wwPDB X-ray Structure Validation Summary Report (i)

Oct 7, 2023 – 06:32 PM EDT

PDB ID	:	6DOI
Title	:	Crystal Structure of Bacillus Halodurans Ribonuclease H1 in Complex with an RNA/DNA Hybrid (1.54 Angstrom wavelength): Soak in 0.5 mM EGTA and 200 mM K+ at 21 C
Authors	:	Samara, N.L.; Yang, W.
Deposited on	:	2018-06-09
Resolution	:	1.95 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:

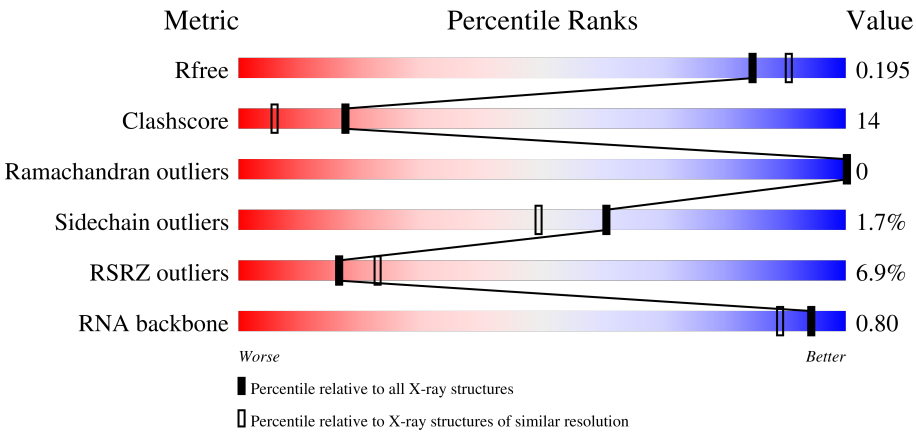
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 1.95 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	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)
RNA backbone	3102	1003 (2.46-1.42)

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% 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	133	8%, 82%, 17%, •
2	B	4	50%, 50%
3	b	2	100%
4	C	6	67%, 33%

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
8	GOL	A	209	-	-	X	-
9	EDO	C	102	-	-	X	-

2 Entry composition (i)

There are 10 unique types of molecules in this entry. The entry contains 2456 atoms, of which 893 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 Ribonuclease H.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	133	2009	719	893	180	215	2	0	5	0

- Molecule 2 is a RNA chain called 5'-R(*AP*CP*AP*U)-3' portion of intact RNA (5'-R(*AP*CP*AP*UP*CP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	4	80	38	15	24	3	0	1	0

- Molecule 3 is a RNA chain called 5'-R(P*CP*G)-3' portion of intact RNA (5'-R(*AP*CP*AP*UP*CP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	b	2	42	19	8	13	2	0	2	0

- Molecule 4 is a DNA chain called DNA (5'-D(*CP*GP*AP*TP*GP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	C	6	121	59	22	35	5	0	0	0

- Molecule 5 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	K		
5	A	2	2	2	0	2
5	C	1	1	1	0	0

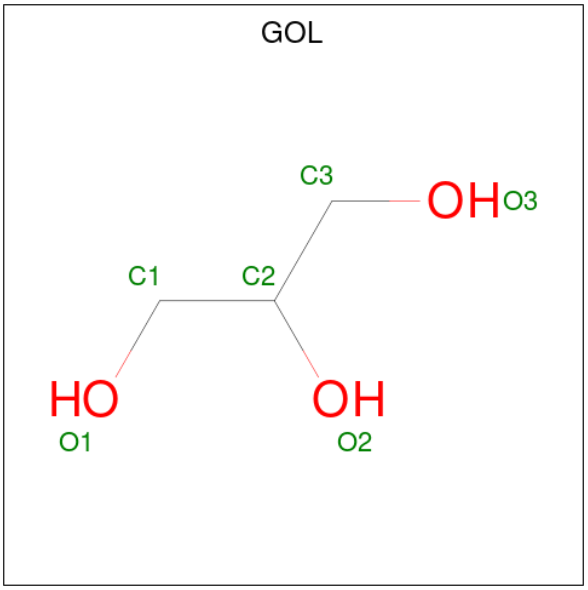
- Molecule 6 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
6	A	1	1	1	0	1

- Molecule 7 is IODIDE ION (three-letter code: IOD) (formula: I).

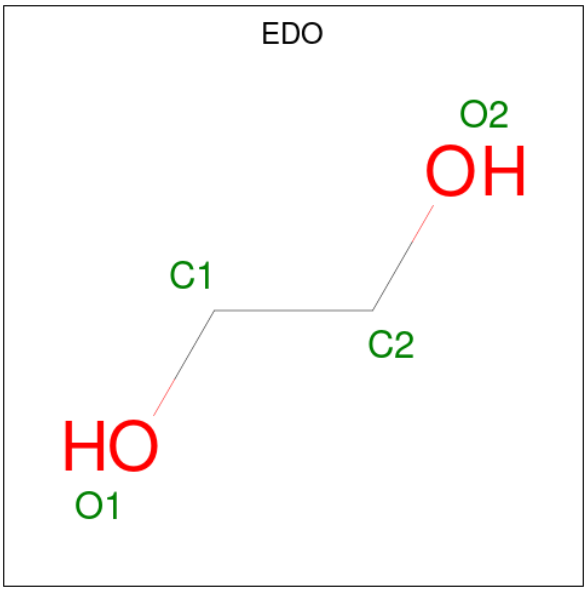
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	I		
7	A	4	4	4	0	0

- Molecule 8 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
8	A	1	6	3	3	0	0
			Total	C	O		
8	A	1	6	3	3	0	0
			Total	C	O		
8	A	1	6	3	3	0	0
			Total	C	O		
8	A	1	6	3	3	0	0

- Molecule 9 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
9	A	1	4	2	2	0	0
			Total	C	O		
9	C	1	4	2	2	0	0

- Molecule 10 is water.

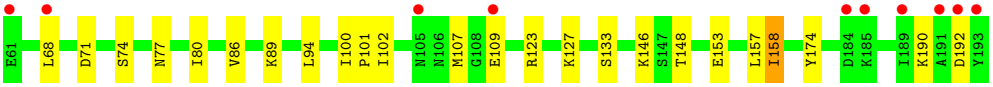
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
10	A	131	131	131	0	0
			Total	O		
10	B	13	13	13	0	0
			Total	O		
10	b	5	5	5	0	0
			Total	O		
10	C	15	15	15	0	1

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: Ribonuclease H

Chain A:



• Molecule 2: 5'-R(*AP*CP*AP*U)-3' portion of intact RNA (5'-R(*AP*CP*AP*UP*CP*G)-3')

Chain B:

• Molecule 3: 5'-R(P*CP*G)-3' portion of intact RNA (5'-R(*AP*CP*AP*UP*CP*G)-3')

Chain b:

There are no outlier residues recorded for this chain.

• Molecule 4: DNA (5'-D(*CP*GP*AP*TP*GP*T)-3')

Chain C:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	81.46Å 38.02Å 62.07Å 90.00° 96.22° 90.00°	Depositor
Resolution (Å)	22.01 – 1.95 22.01 – 1.95	Depositor EDS
% Data completeness (in resolution range)	92.1 (22.01-1.95) 92.2 (22.01-1.95)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.23 (at 1.95Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.147 , 0.195 0.149 , 0.195	Depositor DCC
R_{free} test set	608 reflections (4.69%)	wwPDB-VP
Wilson B-factor (Å^2)	24.5	Xtriage
Anisotropy	0.052	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 63.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	2456	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 12.75% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: CA, EDO, GOL, K, IOD

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	0.46	0/1157	0.61	0/1568
2	B	0.74	0/89	1.05	0/136
3	b	0.85	0/46	1.22	0/69
4	C	1.12	1/135 (0.7%)	1.04	0/207
All	All	0.59	1/1427 (0.1%)	0.73	0/1980

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	2	DG	C3'-O3'	-5.03	1.37	1.44

There are no bond angle outliers.

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	1116	893	1113	28	0
2	B	80	0	44	1	0
3	b	42	0	23	0	0
4	C	121	0	70	2	0
5	A	2	0	0	0	0
5	C	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	1	0	0	0	0
7	A	4	0	0	2	0
8	A	24	0	32	15	0
9	A	4	0	6	0	0
9	C	4	0	6	4	0
10	A	131	0	0	8	0
10	B	13	0	0	0	0
10	C	15	0	0	0	0
10	b	5	0	0	0	0
All	All	1563	893	1294	36	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 14.

The worst 5 of 36 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)
4:C:4:DT:H2'	9:C:102:EDO:H22	1.50	0.94
7:A:206:IOD:I	10:A:430:HOH:O	2.77	0.71
1:A:109[B]:GLU:HG2	1:A:133:SER:HB3	1.79	0.65
1:A:153:GLU:HG3	10:A:411:HOH:O	1.98	0.61
1:A:109[B]:GLU:HG2	1:A:133:SER:CB	2.31	0.60

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	136/133 (102%)	135 (99%)	1 (1%)	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
1	A	123/118 (104%)	121 (98%)	2 (2%)	62 52

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

Mol	Chain	Res	Type
1	A	146	LYS
1	A	158	ILE

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
2	B	2/4 (50%)	0	0
3	b	0/2	-	-
All	All	2/6 (33%)	0	0

There are no RNA backbone outliers to report.

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)

Of 14 ligands modelled in this entry, 8 are monoatomic - leaving 6 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
9	EDO	C	102	-	3,3,3	0.48	0	2,2,2	0.38	0
8	GOL	A	208	-	5,5,5	0.38	0	5,5,5	0.42	0
9	EDO	A	210	-	3,3,3	0.43	0	2,2,2	0.37	0
8	GOL	A	211	-	5,5,5	0.35	0	5,5,5	0.23	0
8	GOL	A	212	-	5,5,5	0.36	0	5,5,5	0.51	0
8	GOL	A	209	-	5,5,5	0.42	0	5,5,5	0.23	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
9	EDO	C	102	-	-	0/1/1/1	-
8	GOL	A	208	-	-	2/4/4/4	-
9	EDO	A	210	-	-	0/1/1/1	-
8	GOL	A	211	-	-	4/4/4/4	-
8	GOL	A	212	-	-	2/4/4/4	-
8	GOL	A	209	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
8	A	211	GOL	C1-C2-C3-O3
8	A	212	GOL	O1-C1-C2-C3
8	A	212	GOL	O1-C1-C2-O2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
8	A	208	GOL	O1-C1-C2-C3
8	A	209	GOL	O1-C1-C2-C3

There are no ring outliers.

5 monomers are involved in 19 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
9	C	102	EDO	4	0
8	A	208	GOL	3	0
8	A	211	GOL	3	0
8	A	212	GOL	1	0
8	A	209	GOL	8	0

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, 95th 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^2)	Q<0.9
1	A	133/133 (100%)	0.21	10 (7%)	14	20	15, 25, 59, 94	0
2	B	4/4 (100%)	-0.01	0	100	100	23, 25, 39, 40	0
3	b	2/2 (100%)	-0.52	0	100	100	24, 24, 24, 32	0
4	C	6/6 (100%)	0.20	0	100	100	23, 30, 40, 43	0
All	All	145/145 (100%)	0.19	10 (6%)	16	23	15, 25, 56, 94	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	68[A]	LEU	3.7
1	A	193	TYR	2.9
1	A	61	GLU	2.9
1	A	192	ASP	2.7
1	A	109[A]	GLU	2.6

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, 95th 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^2)	Q<0.9
8	GOL	A	211	6/6	0.65	0.28	77,77,81,85	0
8	GOL	A	208	6/6	0.81	0.26	66,76,83,89	0
8	GOL	A	209	6/6	0.82	0.24	61,67,68,70	0
6	CA	A	203[B]	1/1	0.82	0.62	46,46,46,46	1
9	EDO	C	102	4/4	0.84	0.18	34,39,41,41	0
8	GOL	A	212	6/6	0.86	0.22	62,65,68,68	0
9	EDO	A	210	4/4	0.93	0.15	54,56,68,70	0
5	K	C	101	1/1	0.94	0.09	32,32,32,32	1
7	IOD	A	205	1/1	0.97	0.05	47,47,47,47	1
7	IOD	A	207	1/1	0.98	0.19	35,35,35,35	1
7	IOD	A	206	1/1	0.99	0.07	40,40,40,40	1
5	K	A	201[A]	1/1	0.99	0.13	21,21,21,21	1
7	IOD	A	204	1/1	1.00	0.12	21,21,21,21	1
5	K	A	202[B]	1/1	1.00	0.17	24,24,24,24	1

6.5 Other polymers (i)

There are no such residues in this entry.