



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2026 – 03:28 AM UTC

PDB ID : 5ITR / pdb_00005itr
Title : Crystal Structure of Human NEIL1(P2G) 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 : 2.46 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 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
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

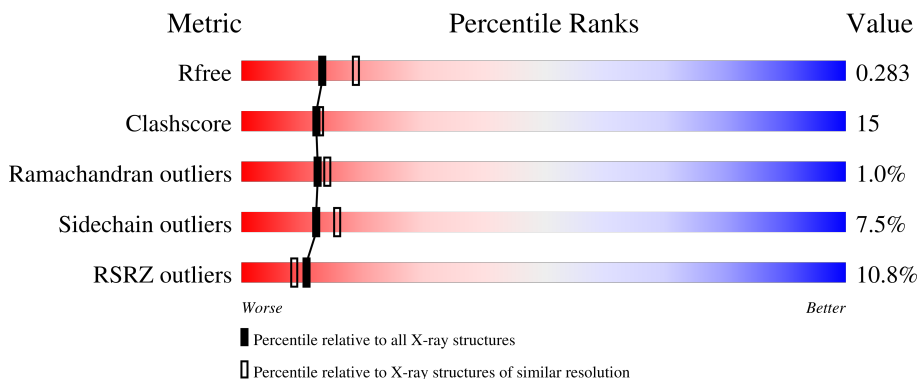
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.46 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	1190 (2.46-2.46)
Clashscore	190562	1229 (2.46-2.46)
Ramachandran outliers	187476	1218 (2.46-2.46)
Sidechain outliers	187428	1218 (2.46-2.46)
RSRZ outliers	180081	1190 (2.46-2.46)

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	400	
1	B	400	
1	C	400	
2	D	26	
2	E	26	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	26	 54% 46%

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 7925 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
			Total	C	N	O	S			
1	A	270	2146	1369	393	374	10	0	2	0
1	B	261	2073	1324	381	358	10	0	0	0
1	C	254	2032	1302	371	349	10	0	0	0

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	E	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	F	26	Total	C	N	O	P	0	0	0
			517	247	94	152	24			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	59	Total	O	0	0
			59	59		
4	B	45	Total	O	0	0
			45	45		
4	C	4	Total	O	0	0
			4	4		
4	D	9	Total	O	0	0
			9	9		
4	E	8	Total	O	0	0
			8	8		

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	73.31Å 109.04Å 169.84Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	91.75 – 2.46 91.75 – 2.46	Depositor EDS
% Data completeness (in resolution range)	99.4 (91.75-2.46) 99.4 (91.75-2.46)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.59 (at 2.45Å)	Xtrriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.211 , 0.263 0.243 , 0.283	Depositor DCC
R_{free} test set	2488 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å ²)	55.4	Xtrriage
Anisotropy	0.521	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 47.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	7925	wwPDB-VP
Average B, all atoms (Å ²)	59.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.03% 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](#)

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	1.32	7/2211 (0.3%)	1.27	13/2989 (0.4%)
1	B	1.28	14/2129 (0.7%)	1.26	13/2879 (0.5%)
1	C	0.68	1/2085 (0.0%)	0.93	5/2816 (0.2%)
2	D	0.45	0/577	0.85	0/886
2	E	0.47	0/577	0.95	1/886 (0.1%)
3	F	0.37	0/578	0.76	0/888
All	All	1.03	22/8157 (0.3%)	1.10	32/11344 (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

The worst 5 of 22 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	10	ALA	CA-C	7.18	1.62	1.52
1	A	47	ILE	C-O	6.93	1.31	1.24
1	B	83	GLY	C-O	-6.59	1.15	1.23
1	A	62	PRO	C-O	-6.08	1.16	1.23
1	B	188	ILE	C-N	6.07	1.40	1.33

The worst 5 of 32 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	104	ALA	CA-C-N	11.77	132.50	120.38
1	B	104	ALA	C-N-CA	11.77	132.50	120.38
1	A	104	ALA	CA-C-N	10.33	131.02	120.38

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	104	ALA	C-N-CA	10.33	131.02	120.38
1	A	105	PRO	N-CA-C	9.75	122.60	110.70

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	67	GLN	Peptide,Mainchain
1	B	67	GLN	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	2146	0	2109	85	0
1	B	2073	0	2055	72	0
1	C	2032	0	2016	56	0
2	D	516	0	290	6	0
2	E	516	0	290	6	0
3	F	517	0	290	10	0
4	A	59	0	0	1	2
4	B	45	0	0	1	2
4	C	4	0	0	0	0
4	D	9	0	0	1	0
4	E	8	0	0	0	0
All	All	7925	0	7050	228	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 228 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:191:PHE:CE2	1:B:289:ALA:HB2	1.16	1.58
1:B:191:PHE:CE2	1:B:289:ALA:CB	1.97	1.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:191:PHE:CZ	1:B:289:ALA:CB	2.10	1.32
1:B:191:PHE:CZ	1:B:289:ALA:HB2	1.75	1.16
1:B:191:PHE:CD2	1:B:289:ALA:HB2	1.81	1.15

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:551:HOH:O	4:B:525:HOH:O[3_555]	1.81	0.39
4:A:558:HOH:O	4:B:537:HOH:O[3_555]	1.99	0.21

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	268/400 (67%)	251 (94%)	15 (6%)	2 (1%)	18	24
1	B	255/400 (64%)	231 (91%)	22 (9%)	2 (1%)	16	20
1	C	244/400 (61%)	215 (88%)	25 (10%)	4 (2%)	7	7
All	All	767/1200 (64%)	697 (91%)	62 (8%)	8 (1%)	12	14

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	68	PRO
1	B	68	PRO
1	B	200	ALA
1	C	36	PRO
1	C	97	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	222/335 (66%)	206 (93%)	16 (7%)	13	17
1	B	216/335 (64%)	201 (93%)	15 (7%)	14	19
1	C	214/335 (64%)	196 (92%)	18 (8%)	10	12
All	All	652/1005 (65%)	603 (92%)	49 (8%)	12	16

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	269	SER
1	C	95	ARG
1	B	272	GLN
1	C	69	GLN
1	C	122	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	130	GLN
1	C	151	ASN
1	B	143	GLN
1	C	282	GLN
1	C	139	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 [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, 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	270/400 (67%)	0.06	15 (5%) 30 27	13, 31, 61, 85	2 (0%)
1	B	261/400 (65%)	0.37	14 (5%) 31 29	19, 41, 74, 124	0
1	C	254/400 (63%)	1.41	62 (24%) 2 1	48, 88, 121, 149	0
2	D	26/26 (100%)	0.35	0 100 100	34, 67, 110, 118	0
2	E	26/26 (100%)	0.48	2 (7%) 19 17	29, 64, 79, 85	0
3	F	26/26 (100%)	0.97	0 100 100	68, 98, 111, 116	0
All	All	863/1278 (67%)	0.60	93 (10%) 11 9	13, 50, 110, 149	2 (0%)

The worst 5 of 93 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	191	PHE	9.5
1	A	33	SER	6.7
1	C	147	ASN	5.8
1	C	220	LEU	5.5
1	C	22	VAL	4.8

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.