# Full wwPDB X-ray Structure Validation Report (i)

Oct 3, 2023 – 07:03 AM EDT

| | | |
|---|---|---|
| PDB ID | : | 6OZF |
| Title | : | Crystal structure of Thermotoga maritima (Tm) Endonuclease V (D110N) in complex with a 12mer DNA containing an inosine followed by a ribo-adenosine |
| Authors | : | Samara, N.L.; Yang, W. |
| Deposited on | : | 2019-05-15 |
| Resolution | : | 1.80 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*
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 | : | FAILED |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | FAILED |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| 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 1.80 Å.

There are no overall percentile quality scores available for this entry.

MolProbity and EDS failed to run properly - the sequence quality summary graphics cannot be shown.

# 2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 4478 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 Endonuclease V.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 1 | A | 221 | 1795 | 1171 | 303 | 315 | 6 | 0 | 4 | 0 |
| 1 | B | 225 | 1852 | 1211 | 311 | 323 | 7 | 0 | 9 | 0 |

There are 2 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 110 | ASN | ASP | engineered mutation | UNP Q9X2H9 |
| B | 110 | ASN | ASP | engineered mutation | UNP Q9X2H9 |

- Molecule 2 is DNA/RNA hybrid called DNA/RNA (5'-D(P*AP*AP*TP*GP*I)-R(P*A)-D(P*GP*AP*TP*GP*CP*T)-3').

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | P | | | |
| 2 | C | 11 | 214 | 99 | 41 | 63 | 11 | 0 | 0 | 0 |
| 2 | D | 12 | 248 | 119 | 48 | 70 | 11 | 0 | 0 | 0 |

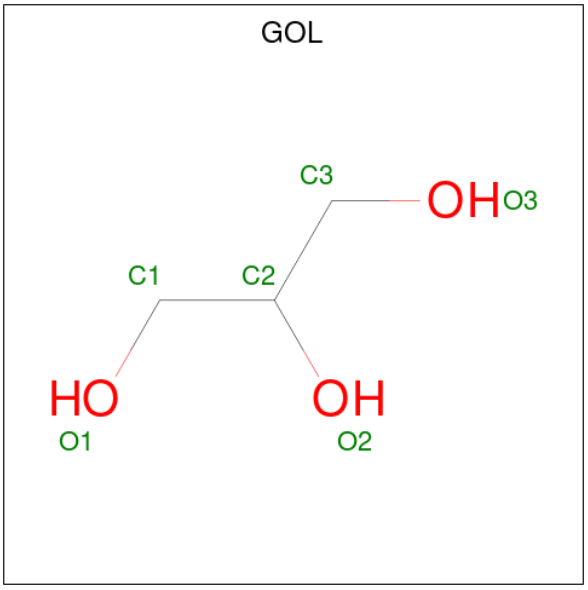
- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| | | | Total | Ca | | |
| 3 | A | 1 | 1 | 1 | 0 | 0 |
| 3 | B | 1 | 1 | 1 | 0 | 0 |
| 3 | C | 1 | 1 | 1 | 0 | 0 |
| 3 | D | 1 | 1 | 1 | 0 | 0 |

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

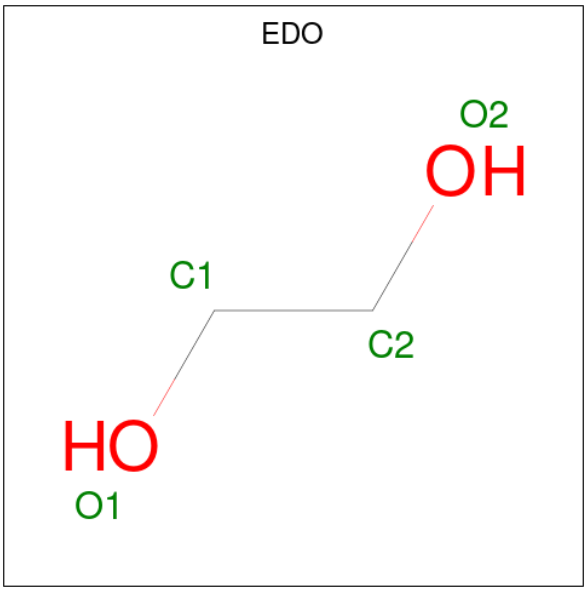
| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| | | | Total | Na | | |
| 4 | A | 1 | 1 | 1 | 0 | 0 |
| | | | Total | Na | | |
| 4 | B | 1 | 1 | 1 | 0 | 0 |

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



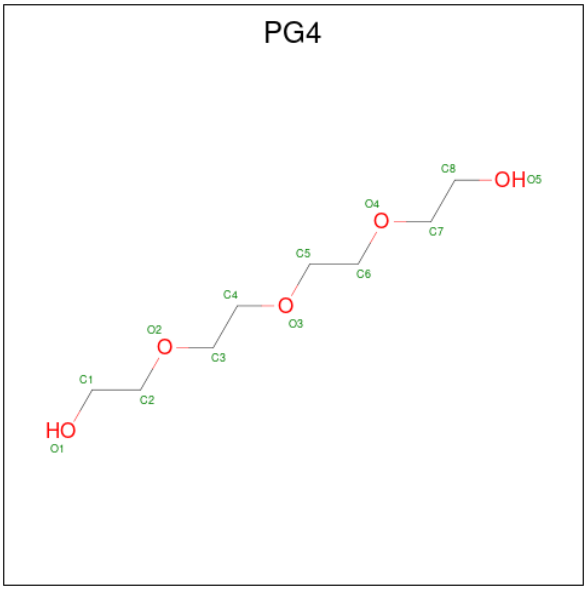

| Mol | Chain | Residues | Atoms | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|
| | | | Total | C | O | | |
| 5 | B | 1 | 6 | 3 | 3 | 0 | 0 |
| | | | Total | C | O | | |
| 5 | B | 1 | 6 | 3 | 3 | 0 | 0 |

- Molecule 6 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).




| Mol | Chain | Residues | Atoms | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|
| | | | Total | C | O | | |
| 6 | B | 1 | 4 | 2 | 2 | 0 | 0 |
| | | | Total | C | O | | |
| 6 | C | 1 | 4 | 2 | 2 | 0 | 0 |

- Molecule 7 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).




| Mol | Chain | Residues | Atoms | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|
| | | | Total | C | O | | |
| 7 | B | 1 | 13 | 8 | 5 | 0 | 0 |

- Molecule 8 is water.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| | | | Total | O | | |
| 8 | A | 115 | 115 | 115 | 0 | 0 |
| | | | Total | O | | |
| 8 | B | 154 | 154 | 154 | 0 | 1 |
| | | | Total | O | | |
| 8 | C | 30 | 30 | 30 | 0 | 0 |
| | | | Total | O | | |
| 8 | D | 31 | 31 | 31 | 0 | 0 |

MolProbity and EDS failed to run properly - this section is therefore empty.

# 3 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

| Property | Value | Source |
|---|---|---|
| Space group | P 2 21 21 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 63.13Å 75.58Å 106.40Å<br>90.00° 90.00° 90.00° | Depositor |
| Resolution (Å) | 28.09 – 1.80 | Depositor |
| % Data completeness<br>(in resolution range) | 99.1 (28.09-1.80) | Depositor |
| $R_{merge}$ | (Not available) | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 2.08 (at 1.80Å) | Xtriage |
| Refinement program | PHENIX (1.12_2829: ???) | Depositor |
| R, $R_{free}$ | 0.167 , 0.198 | Depositor |
| Wilson B-factor (Å$^2$) | 21.0 | Xtriage |
| Anisotropy | 0.369 | Xtriage |
| L-test for twinning[2] | $< \|L\| > = 0.47$, $< L^2 > = 0.30$ | Xtriage |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| Total number of atoms | 4478 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 30.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.32% 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.

# 4 Model quality (i)

## 4.1 Standard geometry (i)

MolProbity failed to run properly - this section is therefore empty.

## 4.2 Too-close contacts (i)

MolProbity failed to run properly - this section is therefore empty.

## 4.3 Torsion angles (i)

### 4.3.1 Protein backbone (i)

MolProbity failed to run properly - this section is therefore empty.

### 4.3.2 Protein sidechains (i)

MolProbity failed to run properly - this section is therefore empty.

### 4.3.3 RNA (i)

MolProbity failed to run properly - this section is therefore empty.

## 4.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

## 4.5 Carbohydrates (i)

There are no monosaccharides in this entry.

## 4.6 Ligand geometry (i)

Of 11 ligands modelled in this entry, 6 are monoatomic - leaving 5 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 | $\#\|Z\| > 2$ |
| 7 | PG4 | B | 305 | - | 12,12,12 | 0.52 | 0 | 11,11,11 | 0.35 | 0 |
| 5 | GOL | B | 302 | - | 5,5,5 | 0.95 | 0 | 5,5,5 | 0.85 | 0 |
| 6 | EDO | C | 102 | - | 3,3,3 | 0.42 | 0 | 2,2,2 | 0.31 | 0 |
| 6 | EDO | B | 304 | - | 3,3,3 | 0.51 | 0 | 2,2,2 | 0.25 | 0 |
| 5 | GOL | B | 303 | - | 5,5,5 | 0.86 | 0 | 5,5,5 | 1.02 | 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 |
|---|---|---|---|---|---|---|---|
| 7 | PG4 | B | 305 | - | - | 3/10/10/10 | - |
| 5 | GOL | B | 302 | - | - | 4/4/4/4 | - |
| 6 | EDO | C | 102 | - | - | 0/1/1/1 | - |
| 6 | EDO | B | 304 | - | - | 0/1/1/1 | - |
| 5 | GOL | B | 303 | - | - | 0/4/4/4 | - |

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (7) torsion outliers are listed below:

| Mol | Chain | Res | Type | Atoms |
|---|---|---|---|---|
| 5 | B | 302 | GOL | C1-C2-C3-O3 |
| 7 | B | 305 | PG4 | C4-C3-O2-C2 |
| 7 | B | 305 | PG4 | O3-C5-C6-O4 |
| 5 | B | 302 | GOL | O2-C2-C3-O3 |
| 5 | B | 302 | GOL | O1-C1-C2-O2 |
| 7 | B | 305 | PG4 | C5-C6-O4-C7 |
| 5 | B | 302 | GOL | O1-C1-C2-C3 |

There are no ring outliers.

No monomer is involved in short contacts.

## 4.7 Other polymers (i)

There are no such residues in this entry.

## 4.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

# 5 Fit of model and data (i)

## 5.1 Protein, DNA and RNA chains (i)

EDS failed to run properly - this section is therefore empty.

## 5.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS failed to run properly - this section is therefore empty.

## 5.3 Carbohydrates (i)

EDS failed to run properly - this section is therefore empty.

## 5.4 Ligands (i)

EDS failed to run properly - this section is therefore empty.

## 5.5 Other polymers (i)

EDS failed to run properly - this section is therefore empty.