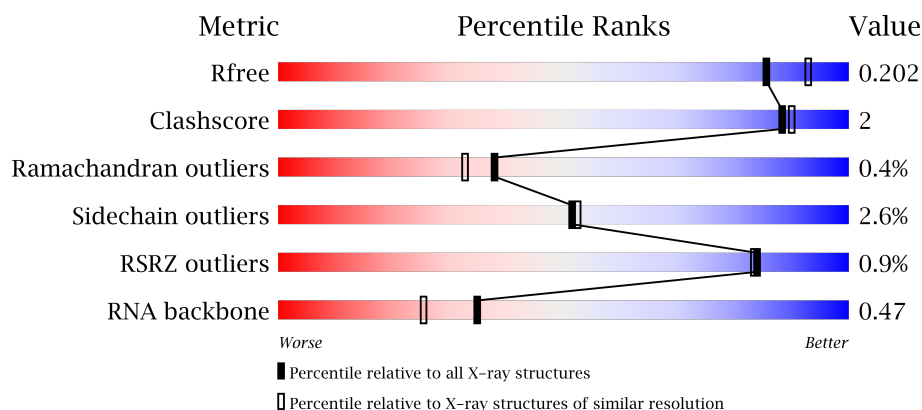
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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))
$R_{free}$	100719	6609 (2.00-2.00)
Clashscore	112137	7775 (2.00-2.00)
Ramachandran outliers	110173	7679 (2.00-2.00)
Sidechain outliers	110143	7678 (2.00-2.00)
RSRZ outliers	101464	6696 (2.00-2.00)
RNA backbone	2435	1011 (2.66-1.34)

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. 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	148	<div> <div>0.1%</div> <div>89%</div> <div>5%</div> <div>6%</div> </div>
1	B	148	<div> <div>0.1%</div> <div>91%</div> <div>•</div> <div>6%</div> </div>
2	V	25	<div> <div>4%</div> <div>60%</div> <div>32%</div> <div>•</div> </div>
2	W	25	<div> <div>68%</div> <div>20%</div> <div>•</div> <div>8%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	ZN	A	201	-	-	-	X
3	ZN	B	201	-	-	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3644 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 lin-28 homolog A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	139	Total	C	N	O	S	0	0	0
			1064	671	195	186	12			
1	B	139	Total	C	N	O	S	0	0	0
			1064	671	195	186	12			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	LYS	deletion	UNP Q9H9Z2
A	?	-	SER	deletion	UNP Q9H9Z2
A	?	-	MET	deletion	UNP Q9H9Z2
A	?	-	GLN	deletion	UNP Q9H9Z2
A	?	-	LYS	deletion	UNP Q9H9Z2
A	?	-	ARG	deletion	UNP Q9H9Z2
A	?	-	ARG	deletion	UNP Q9H9Z2
A	?	-	SER	deletion	UNP Q9H9Z2
A	?	-	LYS	deletion	UNP Q9H9Z2
B	?	-	LYS	deletion	UNP Q9H9Z2
B	?	-	SER	deletion	UNP Q9H9Z2
B	?	-	MET	deletion	UNP Q9H9Z2
B	?	-	GLN	deletion	UNP Q9H9Z2
B	?	-	LYS	deletion	UNP Q9H9Z2
B	?	-	ARG	deletion	UNP Q9H9Z2
B	?	-	ARG	deletion	UNP Q9H9Z2
B	?	-	SER	deletion	UNP Q9H9Z2
B	?	-	LYS	deletion	UNP Q9H9Z2

- Molecule 2 is a RNA chain called let-7f-1 pre-element.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	V	24	Total	C	N	O	P	0	0	0
			514	230	93	168	23			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	W	23	Total	C	N	O	P	0	0	0
			481	215	83	161	22			

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total	Zn	0	0
			2	2		
3	A	2	Total	Zn	0	0
			2	2		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	154	Total	O	0	0
			154	154		
4	B	153	Total	O	0	0
			153	153		
4	V	107	Total	O	0	0
			107	107		
4	W	103	Total	O	0	0
			103	103		



- Molecule 1: Protein lin-28 homolog A



- ALA  
ALA  
D33  
R122  
K150  
E151  
A178  
Q179  
Q180  
GLY  
PRO  
SER  
ALA  
GLN  
GLY  
LYS

- 
- | Category | Color  |
|----------|--------|
| G1       | Green  |
| G7       | Yellow |
| U8       | Green  |
| G9       | Yellow |
| A10      | Green  |
| U11      | Yellow |
| U12      | Green  |
| U13      | Yellow |
| U14      | Orange |
| A15      | Yellow |
| C16      | Yellow |
| A24      | Yellow |
| U        | Grey   |

- 
- Diagram of a 1D polymer chain with 14 beads. The beads are labeled G1, G2, G3, G7, U8, U14, A22, G23, A, and U. The beads are colored: G1, G2, G3 are green; G7, U8 are orange; U14 is yellow; A22, G23 are yellow; A, U are grey. The beads are connected by green lines.

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	76.43Å 76.43Å 105.61Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.52 – 2.00 76.43 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.0 (32.52-2.00) 92.4 (76.43-2.00)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.03 (at 2.00Å)	Xtriage
Refinement program	PHENIX 1.8.2_1309	Depositor
R, $R_{free}$	0.164 , 0.200 0.167 , 0.202	Depositor DCC
$R_{free}$ test set	1839 reflections (4.86%)	DCC
Wilson B-factor (Å <sup>2</sup> )	20.9	Xtriage
Anisotropy	0.003	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 48.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.044 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3644	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.13% 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: ZN

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	0.58	0/1091	0.64	0/1463
1	B	0.50	0/1091	0.61	0/1463
2	V	0.84	0/575	1.27	6/896 (0.7%)
2	W	0.71	0/537	1.08	1/836 (0.1%)
All	All	0.63	0/3294	0.88	7/4658 (0.2%)

There are no bond length outliers.

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	V	11	U	N1-C2-O2	7.12	127.78	122.80
2	W	3	G	O5'-P-OP2	-6.26	100.06	105.70
2	V	15	A	O5'-P-OP1	-5.93	100.36	105.70
2	V	15	A	OP1-P-OP2	5.57	127.95	119.60
2	V	9	G	N1-C6-O6	5.44	123.16	119.90
2	V	11	U	N3-C2-O2	-5.38	118.43	122.20
2	V	14	U	O4'-C1'-N1	5.05	112.24	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	A	1064	0	1050	6	0
1	B	1064	0	1050	1	0
2	V	514	0	259	3	0
2	W	481	0	243	4	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
4	A	154	0	0	4	2
4	B	153	0	0	1	2
4	V	107	0	0	1	1
4	W	103	0	0	4	1
All	All	3644	0	2602	13	3

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 2.

All (13) 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:121:GLU:OE1	4:A:301:HOH:O	1.97	0.81
1:B:122:ARG:NH2	4:B:303:HOH:O	2.23	0.71
2:W:8:U:H5'	4:W:102:HOH:O	1.92	0.67
1:A:39:HIS:HD2	4:A:427:HOH:O	1.89	0.55
2:W:7:G:O3'	4:W:102:HOH:O	2.18	0.54
2:V:13:U:OP1	4:V:101:HOH:O	2.19	0.54
2:W:8:U:C5'	4:W:102:HOH:O	2.55	0.53
1:A:122:ARG:NH2	4:A:304:HOH:O	2.43	0.52
1:A:102:LYS:HE3	4:A:355:HOH:O	2.14	0.48
2:W:22:A:H2'	4:W:164:HOH:O	2.16	0.45
1:A:118:ILE:HD11	1:A:165:GLN:HG3	1.99	0.45
2:V:16:C:H5''	2:V:16:C:H6	1.84	0.42
1:A:102:LYS:HB2	2:V:10:A:C6	2.55	0.42

All (3) 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:V:186:HOH:O	4:W:131:HOH:O[3_645]	2.08	0.12
4:A:394:HOH:O	4:B:320:HOH:O[3_655]	2.17	0.03
4:A:307:HOH:O	4:B:384:HOH:O[3_655]	2.17	0.03

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	137/148 (93%)	134 (98%)	3 (2%)	0	100	100
1	B	137/148 (93%)	131 (96%)	5 (4%)	1 (1%)	25	18
All	All	274/296 (93%)	265 (97%)	8 (3%)	1 (0%)	38	33

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	178	ALA

### 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	114/118 (97%)	112 (98%)	2 (2%)	64	68
1	B	114/118 (97%)	110 (96%)	4 (4%)	41	39
All	All	228/236 (97%)	222 (97%)	6 (3%)	51	52

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	34	GLU
1	A	180	GLN
1	B	122	ARG
1	B	150	LYS
1	B	151	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	180	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	179	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	V	23/25 (92%)	3 (13%)	0
2	W	22/25 (88%)	3 (13%)	0
All	All	45/50 (90%)	6 (13%)	0

All (6) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	V	7	G
2	V	14	U
2	V	24	A
2	W	7	G
2	W	14	U
2	W	23	G

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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 4 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	139/148 (93%)	-0.26	1 (0%) 87 87	6, 12, 34, 58	0
1	B	139/148 (93%)	-0.34	1 (0%) 87 87	8, 16, 40, 53	0
2	V	24/25 (96%)	-0.80	1 (4%) 37 37	10, 18, 41, 66	0
2	W	23/25 (92%)	-0.91	0 100 100	15, 24, 43, 50	0
All	All	325/346 (93%)	-0.38	3 (0%) 84 83	6, 16, 41, 66	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	33	ASP	3.4
2	V	24	A	3.2
1	B	180	GLN	3.0

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ZN	A	201	1/1	1.00	0.14	4.40	9,9,9,9	0
3	ZN	B	201	1/1	1.00	0.13	3.55	13,13,13,13	0
3	ZN	B	200	1/1	0.99	0.10	1.03	15,15,15,15	0
3	ZN	A	200	1/1	1.00	0.11	-0.08	13,13,13,13	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.