



## wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Mar 2, 2017 – 11:37 am GMT

PDB ID : 3JAA  
EMDB ID: : EMD-6339  
Title : HUMAN DNA POLYMERASE ETA in COMPLEX WITH NORMAL DNA  
AND INCO NUCLEOTIDE (NRM)  
Authors : Lau, W.C.Y.; Li, Y.; Zhang, Q.; Huen, M.S.Y.  
Deposited on : 2015-05-19  
Resolution : 22.00 Å (reported)  
Based on PDB ID : 3MR2

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report  
for a publicly released PDB/EMDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/EMValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc29047

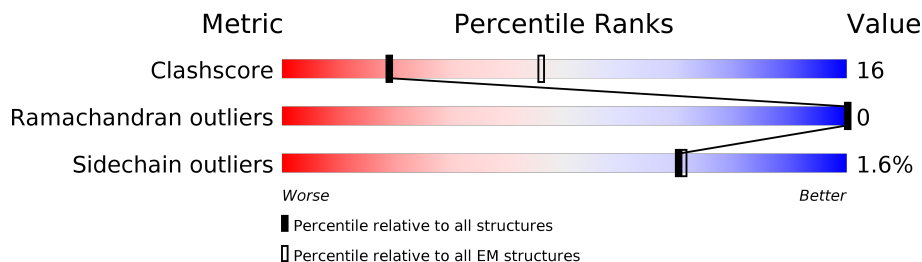
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 22.00 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)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	435	
2	T	13	
3	P	9	

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
4	DZ4	A	509	-	-	X	-

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 4427 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 polymerase eta.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	430	Total	C	N	O	S	32	0
			3555	2226	652	649	28		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP Q9Y253
A	-1	PRO	-	EXPRESSION TAG	UNP Q9Y253
A	0	HIS	-	EXPRESSION TAG	UNP Q9Y253

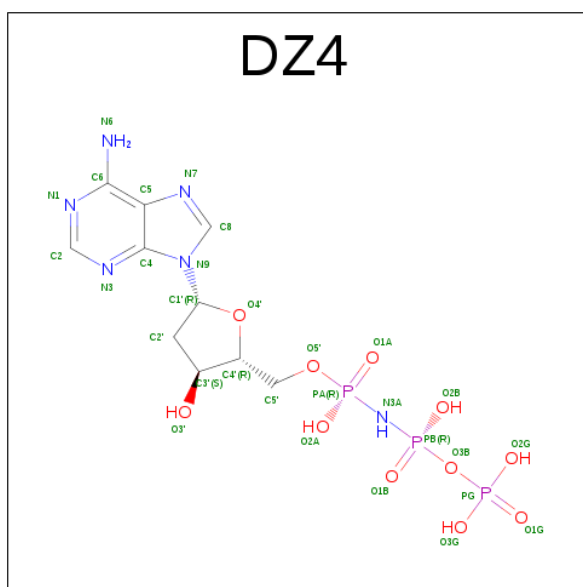
- Molecule 2 is a DNA chain called DNA (5'-D(\*T\*CP\*AP\*TP\*TP\*AP\*TP\*GP\*AP\*CP\*GP\*CP\*T)-3).

Mol	Chain	Residues	Atoms					AltConf	Trace
2	T	12	Total	C	N	O	P	0	1
			226	108	39	68	11		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
3	P	9	Total	C	N	O	P	0	1
			165	78	30	49	8		

- Molecule 4 is 2'-DEOXY-5'-O-[(R)-HYDROXY{[(R)-HYDROXY(PHOSPHONOOXY)P HOSPHORYL]AMINO}PHOSPHORYL]ADENOSINE (three-letter code: DZ4) (formula: C<sub>10</sub>H<sub>17</sub>N<sub>6</sub>O<sub>11</sub>P<sub>3</sub>).

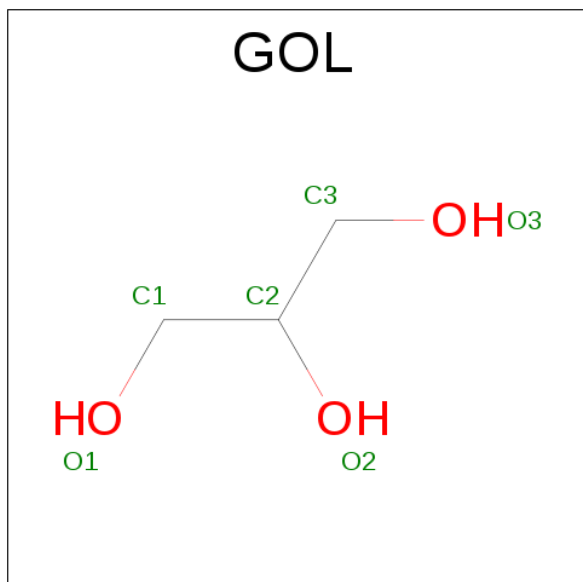


Mol	Chain	Residues	Atoms					AltConf
4	A	1	Total 41	C 16	N 11	O 11	P 3	0
4	A	1	Total 41	C 16	N 11	O 11	P 3	0

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
5	A	2	Total 2	Mg 2	0

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula:  $\text{C}_3\text{H}_8\text{O}_3$ ).



Mol	Chain	Residues	Atoms			AltConf
6	A	1	Total	C	O	0
			30	15	15	
6	A	1	Total	C	O	0
			30	15	15	
6	A	1	Total	C	O	0
			30	15	15	
6	A	1	Total	C	O	0
			30	15	15	
6	T	1	Total	C	O	0
			6	3	3	

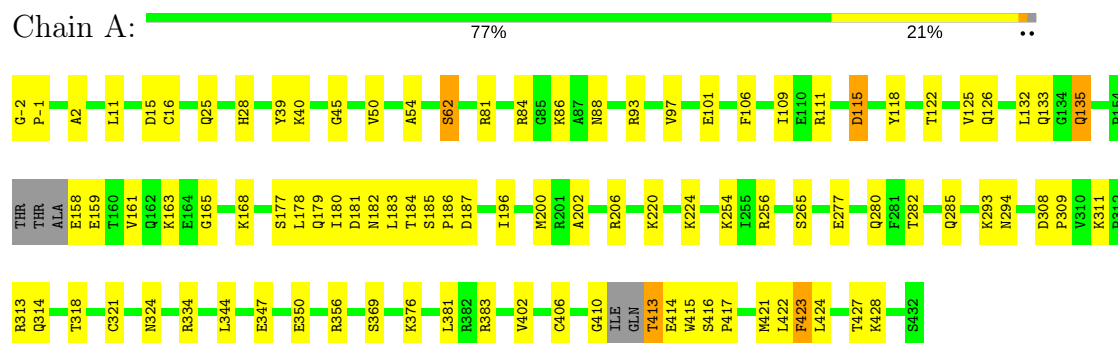
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		AltConf
7	A	345	Total	O	0
			345	345	
7	T	30	Total	O	0
			30	30	
7	P	27	Total	O	0
			27	27	

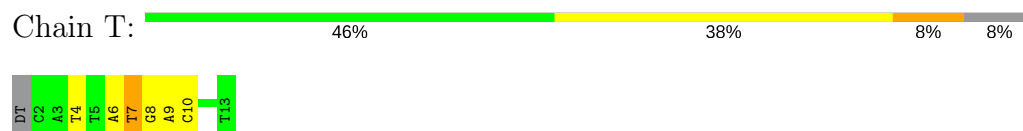
### 3 Residue-property plots [i](#)

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

- Molecule 1: DNA polymerase eta



- Molecule 2: DNA (5'-D(\*T\*CP\*AP\*TP\*TP\*AP\*TP\*GP\*AP\*CP\*GP\*CP\*T)-3



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



There are no outlier residues recorded for this chain.

## 4 Experimental information

Property	Value	Source
Reconstruction method	Not provided	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images used	7330	Depositor
Resolution determination method	FSC 0.143	Depositor
CTF correction method	Particles	Depositor
Microscope	JEOL 2010	Depositor
Voltage (kV)	200	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	18	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	50000	Depositor
Image detector	GATAN ULTRASCAN 4000 (4k x 4k)	Depositor

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, MG, DZ4

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	$\# Z  > 2$	RMSZ	$\# Z  > 2$
1	A	0.90	2/3731 (0.1%)	1.86	4/5024 (0.1%)
2	T	1.41	1/252 (0.4%)	1.35	3/388 (0.8%)
3	P	1.37	0/184	1.13	0/283
All	All	0.96	3/4167 (0.1%)	1.80	7/5695 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	T	6	DA	C2'-C1'	5.69	1.58	1.52
1	A	265[A]	SER	CB-OG	-5.44	1.35	1.42
1	A	265[B]	SER	CB-OG	-5.44	1.35	1.42

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	125[A]	VAL	CG1-CB-CG2	-67.89	2.27	110.90
1	A	125[B]	VAL	CG1-CB-CG2	-67.89	2.27	110.90
1	A	125[C]	VAL	CG1-CB-CG2	-67.89	2.27	110.90
2	T	7	DT	C1'-O4'-C4'	-8.89	101.21	110.10
1	A	115	ASP	CB-CG-OD1	-6.86	112.13	118.30

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	A	3555	0	3642	128	0
2	T	226	0	126	3	0
3	P	165	0	90	0	0
4	A	41	0	16	10	0
5	A	2	0	0	0	0
6	A	30	0	39	1	0
6	T	6	0	7	1	0
7	A	345	0	0	14	0
7	P	27	0	0	0	0
7	T	30	0	0	0	0
All	All	4427	0	3920	129	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 16.

The worst 5 of 129 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:293:LYS:HG2	4:A:509:DZ4:C8	1.75	1.16
1:A:427[B]:THR:HG22	1:A:428:LYS:H	1.11	1.14
1:A:293:LYS:CG	4:A:509:DZ4:H8	1.79	1.13
1:A:293:LYS:HE2	4:A:509:DZ4:C1'	1.82	1.08
1:A:11[A]:LEU:HD23	1:A:224:LYS:HA	1.45	0.98

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 PDB entries followed by that with respect to all EM entries.

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	461/435 (106%)	452 (98%)	9 (2%)	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 PDB entries followed by that with respect to all EM entries.

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	402/372 (108%)	396 (98%)	6 (2%)	70 85

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	115	ASP
1	A	423	PHE
1	A	135	GLN
1	A	62	SER
1	A	413	THR

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

Mol	Chain	Res	Type
1	A	0	HIS
1	A	88	ASN
1	A	294	ASN

### 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

Of 10 ligands modelled in this entry, 2 are monoatomic - leaving 8 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	DZ4	A	501	5	29,32,32	3.50	13 (44%)	32,50,50	2.63	11 (34%)
6	GOL	A	504	-	5,5,5	1.97	1 (20%)	5,5,5	0.93	0
6	GOL	A	505	-	5,5,5	0.42	0	5,5,5	1.04	0
6	GOL	A	506	-	5,5,5	0.95	0	5,5,5	0.45	0
6	GOL	A	507	-	5,5,5	0.25	0	5,5,5	0.42	0
6	GOL	A	508	-	5,5,5	1.21	1 (20%)	5,5,5	0.98	0
4	DZ4	A	509	-	10,12,32	1.61	2 (20%)	8,17,50	4.44	3 (37%)
6	GOL	T	101	-	5,5,5	0.96	0	5,5,5	1.49	1 (20%)

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	DZ4	A	501	5	-	0/14/34/34	0/3/3/3
6	GOL	A	504	-	-	0/4/4/4	0/0/0/0
6	GOL	A	505	-	-	0/4/4/4	0/0/0/0
6	GOL	A	506	-	-	0/4/4/4	0/0/0/0
6	GOL	A	507	-	-	0/4/4/4	0/0/0/0
6	GOL	A	508	-	-	0/4/4/4	0/0/0/0
4	DZ4	A	509	-	-	0/0/0/34	0/2/2/3
6	GOL	T	101	-	-	0/4/4/4	0/0/0/0

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	501	DZ4	PA-O1A	-10.42	1.34	1.46
4	A	501	DZ4	PB-O2B	-10.29	1.28	1.56
4	A	501	DZ4	PB-O1B	-5.10	1.40	1.46

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	501	DZ4	PA-O2A	-4.34	1.44	1.56
4	A	501	DZ4	PG-O3G	-3.57	1.40	1.54

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	509	DZ4	N3-C2-N1	-11.67	118.70	128.86
4	A	501	DZ4	N3-C2-N1	-10.37	119.83	128.86
4	A	501	DZ4	O1B-PB-N3A	-3.90	105.95	111.79
4	A	509	DZ4	C4-C5-N7	-3.13	106.38	109.41
4	A	501	DZ4	C4-C5-N7	-2.72	106.78	109.41

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	507	GOL	1	0
4	A	509	DZ4	10	0
6	T	101	GOL	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.