wwPDB X-ray Structure Validation Summary Report (i)

Oct 19, 2023 – 01:52 AM EDT

PDB ID	:	2OPF
Title	:	Crystal structure of the DNA repair enzyme endonuclease-VIII (Nei) from E. coli (R252A) in complex with AP-site containing DNA substrate
Authors	:	Golan, G.; Zharov, D.O.; Grollman, A.P.; Shoham, G.
Deposited on	:	2007-01-29
Resolution	:	1.85 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:

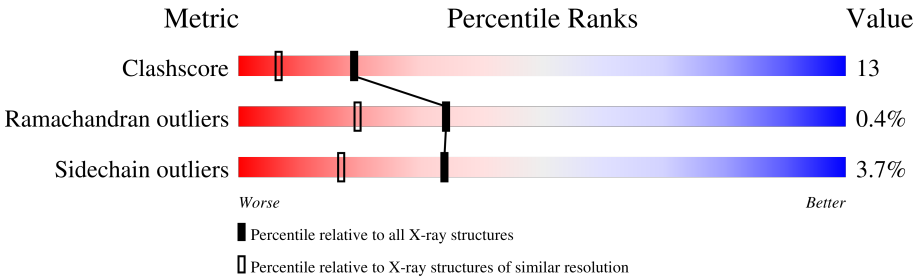
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.85 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))
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	B	12	50% 42% 8%
2	C	12	17% 83%
3	A	262	79% 17% • •

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 2910 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 DNA chain called 5'-D(*GP*GP*CP*TP*TP*CP*AP*TP*CP*CP*TP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	11	Total	C	N	O	P	0	0	0
			221	106	35	69	11			

- Molecule 2 is a DNA chain called 5'-D(P*CP*AP*GP*GP*AP*(PED)P*GP*AP*AP*GP*CP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	12	Total	C	N	O	P	0	0	0
			240	112	49	67	12			

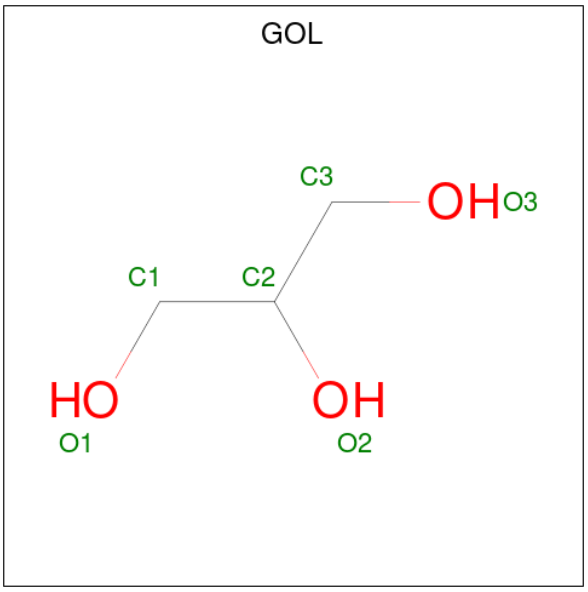
- Molecule 3 is a protein called Endonuclease VIII.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	256	Total	C	N	O	S	0	15	0
			2170	1381	393	391	5			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	34	THR	PRO	SEE REMARK 999	UNP P50465
A	112	ARG	THR	SEE REMARK 999	UNP P50465
A	252	ALA	ARG	engineered mutation	UNP P50465

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

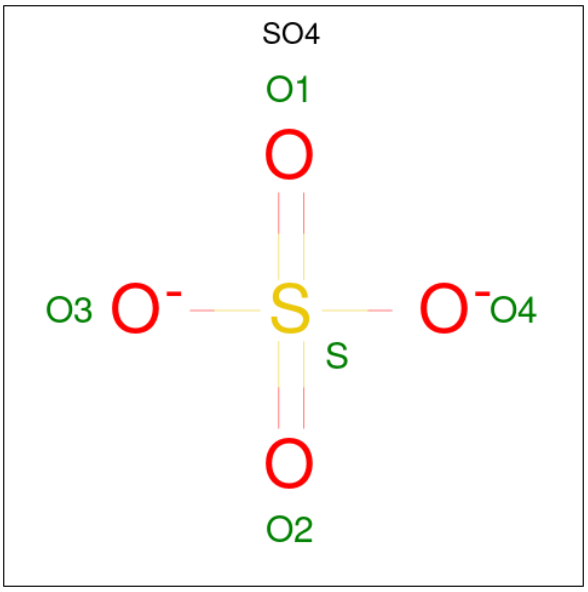


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	C	1	Total C O 6 3 3	0	0
4	A	1	Total C O 6 3 3	0	0
4	A	1	Total C O 6 3 3	0	0
4	A	1	Total C O 6 3 3	0	0
4	A	1	Total C O 6 3 3	0	0

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Zn 1 1	0	0

- Molecule 6 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total O S 5 4 1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	B	23	Total O 23 23	0	0
7	C	30	Total O 30 30	0	0
7	A	190	Total O 190 190	0	0

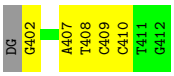
3 Residue-property plots (i)

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. 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. 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.

Note EDS was not executed.

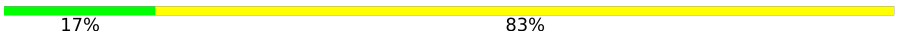
• Molecule 1: 5'-D(*GP*GP*CP*TP*TP*CP*AP*TP*CP*CP*TP*G)-3'

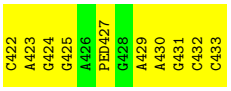
Chain B:



• Molecule 2: 5'-D(P*CP*AP*GP*GP*AP*(PED)P*GP*AP*AP*GP*CP*C)-3'

Chain C:





• Molecule 3: Endonuclease VIII

Chain A:

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 43 2 2	Depositor
Cell constants a, b, c, α, β, γ	73.56Å 73.56Å 170.11Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.41 – 1.85	Depositor
% Data completeness (in resolution range)	89.4 (20.41-1.85)	Depositor
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.240 , 0.267	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2910	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: SO4, ZN, PED, GOL

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	0.41	0/245	0.73	0/375
2	C	0.40	0/257	0.76	0/392
3	A	0.32	0/2219	0.62	1/3010 (0.0%)
All	All	0.33	0/2721	0.65	1/3777 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	83	PRO	N-CA-CB	5.29	109.65	103.30

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	221	0	126	8	0
2	C	240	0	131	17	0
3	A	2170	0	2153	46	1
4	A	24	0	32	6	0
4	C	6	0	8	1	0
5	A	1	0	0	0	0
6	A	5	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	A	190	0	0	3	0
7	B	23	0	0	5	0
7	C	30	0	0	0	1
All	All	2910	0	2450	67	2

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 13.

The worst 5 of 67 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:402:DG:H1	2:C:432:DC:H42	1.16	0.88
2:C:432:DC:H2"	2:C:433:DC:H5"	1.56	0.86
2:C:431:DG:H2"	2:C:432:DC:H2'	1.56	0.85
2:C:432:DC:H3'	2:C:432:DC:OP2	1.78	0.83
3:A:9:ALA:HB2	4:A:703:GOL:H12	1.60	0.82

All (2) 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)
3:A:137:GLU:CG	3:A:137:GLU:CG[7_555]	2.05	0.15
7:C:1076:HOH:O	7:C:1076:HOH:O[5_655]	2.09	0.11

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
3	A	267/262 (102%)	261 (98%)	5 (2%)	1 (0%)	34 19

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	215	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
3	A	232/225 (103%)	223 (96%)	9 (4%)	32 15

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	214	GLN
3	A	215	VAL
3	A	116	LEU
3	A	169	TYR
3	A	171	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	31	GLN
3	A	68	ASN
3	A	123	GLN
3	A	152	GLN
3	A	214	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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 7 ligands modelled in this entry, 1 is 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	$\#\|Z\| > 2$
4	GOL	A	703	-	5,5,5	0.87	0	5,5,5	0.47	0
4	GOL	C	702	-	5,5,5	0.78	0	5,5,5	0.42	0
6	SO4	A	601	-	4,4,4	0.25	0	6,6,6	0.10	0
4	GOL	A	705	-	5,5,5	0.80	0	5,5,5	0.46	0
4	GOL	A	701	-	5,5,5	0.87	0	5,5,5	0.41	0
4	GOL	A	704	-	5,5,5	0.77	0	5,5,5	0.44	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
4	GOL	A	703	-	-	4/4/4/4	-
4	GOL	C	702	-	-	2/4/4/4	-
4	GOL	A	705	-	-	0/4/4/4	-
4	GOL	A	701	-	-	2/4/4/4	-
4	GOL	A	704	-	-	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
4	C	702	GOL	C1-C2-C3-O3
4	A	701	GOL	O1-C1-C2-C3
4	A	703	GOL	C1-C2-C3-O3
4	A	704	GOL	O1-C1-C2-O2
4	A	704	GOL	O1-C1-C2-C3

There are no ring outliers.

3 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	703	GOL	3	0
4	C	702	GOL	1	0
4	A	701	GOL	3	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	427:PED	O3'	428:DG	P	1.76

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.