



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 8, 2026 – 03:16 PM UTC

PDB ID : 9YHV / pdb_00009yhv
Title : DNA ligase 1 E346A/E592A in complex with nick containing 3'-8oxorG:C
captured at pre-catalytic stage
Authors : KanalElamparithi, B.; Caglayan, M.
Deposited on : 2025-10-01
Resolution : 2.81 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 ⓘ 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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

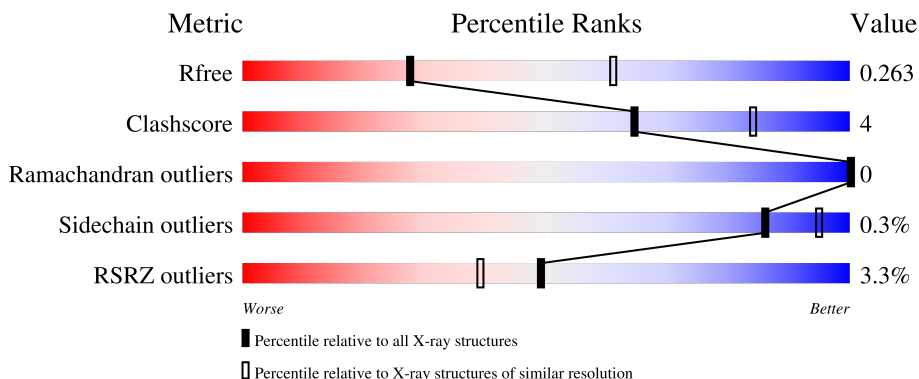
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

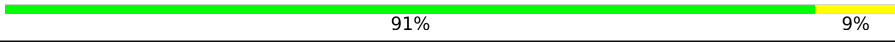
The reported resolution of this entry is 2.81 Å.

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}	180053	4591 (2.84-2.80)
Clashscore	190562	5010 (2.84-2.80)
Ramachandran outliers	187476	4916 (2.84-2.80)
Sidechain outliers	187428	4918 (2.84-2.80)
RSRZ outliers	180081	4594 (2.84-2.80)

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 $\leq 5\%$. 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	A	668	
2	B	11	
3	C	8	
4	D	18	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 4993 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 DNA ligase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	610	Total	C	N	O	S	0	0	1
			4184	2644	743	784	13			

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	346	ALA	GLU	engineered mutation	UNP P18858
A	592	ALA	GLU	engineered mutation	UNP P18858
A	919	ALA	-	expression tag	UNP P18858
A	920	ALA	-	expression tag	UNP P18858
A	921	ALA	-	expression tag	UNP P18858
A	922	LEU	-	expression tag	UNP P18858
A	923	GLU	-	expression tag	UNP P18858
A	924	HIS	-	expression tag	UNP P18858
A	925	HIS	-	expression tag	UNP P18858
A	926	HIS	-	expression tag	UNP P18858
A	927	HIS	-	expression tag	UNP P18858
A	928	HIS	-	expression tag	UNP P18858
A	929	HIS	-	expression tag	UNP P18858

- Molecule 2 is a DNA chain called DNA (5'-D(P*GP*CP*TP*GP*AP*TP*GP*CP*GP*TP*(8GM))-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	11	Total	C	N	O	P	0	0	0
			228	108	42	68	10			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	8	Total	C	N	O	P	0	0	0
			168	78	33	49	8			

- Molecule 4 is a DNA chain called DNA (5'-D(P*GP*TP*CP*CP*GP*AP*CP*CP*AP*CP*GP*CP*AP*TP*CP*AP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	18	Total 361	C 172	N 68	O 104	P 17	0	0	0

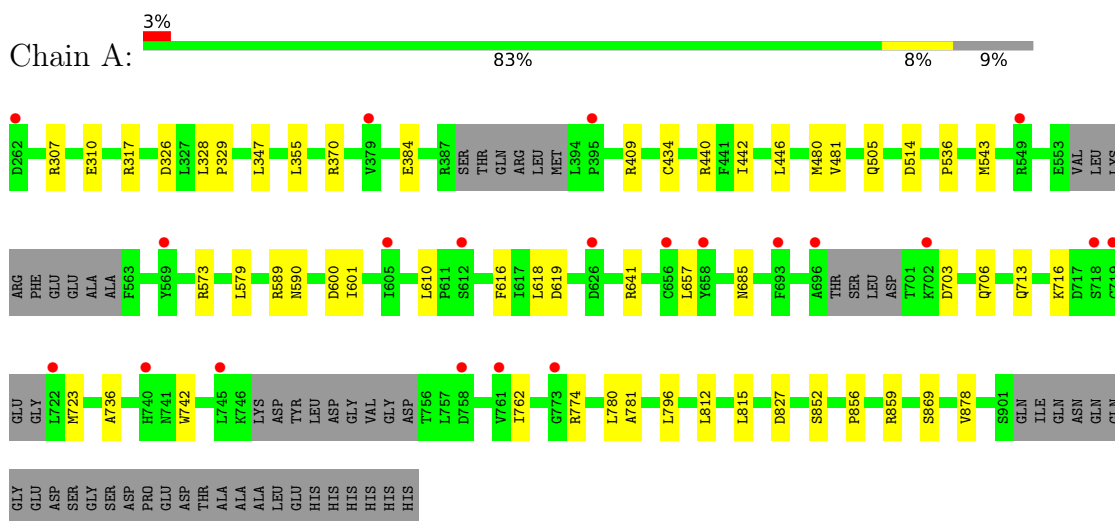
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	30	Total 30	O 30	0	0
5	B	4	Total 4	O 4	0	0
5	C	8	Total 8	O 8	0	0
5	D	10	Total 10	O 10	0	0

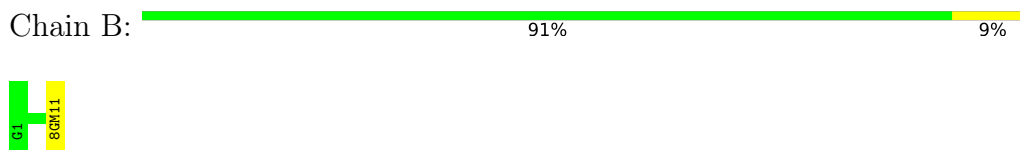
3 Residue-property plots [i](#)

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 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.

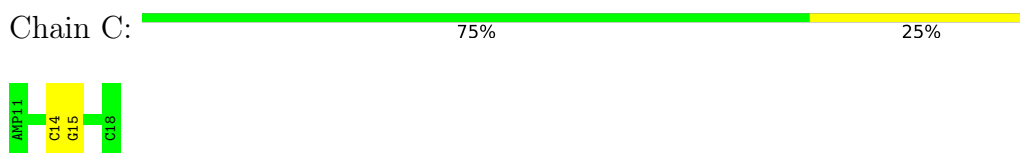
- Molecule 1: DNA ligase 1



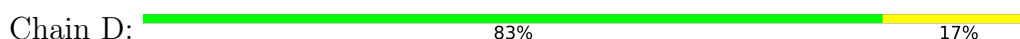
- Molecule 2: DNA (5'-D(P*GP*CP*TP*GP*AP*TP*GP*CP*GP*TP*(8GM))-3')



- Molecule 3: DNA (5'-D(P*(AMP)P*GP*TP*CP*GP*GP*AP*C)-3')



- Molecule 4: DNA (5'-D(P*GP*TP*CP*CP*GP*AP*CP*CP*AP*CP*GP*CP*AP*TP*CP*AP*GP*C)-3')





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	64.12Å 115.58Å 125.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.83 – 2.81 36.83 – 2.81	Depositor EDS
% Data completeness (in resolution range)	99.2 (36.83-2.81) 96.6 (36.83-2.81)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.04 (at 2.81Å)	Xtriage
Refinement program	PHENIX (1.21.1_5286: ???)	Depositor
R, R_{free}	0.221 , 0.262 0.223 , 0.263	Depositor DCC
R_{free} test set	2004 reflections (8.53%)	wwPDB-VP
Wilson B-factor (Å ²)	72.3	Xtriage
Anisotropy	0.698	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 109.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	4993	wwPDB-VP
Average B, all atoms (Å ²)	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.99% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 8GM, AMP

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	0.09	0/4259	0.24	0/5815
2	B	0.19	0/228	0.40	0/351
3	C	0.21	0/162	0.43	0/248
4	D	0.19	0/404	0.33	0/620
All	All	0.11	0/5053	0.26	0/7034

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 [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	4184	0	3660	32	0
2	B	228	0	114	1	0
3	C	168	0	91	1	0
4	D	361	0	202	2	0
5	A	30	0	0	1	0
5	B	4	0	0	0	0
5	C	8	0	0	0	0
5	D	10	0	0	0	0
All	All	4993	0	4067	35	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 4.

All (35) 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:A:703:ASP:HB2	1:A:706:GLN:HB2	1.72	0.70
1:A:307:ARG:HH21	1:A:310:GLU:HB3	1.60	0.66
1:A:573:ARG:NH2	1:A:619:ASP:OD2	2.30	0.64
1:A:762:ILE:HD13	1:A:781:ALA:HB2	1.86	0.58
1:A:505:GLN:HG2	1:A:736:ALA:HB3	1.86	0.58
1:A:827:ASP:OD2	1:A:859:ARG:NH1	2.39	0.56
1:A:780:LEU:HD11	1:A:796:LEU:HD22	1.92	0.52
1:A:347:LEU:HD21	1:A:446:LEU:HD23	1.92	0.52
1:A:384:GLU:OE1	1:A:440:ARG:NE	2.43	0.51
1:A:536:PRO:HB2	1:A:579:LEU:HG	1.91	0.51
1:A:384:GLU:OE1	1:A:440:ARG:HB2	2.10	0.51
2:B:11:8GM:N1	4:D:8:DC:N3	2.59	0.51
1:A:812:LEU:HA	1:A:815:LEU:HD13	1.94	0.50
1:A:723:MET:HA	1:A:723:MET:HE2	1.94	0.50
1:A:317:ARG:NH2	1:A:481:VAL:O	2.45	0.49
1:A:852:SER:HB2	1:A:869:SER:HB2	1.96	0.48
1:A:409:ARG:NH1	5:A:1002:HOH:O	2.42	0.48
1:A:601:ILE:HD11	1:A:657:LEU:HD21	1.96	0.48
1:A:600:ASP:OD1	1:A:600:ASP:N	2.49	0.46
1:A:328:LEU:HD11	1:A:514:ASP:HA	1.98	0.45
1:A:355:LEU:HD13	1:A:370:ARG:HG2	1.98	0.45
1:A:543:MET:HG2	1:A:742:TRP:CE2	2.51	0.45
1:A:589:ARG:NH1	1:A:590:ASN:OD1	2.50	0.44
1:A:616:PHE:CE1	1:A:618:LEU:HD12	2.51	0.44
1:A:326:ASP:C	1:A:329:PRO:HD2	2.43	0.44
1:A:434:CYS:SG	1:A:442:ILE:HD12	2.59	0.42
1:A:713:GLN:HA	1:A:716:LYS:HB3	2.01	0.42
1:A:827:ASP:HB2	1:A:856:PRO:HB3	2.01	0.42
1:A:480:MET:SD	1:A:481:VAL:N	2.93	0.42
4:D:10:DC:H2''	4:D:11:DG:C8	2.54	0.42
1:A:641:ARG:HE	1:A:641:ARG:HB3	1.76	0.41
1:A:774:ARG:H	1:A:774:ARG:HG2	1.75	0.41
1:A:328:LEU:HD23	1:A:328:LEU:HA	1.90	0.41
1:A:610:LEU:HD12	1:A:685:ASN:HA	2.03	0.41
3:C:14:DC:H2'	3:C:15:DG:C8	2.56	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	
1	A	598/668 (90%)	568 (95%)	30 (5%)	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	340/565 (60%)	339 (100%)	1 (0%)	86	95

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	878	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	499	GLN
1	A	706	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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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

6.1 Protein, DNA and RNA chains

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	610/668 (91%)	0.40	21 (3%)	48	38	52, 92, 143, 174	51 (8%)
2	B	10/11 (90%)	-0.25	0	100	100	62, 72, 86, 87	0
3	C	7/8 (87%)	-0.46	0	100	100	54, 56, 65, 73	0
4	D	18/18 (100%)	-0.71	0	100	100	49, 63, 76, 81	0
All	All	645/705 (91%)	0.35	21 (3%)	49	39	49, 91, 142, 174	51 (7%)

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	658	TYR	4.4
1	A	719	CYS	3.7
1	A	262	ASP	3.4
1	A	773	GLY	2.9
1	A	605	ILE	2.7
1	A	656	CYS	2.7
1	A	612	SER	2.6
1	A	569	TYR	2.5
1	A	693	PHE	2.4
1	A	745	LEU	2.3
1	A	718	SER	2.3
1	A	761	VAL	2.3
1	A	395	PRO	2.2
1	A	379	VAL	2.2
1	A	702	LYS	2.2
1	A	740	HIS	2.2
1	A	549	ARG	2.1
1	A	696	ALA	2.1
1	A	722	LEU	2.1
1	A	626	ASP	2.1
1	A	758	ASP	2.1

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 oligosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.