WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 07:16 am BST

PDB ID	:	5ITU
Title	:	Crystal Structure of Human NEIL1(242K) 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	:	2.41 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
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:

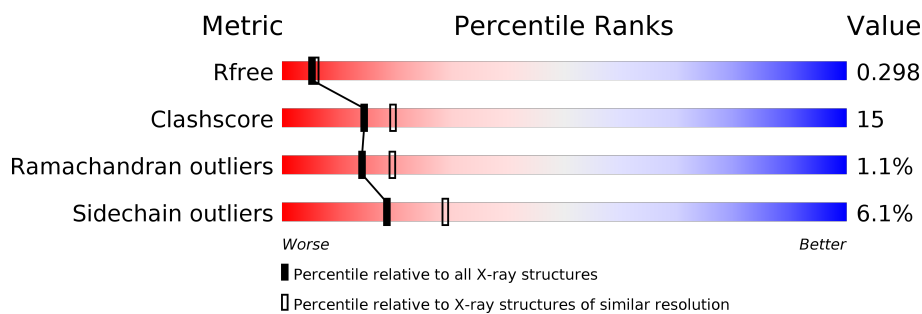|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.41 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.
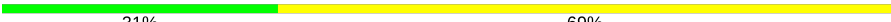
Metric	Percentile Ranks	Value
Rfree		0.298
Clashscore		15
Ramachandran outliers		1.1%
Sidechain outliers		6.1%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)

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 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 <=5%

Mol	Chain	Length	Quality of chain
1	A	400	53% 11% • • 33%
1	B	400	50% 13% • 36%
1	C	400	49% 17% • 32%
2	D	13	46% 54%
2	F	13	38% 62%
2	H	13	31% 69%
3	E	13	54% 38% 8%

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	13	31%　　　　　　　　　　　69%
3	I	13	23%　　　　　　　　　　　77%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 8023 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	268	Total 2120	C 1355	N 382	O 372	S 11	0	1	0
1	B	255	Total 2033	C 1302	N 371	O 350	S 10	0	0	0
1	C	271	Total 2143	C 1373	N 382	O 378	S 10	0	0	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
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	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	391	ALA	-	expression tag	UNP Q96FI4
C	392	ALA	-	expression tag	UNP Q96FI4
C	393	LEU	-	expression tag	UNP Q96FI4

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
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 (5'-D(*CP*GP*TP*CP*CP*AP*CP*GP*TP*CP*TP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	13	Total 249	C 120	N 41	O 76	P 12	0	0	0
2	F	13	Total 249	C 120	N 41	O 76	P 12	0	0	0
2	H	13	Total 249	C 120	N 41	O 76	P 12	0	0	0

- 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	E	13	Total 267	C 127	N 53	O 75	P 12	0	0	0
3	G	13	Total 267	C 127	N 53	O 75	P 12	0	0	0
3	I	13	Total 267	C 127	N 53	O 75	P 12	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	97	Total 97	O 97	0	0
4	B	58	Total 58	O 58	0	0
4	D	6	Total 6	O 6	0	0
4	E	8	Total 8	O 8	0	0

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	F	6	Total 6	O 6	0	0
4	G	4	Total 4	O 4	0	0

3 Residue-property plots ⓘ

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: Endonuclease 8-like 1

- Molecule 1: Endonuclease 8-like 1

- Molecule 1: Endonuclease 8-like 1

C113 I117 R118 W123 D124 L125 G126 GLY K128 W129 Q130 P131 G132 R133 L138 Y141 R145 E446 R150 N151 L152 ALA D154 R169 Y177 L178 R179 A180 Q202 GLN HIS ARG PRO SER PRO GLU LEU T211 L212 S213 Q214 K215 I216 ARG T218 N222 P223 E227 L228

V232 P233 K234 E235 V236 L239 K242 GLY Y244 G245 SER E247 F253 Y263 G267 M268 S269 S270 L271 Q272 D273 ARG HIS G276 R277 D284 P285 P290 LYS GLY ARG LYS SER ARG LYS LYS SER LYS ALA THR GLN LEU SER PRO GLU ASP ARG VAL GLU ASP ALA

LEU PRO PRO SER LYS ALA PRO SER ARG THR ARG ARG ALA LYS ARG ASP PRO LYS ARG THR ALA THR GLN ARG ARG GLU PRO GLU GLY THR SER LEU GLN GLN ASP PRO GLU ALA PRO THR VAL PRO LYS LYS GLY ARG ARG LYS GLY ARG GLN ALA ALA GLY SER HIS CYS ARG PRO ARG LYS

VAL LYS ALA ASP ILE PRO SER GLU LEU PRO GLU GLY THR SER ALA SER ALA ALA LEU GLY HIS HIS HIS HIS HIS HIS

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

Chain D: 46% 54%

C1 G2 T3 A6 T9 C10 T11 A12 C13

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

Chain F: 38% 62%

C1 G2 T3 C4 C5 A6 C7 C10 C13

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

Chain H: 31% 69%

C1 G2 T3 C4 C5 A6 C7 G8 T11 A12 C13

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

Chain E: 54% 38% 8%

T1 A2 G3 A4 I7 G8 G13

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

Chain G: 31% 69%

T1 A2 C6 T7 G8 G9 A10 C11 G12 G13

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

Chain I: 23% 77%

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	73.70Å 109.29Å 171.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	92.23 – 2.41 39.54 – 2.41	Depositor EDS
% Data completeness (in resolution range)	98.8 (92.23-2.41) 98.8 (39.54-2.41)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.06	Depositor
$<I/\sigma(I)>$ [1]	2.89 (at 2.42Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.210 , 0.260 0.262 , 0.298	Depositor DCC
R_{free} test set	2721 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	56.4	Xtriage
Anisotropy	0.347	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 46.3	EDS
L-test for twinning[2]	$<|L|> = 0.49$, $<L^2> = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	8023	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.59% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $<|L|>$, $<L^2>$ 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 ⓘ

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	1.06	2/2181 (0.1%)	1.24	16/2949 (0.5%)
1	B	0.94	1/2089 (0.0%)	1.15	15/2826 (0.5%)
1	C	0.53	0/2192	0.77	3/2952 (0.1%)
2	D	0.60	0/277	0.90	0/424
2	F	0.65	0/277	0.99	0/424
2	H	0.43	0/277	0.85	0/424
3	E	0.50	0/300	0.93	1/462 (0.2%)
3	G	0.58	0/300	0.92	0/462
3	I	0.33	0/300	0.77	0/462
All	All	0.81	3/8193 (0.0%)	1.03	35/11385 (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	3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	31	SER	CB-OG	-5.64	1.34	1.42
1	B	68	PRO	N-CD	5.14	1.55	1.47
1	A	45	TYR	CE2-CZ	5.04	1.45	1.38

The worst 5 of 35 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	133	ARG	NE-CZ-NH2	-18.24	111.18	120.30
1	B	133	ARG	NE-CZ-NH2	-16.44	112.08	120.30
1	A	133	ARG	NE-CZ-NH1	13.52	127.06	120.30
1	B	119	ARG	NE-CZ-NH2	-12.41	114.09	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	128	LYS	N-CA-C	11.17	141.15	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	248	SER	Peptide
1	A	67	GLN	Mainchain,Peptide

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	2120	0	2091	46	1
1	B	2033	0	2026	39	1
1	C	2143	0	2109	75	0
2	D	249	0	143	6	0
2	F	249	0	143	14	0
2	H	249	0	143	27	0
3	E	267	0	147	19	0
3	G	267	0	147	12	0
3	I	267	0	147	44	0
4	A	97	0	0	4	1
4	B	58	0	0	1	0
4	D	6	0	0	0	0
4	E	8	0	0	0	0
4	F	6	0	0	0	0
4	G	4	0	0	0	0
All	All	8023	0	7096	229	2

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 15.

The worst 5 of 229 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:C:118:ARG:NH2	2:H:6:DA:C2	1.83	1.45
1:C:118:ARG:NH2	2:H:6:DA:N1	1.74	1.35
1:C:117:ILE:CD1	3:I:8:DG:H5'	1.60	1.30
2:H:2:DG:H1	3:I:11:DC:N4	1.43	1.14
3:E:1:DT:H5"	3:I:13:DG:H2"	1.30	1.09

All (2) 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)
4:A:501:HOH:O	4:A:503:HOH:O[3_554]	1.99	0.21
1:A:284:ASP:OD2	1:B:95:ARG:NH2[3_454]	2.09	0.11

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	263/400 (66%)	249 (95%)	8 (3%)	6 (2%)	6	6
1	B	249/400 (62%)	232 (93%)	16 (6%)	1 (0%)	34	47
1	C	249/400 (62%)	217 (87%)	31 (12%)	1 (0%)	34	47
All	All	761/1200 (63%)	698 (92%)	55 (7%)	8 (1%)	14	19

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	68	PRO
1	A	105	PRO
1	A	200	ALA
1	A	250	GLU
1	A	251	GLU

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%)	211 (95%)	11 (5%)	24	38
1	B	214/335 (64%)	205 (96%)	9 (4%)	30	46
1	C	225/335 (67%)	205 (91%)	20 (9%)	9	14
All	All	661/1005 (66%)	621 (94%)	40 (6%)	18	29

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	265	MET
1	C	43	SER
1	C	239	LEU
1	C	12	GLN
1	C	63	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	12	GLN
1	A	139	GLN
1	B	12	GLN
1	B	275	HIS
1	C	151	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 (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 (i)

6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.