wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 01:45 AM GMT

PDB ID	:	2EA0
Title	:	Crystal structure of the DNA repair enzyme endonuclease-VIII (Nei) from E. coli 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	:	1.40 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

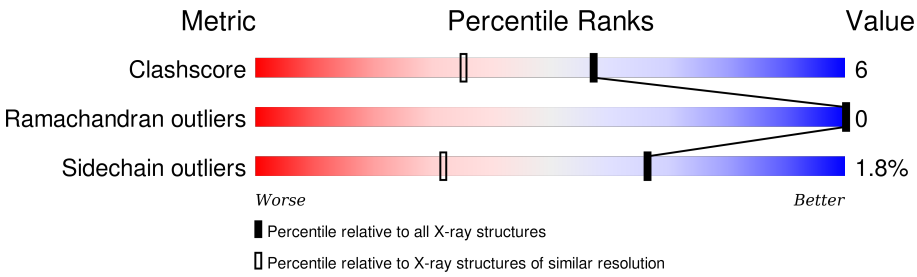
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 1.40 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	102246	1295 (1.40-1.40)
Ramachandran outliers	100387	1259 (1.40-1.40)
Sidechain outliers	100360	1258 (1.40-1.40)

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 >=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 <=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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	B	12	75% 25%
2	C	12	8% 33% 58%
3	A	262	85% 10% • •

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
6	GOL	A	511	-	-	X	-

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 3023 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(P*GP*GP*CP*TP*TP*CP*AP*TP*CP*CP*TP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	B	12	243	116	40	75	12	0	0	0

- 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
			Total	C	N	O	P			
2	C	12	240	112	49	67	12	0	0	0

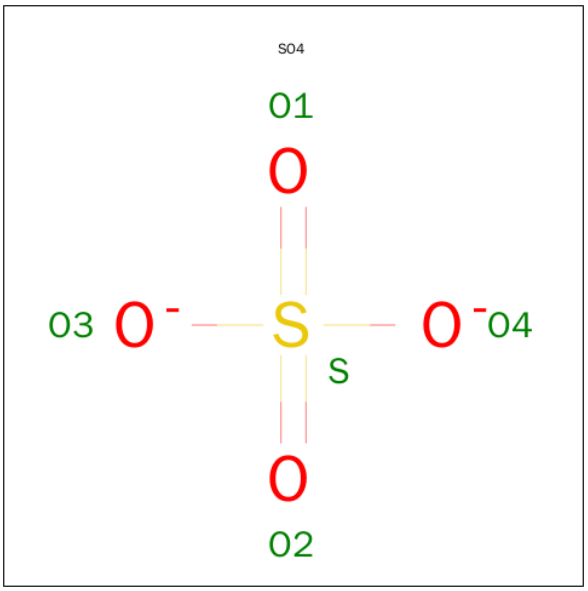
- Molecule 3 is a protein called Endonuclease VIII.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	254	2096	1338	378	375	5	0	15	0

There are 2 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

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

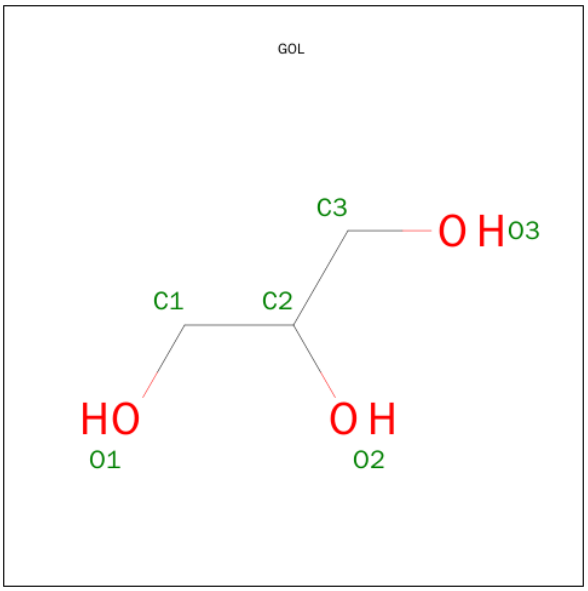


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
4	C	1	5	4	1	0	0
			Total	O	S		
4	A	1	5	4	1	0	0
			Total	O	S		
4	A	1	5	4	1	0	0
			Total	O	S		
4	A	1	5	4	1	0	0
			Total	O	S		
4	A	1	5	4	1	0	0

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

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

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	B	34	34	34	0	0
			Total	O		
7	C	18	18	18	0	0
			Total	O		
7	A	360	360	360	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

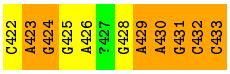
Note EDS was not executed.

- Molecule 1: 5'-D(P*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 will therefore be incomplete.

Property	Value	Source
Space group	P 43 2 2	Depositor
Cell constants a, b, c, α, β, γ	76.18Å 76.18Å 164.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 1.40	Depositor
% Data completeness (in resolution range)	95.8 (30.00-1.40)	Depositor
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	SHELXL-97	Depositor
R, R_{free}	0.160 , 0.204	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3023	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

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: PED, GOL, ZN, SO4

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	3.14	16/270 (5.9%)	5.86	101/414 (24.4%)
2	C	1.97	6/257 (2.3%)	5.73	107/392 (27.3%)
3	A	0.68	0/2217	1.25	20/3006 (0.7%)
All	All	1.31	22/2744 (0.8%)	2.89	228/3812 (6.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	C	1	0

The worst 5 of 22 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	401	DG	O3'-P	18.25	1.83	1.61
1	B	404	DT	C5-C7	18.21	1.60	1.50
1	B	411	DT	C5-C7	16.45	1.59	1.50
1	B	405	DT	C5-C7	16.27	1.59	1.50
1	B	408	DT	C5-C7	14.69	1.58	1.50

The worst 5 of 228 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	402	DG	N3-C4-C5	-22.43	117.39	128.60
2	C	429	DA	OP2-P-O3'	22.16	153.95	105.20
2	C	424	DG	N7-C8-N9	21.70	123.95	113.10
1	B	402	DG	N7-C8-N9	21.44	123.82	113.10
1	B	412	DG	N7-C8-N9	20.95	123.57	113.10

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	C	429	DA	C3'

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	B	243	0	137	4	0
2	C	240	0	130	8	1
3	A	2096	0	2100	18	0
4	A	20	0	0	1	0
4	C	5	0	0	0	0
5	A	1	0	0	0	0
6	A	6	0	7	6	0
7	A	360	0	0	4	0
7	B	34	0	0	2	0
7	C	18	0	0	1	0
All	All	3023	0	2374	30	1

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 6.

The worst 5 of 30 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)
3:A:147:ARG:HH21	6:A:511:GOL:H32	1.18	1.02
3:A:187:LYS:HD2	4:A:553:SO4:O3	1.88	0.72
1:B:405:DT:H5"	7:B:989:HOH:O	1.91	0.70
3:A:112:ARG:H	3:A:115:GLN:HE21	1.41	0.69
3:A:112:ARG:H	3:A:115:GLN:NE2	1.98	0.61

All (1) 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)
2:C:433:DC:O2	2:C:433:DC:O2[6_553]	1.69	0.51

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	266/262 (102%)	257 (97%)	9 (3%)	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
3	A	233/226 (103%)	228 (98%)	5 (2%)	61 24

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	69	GLN
3	A	124[A]	ARG
3	A	124[B]	ARG
3	A	169	TYR
3	A	231	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	68	ASN
3	A	115	GLN
3	A	231	HIS

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 carbohydrates 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	$\#\|Z\| > 2$
6	GOL	A	511	-	5,5,5	0.61	0	5,5,5	0.64	0
4	SO4	A	551	-	4,4,4	0.14	0	6,6,6	0.24	0
4	SO4	A	553	-	4,4,4	0.14	0	6,6,6	0.27	0
4	SO4	A	554	-	4,4,4	0.44	0	6,6,6	1.15	1 (16%)
4	SO4	A	555	-	4,4,4	0.35	0	6,6,6	0.41	0
4	SO4	C	552	-	4,4,4	0.13	0	6,6,6	0.17	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
6	GOL	A	511	-	-	0/4/4/4	0/0/0/0
4	SO4	A	551	-	-	0/0/0/0	0/0/0/0
4	SO4	A	553	-	-	0/0/0/0	0/0/0/0
4	SO4	A	554	-	-	0/0/0/0	0/0/0/0
4	SO4	A	555	-	-	0/0/0/0	0/0/0/0
4	SO4	C	552	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	554	SO4	O4-S-O3	2.11	117.54	108.98

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	511	GOL	6	0
4	A	553	SO4	1	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)

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

6.3 Carbohydrates (i)

EDS was not executed - this section will therefore be empty.

6.4 Ligands (i)

EDS was not executed - this section will therefore be empty.

6.5 Other polymers (i)

EDS was not executed - this section will therefore be empty.