



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 27, 2016 – 03:47 PM EDT

PDB ID : 5ITT  
Title : Crystal Structure of Human NEIL1 bound to duplex DNA containing THF  
Authors : Zhu, C.; Lu, L.; Zhang, J.; Yue, Z.; Song, J.; Zong, S.; Liu, M.; Stovicek, O.;  
Gao, Y.; Yi, C.  
Deposited on : 2016-03-17  
Resolution : 2.53 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

<http://wwpdb.org/validation/2016/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 : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027939  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027939

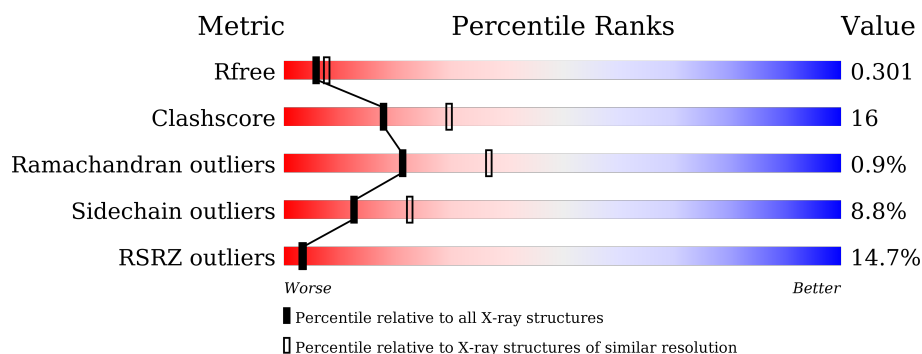
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.53 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))
$R_{free}$	91344	4241 (2.54-2.50)
Clashscore	102246	4968 (2.54-2.50)
Ramachandran outliers	100387	4873 (2.54-2.50)
Sidechain outliers	100360	4875 (2.54-2.50)
RSRZ outliers	91569	4253 (2.54-2.50)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	400	<div> <div>5%</div> <div>52% 12% 32%</div> </div>
1	B	400	<div> <div>5%</div> <div>51% 12% 35%</div> </div>
1	C	400	<div> <div>19%</div> <div>44% 17% 35%</div> </div>
2	D	26	<div> <div>8%</div> <div>77% 23%</div> </div>
2	F	26	<div> <div>38%</div> <div>31% 65%</div> </div>
3	E	26	<div> <div>8%</div> <div>58% 38%</div> </div>

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	GOL	A	501	-	-	-	X
4	GOL	B	501	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8033 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	271	Total	C	N	O	S	0	0	0
			2143	1367	389	377	10			
1	B	261	Total	C	N	O	S	0	0	0
			2076	1327	381	358	10			
1	C	259	Total	C	N	O	S	0	0	0
			2074	1330	378	356	10			

There are 33 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	242	ARG	LYS	engineered mutation	UNP Q96FI4
A	391	ALA	-	expression tag	UNP Q96FI4
A	392	ALA	-	expression tag	UNP Q96FI4
A	393	LEU	-	expression tag	UNP Q96FI4
A	394	GLY	-	expression tag	UNP Q96FI4
A	395	HIS	-	expression tag	UNP Q96FI4
A	396	HIS	-	expression tag	UNP Q96FI4
A	397	HIS	-	expression tag	UNP Q96FI4
A	398	HIS	-	expression tag	UNP Q96FI4
A	399	HIS	-	expression tag	UNP Q96FI4
A	400	HIS	-	expression tag	UNP Q96FI4
B	242	ARG	LYS	engineered mutation	UNP Q96FI4
B	391	ALA	-	expression tag	UNP Q96FI4
B	392	ALA	-	expression tag	UNP Q96FI4
B	393	LEU	-	expression tag	UNP Q96FI4
B	394	GLY	-	expression tag	UNP Q96FI4
B	395	HIS	-	expression tag	UNP Q96FI4
B	396	HIS	-	expression tag	UNP Q96FI4
B	397	HIS	-	expression tag	UNP Q96FI4
B	398	HIS	-	expression tag	UNP Q96FI4
B	399	HIS	-	expression tag	UNP Q96FI4
B	400	HIS	-	expression tag	UNP Q96FI4
C	242	ARG	LYS	engineered mutation	UNP Q96FI4

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	391	ALA	-	expression tag	UNP Q96FI4
C	392	ALA	-	expression tag	UNP Q96FI4
C	393	LEU	-	expression tag	UNP Q96FI4
C	394	GLY	-	expression tag	UNP Q96FI4
C	395	HIS	-	expression tag	UNP Q96FI4
C	396	HIS	-	expression tag	UNP Q96FI4
C	397	HIS	-	expression tag	UNP Q96FI4
C	398	HIS	-	expression tag	UNP Q96FI4
C	399	HIS	-	expression tag	UNP Q96FI4
C	400	HIS	-	expression tag	UNP Q96FI4

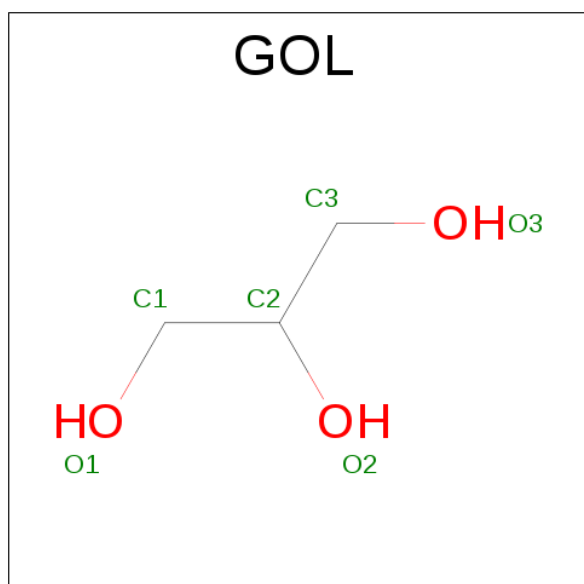
- Molecule 2 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	26	Total	C	N	O	P	0	0	0
			516	247	94	151	24			
2	F	26	Total	C	N	O	P	0	0	0
			516	247	94	151	24			

- Molecule 3 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	26	Total	C	N	O	P	0	0	0
			514	247	94	149	24			

- 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		
4	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	108	Total	O	0	0
			108	108		
5	B	48	Total	O	0	0
			48	48		
5	C	4	Total	O	0	0
			4	4		
5	D	14	Total	O	0	0
			14	14		
5	E	6	Total	O	0	0
			6	6		
5	F	2	Total	O	0	0
			2	2		







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.86Å 108.75Å 171.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	81.81 – 2.53 32.54 – 2.53	Depositor EDS
% Data completeness (in resolution range)	97.9 (81.81-2.53) 97.9 (32.54-2.53)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.19 (at 2.54Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, $R_{free}$	0.210 , 0.278 0.264 , 0.301	Depositor DCC
$R_{free}$ test set	2319 reflections (5.05%)	DCC
Wilson B-factor (Å <sup>2</sup> )	56.2	Xtriage
Anisotropy	0.418	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 31.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	8033	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.48% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 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	A	1.10	2/2202 (0.1%)	1.11	16/2979 (0.5%)
1	B	0.94	1/2133 (0.0%)	1.05	10/2885 (0.3%)
1	C	0.57	2/2129 (0.1%)	0.78	4/2878 (0.1%)
2	D	0.55	0/577	0.91	1/886 (0.1%)
2	F	0.42	0/577	0.81	2/886 (0.2%)
3	E	0.61	0/575	0.99	1/882 (0.1%)
All	All	0.83	5/8193 (0.1%)	0.98	34/11396 (0.3%)

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	B	0	1
All	All	0	3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	223	PRO	N-CD	5.52	1.55	1.47
1	A	280	TRP	CE3-CZ3	5.40	1.47	1.38
1	C	64	PRO	N-CD	5.29	1.55	1.47
1	C	68	PRO	N-CD	5.24	1.55	1.47
1	A	68	PRO	N-CD	5.05	1.54	1.47

The worst 5 of 34 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	133	ARG	NE-CZ-NH2	-11.66	114.47	120.30
1	A	133	ARG	NE-CZ-NH1	11.19	125.89	120.30
1	A	133	ARG	NE-CZ-NH2	-8.87	115.86	120.30
1	A	119	ARG	NE-CZ-NH2	-8.86	115.87	120.30
1	B	133	ARG	NE-CZ-NH1	8.61	124.60	120.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	249	GLY	Peptide
1	A	94	PRO	Peptide
1	B	67	GLN	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	2143	0	2103	84	0
1	B	2076	0	2061	46	0
1	C	2074	0	2064	80	0
2	D	516	0	290	7	0
2	F	516	0	290	23	0
3	E	514	0	290	8	0
4	A	6	0	8	1	0
4	B	6	0	8	0	0
5	A	108	0	0	4	0
5	B	48	0	0	2	0
5	C	4	0	0	0	0
5	D	14	0	0	0	0
5	E	6	0	0	0	0
5	F	2	0	0	0	0
All	All	8033	0	7114	236	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 236 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:199:GLU:OE1	1:B:202:GLN:NE2	1.57	1.35
1:A:250:GLU:HG3	1:A:253:PHE:HD1	1.03	1.16
1:A:250:GLU:HG3	1:A:253:PHE:CD1	1.85	1.11
1:A:250:GLU:HB2	1:A:253:PHE:HB3	1.08	1.08
1:C:282:GLN:OE1	1:C:284:ASP:N	1.88	1.07

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	267/400 (67%)	254 (95%)	13 (5%)	0	100	100
1	B	255/400 (64%)	228 (89%)	25 (10%)	2 (1%)	24	40
1	C	249/400 (62%)	215 (86%)	29 (12%)	5 (2%)	9	14
All	All	771/1200 (64%)	697 (90%)	67 (9%)	7 (1%)	21	37

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	68	PRO
1	C	199	GLU
1	C	167	ASP
1	C	198	LEU
1	C	262	CYS

### 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	222/335 (66%)	205 (92%)	17 (8%)	16	29
1	B	217/335 (65%)	202 (93%)	15 (7%)	19	34
1	C	219/335 (65%)	193 (88%)	26 (12%)	6	11
All	All	658/1005 (66%)	600 (91%)	58 (9%)	12	22

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	155	LYS
1	C	12	GLN
1	C	253	PHE
1	B	187	LYS
1	B	222	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	275	HIS
1	B	130	GLN
1	C	70	GLN
1	A	147	ASN
1	C	69	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 carbohydrates in this entry.

## 5.6 Ligand geometry

2 ligands are 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	501	-	5,5,5	0.92	0	5,5,5	1.02	0
4	GOL	B	501	-	5,5,5	0.49	0	5,5,5	0.53	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	501	-	-	0/4/4/4	0/0/0/0
4	GOL	B	501	-	-	0/4/4/4	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	501	GOL	1	0

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

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	271/400 (67%)	0.43	18 (6%) 22 24	2, 2, 17, 39	0
1	B	261/400 (65%)	0.76	19 (7%) 18 20	2, 12, 29, 42	0
1	C	259/400 (64%)	1.56	77 (29%) 1 0	29, 54, 74, 93	0
2	D	26/26 (100%)	0.55	2 (7%) 16 18	4, 25, 48, 53	0
2	F	26/26 (100%)	1.53	10 (38%) 0 0	30, 55, 64, 68	0
3	E	26/26 (100%)	0.57	2 (7%) 16 18	8, 26, 36, 39	0
All	All	869/1278 (67%)	0.90	128 (14%) 3 3	2, 15, 65, 93	0

The worst 5 of 128 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	264	GLY	8.6
1	C	44	ALA	7.2
1	C	20	ALA	6.4
1	C	280	TRP	5.7
1	C	46	ARG	4.9

### 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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	GOL	B	501	6/6	0.80	0.48	11.42	64,73,79,79	0
4	GOL	A	501	6/6	0.72	0.52	5.77	58,61,63,78	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.