



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 23, 2017 – 12:24 PM EDT

PDB ID : 3HJY  
Title : Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA  
Authors : Liang, B.; Zhou, J.; Kahen, E.; Terns, R.M.; Terns, M.P.; Li, H.  
Deposited on : unknown  
Resolution : 3.65 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

<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-20030345  
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-20030345

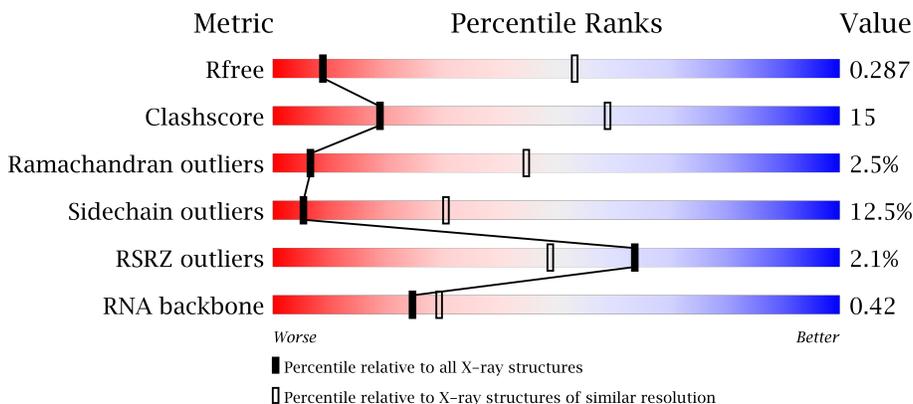
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.65 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	1134 (3.82-3.50)
Clashscore	112137	1267 (3.82-3.50)
Ramachandran outliers	110173	1219 (3.82-3.50)
Sidechain outliers	110143	1218 (3.82-3.50)
RSRZ outliers	101464	1160 (3.82-3.50)
RNA backbone	2435	1004 (4.40-2.90)

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	327	 .% 57% 37% . .
2	B	53	 55% 40% . .
3	C	21	 14% 38% 57% 5%
4	D	25	 60% 28% 12%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
5	E	14	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into four segments: a red segment on the far left labeled '14%', a green segment labeled '50%', a yellow segment labeled '43%', and a red segment on the far right labeled '7%'.</p>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 4260 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 pseudouridine synthase Cbf5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	320	2544	1645	435	454	10	0	0	0

- Molecule 2 is a protein called Ribosome biogenesis protein Nop10.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	53	445	283	86	72	4	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	21	450	201	86	143	20	0	0	0

- Molecule 4 is a RNA chain called RNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	25	521	234	89	174	24	0	0	0

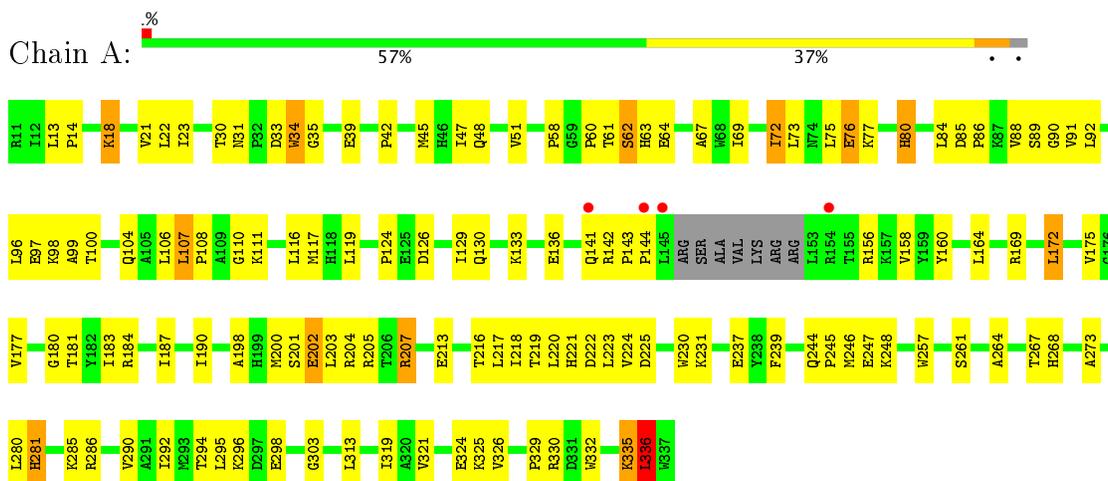
- Molecule 5 is a RNA chain called 5'-R(\*GP\*GP\*AP\*GP\*CP\*GP\*UP\*GP\*CP\*GP\*GP\*UP\*UP\*U)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	E	14	300	134	54	99	13	0	0	0

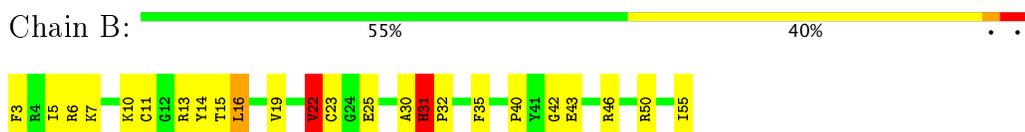
### 3 Residue-property plots

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: pseudouridine synthase Cbf5



- Molecule 2: Ribosome biogenesis protein Nop10



- Molecule 3: 5'-R(\*GP\*GP\*GP\*CP\*UP\*CP\*CP\*GP\*GP\*AP\*AP\*AP\*CP\*CP\*GP\*CP\*GP\*GP\*CP\*GP\*C)-3'



- Molecule 4: RNA (25-MER)



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



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 62 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	238.34Å 238.34Å 127.37Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.33 – 3.65 49.33 – 3.65	Depositor EDS
% Data completeness (in resolution range)	97.8 (49.33-3.65) 97.8 (49.33-3.65)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.00 (at 3.67Å)	Xtrriage
Refinement program	CNS, REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.276 , 0.306 0.268 , 0.287	Depositor DCC
$R_{free}$ test set	1208 reflections (5.10%)	DCC
Wilson B-factor (Å <sup>2</sup> )	113.4	Xtrriage
Anisotropy	0.448	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 63.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	4260	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	132.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.20% 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	0.68	1/2603 (0.0%)	0.73	0/3523
2	B	0.85	0/458	0.85	0/613
3	C	1.21	6/503 (1.2%)	1.33	2/784 (0.3%)
4	D	0.89	0/579	1.58	10/899 (1.1%)
5	E	0.81	0/335	1.34	2/522 (0.4%)
All	All	0.81	7/4478 (0.2%)	1.04	14/6341 (0.2%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	1	G	C2'-C1'	9.82	1.64	1.53
3	C	1	G	C2'-O2'	7.90	1.51	1.41
3	C	2	G	C2'-O2'	6.76	1.50	1.41
1	A	202	GLU	CG-CD	6.69	1.61	1.51
3	C	4	C	C2'-O2'	6.33	1.49	1.41

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	8	G	C5-N7-C8	-7.58	100.51	104.30
4	D	21	A	P-O3'-C3'	6.53	127.54	119.70
4	D	8	G	C4-C5-N7	6.26	113.30	110.80
3	C	5	U	O4'-C1'-N1	6.21	113.17	108.20
5	E	9	G	O4'-C1'-N9	6.21	113.17	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	2544	0	2613	98	0
2	B	445	0	453	23	0
3	C	450	0	232	1	0
4	D	521	0	273	3	0
5	E	300	0	152	1	0
All	All	4260	0	3723	116	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 15.

The worst 5 of 116 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:85:ASP:HB3	1:A:86:PRO:HD2	1.44	0.95
1:A:18:LYS:HE3	1:A:18:LYS:H	1.31	0.93
1:A:164:LEU:CD1	1:A:172:LEU:HD22	2.05	0.85
2:B:11:CYS:HG	2:B:23:CYS:HG	1.07	0.85
2:B:16:LEU:H	2:B:16:LEU:HD12	1.41	0.84

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	316/327 (97%)	265 (84%)	44 (14%)	7 (2%)	8	47
2	B	51/53 (96%)	42 (82%)	7 (14%)	2 (4%)	3	34
All	All	367/380 (97%)	307 (84%)	51 (14%)	9 (2%)	6	45

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	77	LYS
1	A	136	GLU
1	A	246	MET
1	A	336	LEU
2	B	22	VAL

### 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	271/277 (98%)	239 (88%)	32 (12%)	6	33
2	B	48/48 (100%)	40 (83%)	8 (17%)	2	17
All	All	319/325 (98%)	279 (88%)	40 (12%)	5	30

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	142	ARG
1	A	184	ARG
2	B	22	VAL
1	A	172	LEU
1	A	207	ARG

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

Mol	Chain	Res	Type
1	A	141	GLN
2	B	31	HIS
1	A	199	HIS
1	A	130	GLN
1	A	189	HIS

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	C	20/21 (95%)	9 (45%)	1 (5%)
4	D	24/25 (96%)	5 (20%)	2 (8%)
5	E	13/14 (92%)	6 (46%)	2 (15%)
All	All	57/60 (95%)	20 (35%)	5 (8%)

5 of 20 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	C	9	G
3	C	10	A
3	C	11	A
3	C	12	A
3	C	13	C

All (5) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
3	C	9	G
4	D	21	A
4	D	23	A
5	E	9	G
5	E	10	U

#### 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](#)

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

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	320/327 (97%)	-0.01	4 (1%) 77 64	67, 108, 164, 185	0
2	B	53/53 (100%)	-0.33	0 100 100	86, 101, 130, 132	0
3	C	21/21 (100%)	0.39	3 (14%) 3 3	132, 202, 249, 257	0
4	D	25/25 (100%)	-0.25	0 100 100	119, 145, 165, 219	0
5	E	14/14 (100%)	0.11	2 (14%) 3 3	157, 194, 276, 288	0
All	All	433/440 (98%)	-0.04	9 (2%) 64 49	67, 110, 194, 288	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	10	A	7.4
5	E	16	U	3.7
1	A	144	PRO	3.6
5	E	17	U	3.4
1	A	154	ARG	2.8

### 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 carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.