



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 1, 2021 – 02:08 PM JST

PDB ID : 6LWK  
Title : Crystal structure of human NEIL1(P2G, E3Q, R242) bound to duplex DNA containing dihydrouracil (DHU)  
Authors : Liu, M.H.; Zhang, J.; Zhu, C.X.; Zhang, X.X.; Gao, Y.Q.; Yi, C.Q.  
Deposited on : 2020-02-07  
Resolution : 2.88 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](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

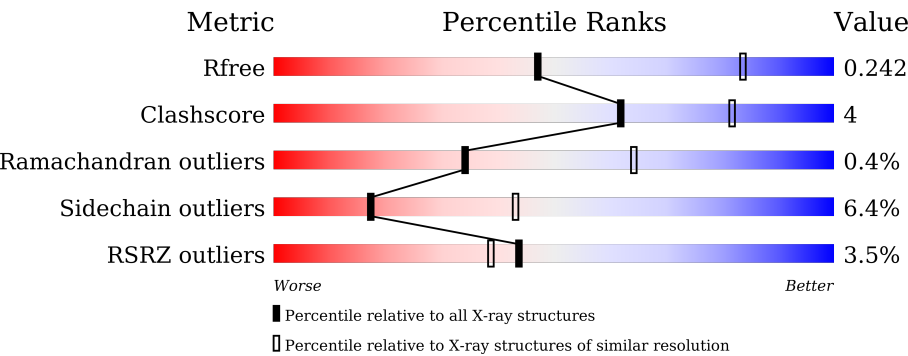
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.19
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.19

# 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 2.88 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))
$R_{free}$	130704	2691 (2.90-2.86)
Clashscore	141614	2947 (2.90-2.86)
Ramachandran outliers	138981	2868 (2.90-2.86)
Sidechain outliers	138945	2871 (2.90-2.86)
RSRZ outliers	127900	2629 (2.90-2.86)

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  $\geq 3$ , 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	295	<div><div>%</div><div><div></div><div>77%</div><div>13%</div><div>•</div><div>8%</div></div></div>
1	D	295	<div><div>%</div><div><div></div><div>81%</div><div>9%</div><div>•</div><div>9%</div></div></div>
1	G	295	<div><div>8%</div><div><div></div><div>71%</div><div>7%</div><div></div><div>21%</div></div></div>
2	B	13	<div><div></div><div><div></div><div>92%</div><div>8%</div></div></div>
2	E	13	<div><div></div><div><div></div><div>100%</div></div></div>
2	H	13	<div><div></div><div><div></div><div>77%</div><div>23%</div></div></div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	C	13	 54% 46%
3	F	13	 62% 31% 8%
3	I	13	 69% 23% 8%

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 7672 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 8-like 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	270	Total	C	N	O	S	0	0	0
			2139	1362	391	376	10			
1	D	269	Total	C	N	O	S	0	0	0
			2135	1360	390	375	10			
1	G	232	Total	C	N	O	S	0	0	0
			1674	1028	326	313	7			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	GLY	PRO	engineered mutation	UNP Q96FI4
A	3	GLN	GLU	engineered mutation	UNP Q96FI4
A	242	ARG	LYS	variant	UNP Q96FI4
D	2	GLY	PRO	engineered mutation	UNP Q96FI4
D	3	GLN	GLU	engineered mutation	UNP Q96FI4
D	242	ARG	LYS	variant	UNP Q96FI4
G	2	GLY	PRO	engineered mutation	UNP Q96FI4
G	3	GLN	GLU	engineered mutation	UNP Q96FI4
G	242	ARG	LYS	variant	UNP Q96FI4

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

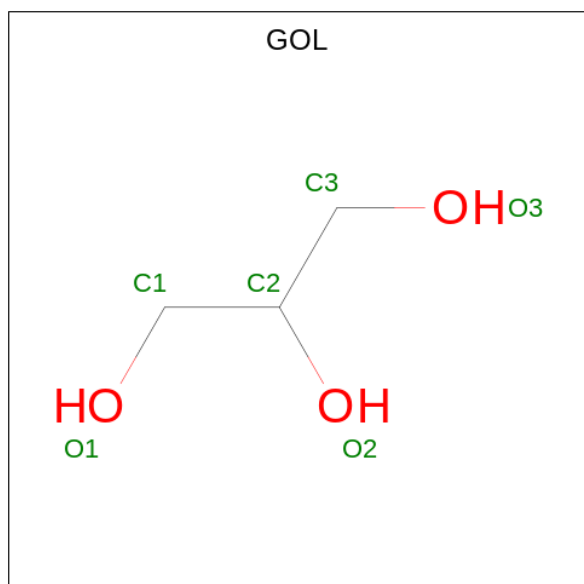
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	13	Total	C	N	O	P	0	0	0
			257	124	43	78	12			
2	E	13	Total	C	N	O	P	0	0	0
			257	124	43	78	12			
2	H	13	Total	C	N	O	P	0	0	0
			257	124	43	78	12			

- Molecule 3 is a DNA chain called DNA (5'-D(\*TP\*AP\*GP\*AP\*CP\*CP\*TP\*GP\*GP\*AP\*

CP\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	13	Total	C	N	O	P	0	0	0
			267	127	53	75	12			
3	F	13	Total	C	N	O	P	0	0	0
			267	127	53	75	12			
3	I	13	Total	C	N	O	P	0	0	0
			267	127	53	75	12			

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



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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	72	Total	O	0	0
			72	72		
5	B	10	Total	O	0	0
			10	10		
5	C	7	Total	O	0	0
			7	7		
5	D	42	Total	O	0	0
			42	42		
5	E	8	Total	O	0	0
			8	8		

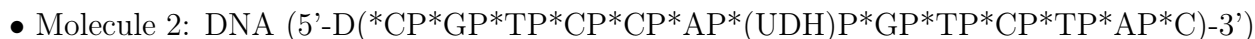
*Continued on next page...*

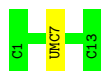
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	F	6	Total	O	0	0
			6	6		
5	I	1	Total	O	0	0
			1	1		



- Molecule 1: Endonuclease 8-like 1





- Molecule 2: DNA (5'-D(\*CP\*GP\*TP\*CP\*CP\*AP\*(UDH)P\*GP\*TP\*CP\*TP\*AP\*C)-3')

Chain E: 100%

There are no outlier residues recorded for this chain.

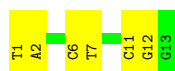
- Molecule 2: DNA (5'-D(\*CP\*GP\*TP\*CP\*CP\*AP\*(UDH)P\*GP\*TP\*CP\*TP\*AP\*C)-3')

Chain H: 77% 23%



- Molecule 3: DNA (5'-D(\*TP\*AP\*GP\*AP\*CP\*CP\*TP\*GP\*GP\*AP\*CP\*GP\*G)-3')

Chain C: 54% 46%



- Molecule 3: DNA (5'-D(\*TP\*AP\*GP\*AP\*CP\*CP\*TP\*GP\*GP\*AP\*CP\*GP\*G)-3')

Chain F: 62% 31% 8%



- Molecule 3: DNA (5'-D(\*TP\*AP\*GP\*AP\*CP\*CP\*TP\*GP\*GP\*AP\*CP\*GP\*G)-3')

Chain I: 69% 23% 8%





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.87Å 109.10Å 169.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.63 – 2.88 49.58 – 2.88	Depositor EDS
% Data completeness (in resolution range)	99.3 (49.63-2.88) 99.4 (49.58-2.88)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.94 (at 2.86Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.201 , 0.240 0.202 , 0.242	Depositor DCC
$R_{free}$ test set	1612 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	67.1	Xtriage
Anisotropy	0.530	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 55.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	7672	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.89% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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

### 5.1 Standard geometry

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

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	0.71	0/2197	0.87	0/2971
1	D	0.70	0/2192	0.86	0/2963
1	G	0.68	0/1707	0.80	0/2282
2	B	0.55	0/264	0.83	0/402
2	E	0.56	0/264	0.80	0/402
2	H	0.48	0/264	0.83	0/402
3	C	0.63	0/300	0.91	0/462
3	F	0.57	0/300	0.90	1/462 (0.2%)
3	I	0.46	0/300	0.87	1/462 (0.2%)
All	All	0.67	0/7788	0.85	2/10808 (0.0%)

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
1	A	0	2
1	D	0	1
1	G	0	1
All	All	0	4

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	2	DA	C2'-C3'-O3'	-6.99	89.54	112.60
3	I	2	DA	C4'-C3'-O3'	6.65	126.32	109.70

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	105	PRO	Peptide
1	A	222	ASN	Peptide
1	D	105	PRO	Peptide
1	G	105	PRO	Peptide

## 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	2139	0	2107	23	0
1	D	2135	0	2103	13	0
1	G	1674	0	1453	9	0
2	B	257	0	149	1	0
2	E	257	0	149	0	0
2	H	257	0	149	5	0
3	C	267	0	147	5	0
3	F	267	0	147	6	0
3	I	267	0	147	4	0
4	A	6	0	8	0	0
5	A	72	0	0	3	0
5	B	10	0	0	0	0
5	C	7	0	0	0	0
5	D	42	0	0	0	0
5	E	8	0	0	0	0
5	F	6	0	0	1	0
5	I	1	0	0	0	0
All	All	7672	0	6559	62	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.

The worst 5 of 62 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)
2:H:11:DT:H2'	2:H:12:DA:N7	1.93	0.83

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:H:11:DT:C2'	2:H:12:DA:C8	2.69	0.76
2:H:11:DT:H2'	2:H:12:DA:C8	2.20	0.76
1:A:246:SER:HB3	1:A:253:PHE:HB2	1.70	0.72
1:A:47:ILE:O	1:D:150:ARG:NH2	2.24	0.70

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	266/295 (90%)	257 (97%)	9 (3%)	0	100	100
1	D	263/295 (89%)	249 (95%)	12 (5%)	2 (1%)	19	48
1	G	214/295 (72%)	201 (94%)	12 (6%)	1 (0%)	29	59
All	All	743/885 (84%)	707 (95%)	33 (4%)	3 (0%)	34	64

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	105	PRO
1	D	106	PRO
1	G	200	ALA

### 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	223/247 (90%)	209 (94%)	14 (6%)	18	44
1	D	223/247 (90%)	211 (95%)	12 (5%)	22	51
1	G	150/247 (61%)	138 (92%)	12 (8%)	12	32
All	All	596/741 (80%)	558 (94%)	38 (6%)	17	43

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	90	ARG
1	G	252	ASP
1	G	95	ARG
1	G	144	PHE
1	G	274	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	130	GLN
1	G	142	GLN
1	G	238	GLN
1	A	202	GLN
1	A	130	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

## 5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

1 ligand is modelled in this entry.

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	301	-	5,5,5	0.15	0	5,5,5	0.36	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	301	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	301	GOL	O1-C1-C2-C3
4	A	301	GOL	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

## 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](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	270/295 (91%)	-0.08	3 (1%) 80 80	41, 62, 103, 133	0
1	D	269/295 (91%)	-0.01	3 (1%) 80 80	42, 77, 127, 153	0
1	G	232/295 (78%)	0.73	24 (10%) 6 4	89, 125, 159, 192	0
2	B	12/13 (92%)	-0.63	0 100 100	63, 104, 140, 158	0
2	E	12/13 (92%)	-0.57	0 100 100	61, 94, 118, 140	0
2	H	12/13 (92%)	-0.64	0 100 100	108, 130, 139, 150	0
3	C	13/13 (100%)	-0.67	0 100 100	80, 96, 126, 142	0
3	F	13/13 (100%)	-0.43	0 100 100	61, 88, 139, 160	0
3	I	13/13 (100%)	-0.43	0 100 100	93, 125, 150, 152	0
All	All	846/963 (87%)	0.12	30 (3%) 44 39	41, 88, 148, 192	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	58	LEU	5.8
1	G	100	ARG	5.0
1	G	75	LEU	4.7
1	A	245	GLY	4.5
1	A	246	SER	4.4

### 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 monosaccharides in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
4	GOL	A	301	6/6	0.81	0.30	92,97,105,106	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.