



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 29, 2024 – 04:16 PM EDT

PDB ID : 4PR6  
Title : A Second Look at the HDV Ribozyme Structure and Dynamics.  
Authors : Kapral, G.J.; Jain, S.; Noeske, J.; Doudna, J.A.; Richardson, D.C.; Richardson, J.S.  
Deposited on : 2014-03-05  
Resolution : 2.30 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

<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  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 1.20.1  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.003 (Gargrove)  
Density-Fitness : 1.0.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.39

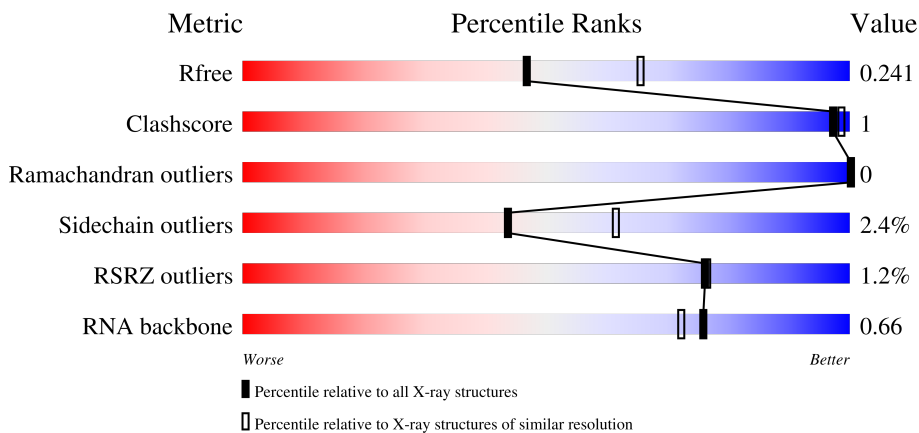
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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))
$R_{free}$	164625	5963 (2.30-2.30)
Clashscore	180529	6698 (2.30-2.30)
Ramachandran outliers	177936	6640 (2.30-2.30)
Sidechain outliers	177891	6640 (2.30-2.30)
RSRZ outliers	164620	5963 (2.30-2.30)
RNA backbone	3690	1032 (2.60-2.00)

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	B	144	 44% 6% 50%
2	A	95	 94% 6%

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4061 atoms, of which 1608 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 HDV RIBOZYME SELF-CLEAVED.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				P
1	B	72	2375	701	803	280	518	73	0	2	0

- Molecule 2 is a protein called U1 small nuclear ribonucleoprotein A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				Se
2	A	95	1586	499	805	139	139	4	0	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	31	HIS	TYR	engineered mutation	UNP P09012
A	36	ARG	GLN	engineered mutation	UNP P09012
A	97	MSE	-	expression tag	UNP P09012
A	98	ALA	-	expression tag	UNP P09012

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	13	Total	Mg	0	1
			14	14		
3	A	3	Total	Mg	0	0
			3	3		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	20	Total	O	0	0
			20	20		
4	A	63	Total	O	0	0
			63	63		



## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	109.35Å 109.35Å 190.68Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	18.58 – 2.30 18.58 – 2.30	Depositor EDS
% Data completeness (in resolution range)	95.7 (18.58-2.30) 95.5 (18.58-2.30)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.39 (at 2.30Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.199 , 0.241 0.204 , 0.241	Depositor DCC
$R_{free}$ test set	1863 reflections (9.86%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	53.1	Xtriage
Anisotropy	0.076	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 59.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4061	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	87.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.24% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG

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	B	0.24	0/1754	0.73	0/2733
2	A	0.33	0/797	0.48	0/1059
All	All	0.27	0/2551	0.67	0/3792

There are no bond length outliers.

There are no bond angle outliers.

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	B	1572	803	804	1	0
2	A	781	805	799	3	0
3	A	3	0	0	0	0
3	B	14	0	0	0	0
4	A	63	0	0	0	0
4	B	20	0	0	0	0
All	All	2453	1608	1603	3	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 1.

All (3) 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:B:150:U:O2	2:A:83:ARG:NH2	2.50	0.44
2:A:9:ASN:ND2	2:A:90:ASP:OD1	2.51	0.40
2:A:89:THR:HG22	2:A:90:ASP:N	2.37	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
2	A	95/95 (100%)	93 (98%)	2 (2%)	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
2	A	86/81 (106%)	84 (98%)	2 (2%)	45 63

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	6	THR
2	A	7	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	B	69/144 (47%)	7 (10%)	0

All (7) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	B	110	G
1	B	126	C
1	B	127	U
1	B	141	C
1	B	155	U
1	B	158	G
1	B	167	U

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 oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 17 ligands modelled in this entry, 17 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	B	72/144 (50%)	0.51	2 (2%) 55 56	51, 104, 133, 208	2 (2%)
2	A	91/95 (95%)	-0.50	0 100 100	24, 54, 96, 166	1 (1%)
All	All	163/239 (68%)	-0.06	2 (1%) 76 76	24, 67, 129, 208	3 (1%)

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	125	G	3.4
1	B	128	G	2.2

### 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](#)

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, 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	B-factors(Å <sup>2</sup> )	Q<0.9
3	MG	B	410[A]	1/1	0.54	0.15	143,143,143,143	1
3	MG	B	410[B]	1/1	0.54	0.15	137,137,137,137	1
3	MG	B	409	1/1	0.62	0.18	118,118,118,118	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
3	MG	B	412	1/1	0.73	0.11	103,103,103,103	0
3	MG	B	402	1/1	0.82	0.09	92,92,92,92	0
3	MG	B	406	1/1	0.82	0.08	102,102,102,102	0
3	MG	B	413	1/1	0.85	0.10	98,98,98,98	0
3	MG	B	411	1/1	0.87	0.07	98,98,98,98	0
3	MG	B	407	1/1	0.89	0.12	94,94,94,94	0
3	MG	A	101	1/1	0.89	0.07	71,71,71,71	0
3	MG	B	408	1/1	0.90	0.09	94,94,94,94	0
3	MG	B	401	1/1	0.90	0.20	85,85,85,85	0
3	MG	B	403	1/1	0.94	0.09	77,77,77,77	0
3	MG	B	405	1/1	0.94	0.14	87,87,87,87	0
3	MG	A	102	1/1	0.94	0.12	92,92,92,92	0
3	MG	B	404	1/1	0.95	0.10	80,80,80,80	0
3	MG	A	103	1/1	0.95	0.19	66,66,66,66	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.