WORLDWIDE PDB
PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report ⓘ

Mar 6, 2022 – 08:30 AM EST

PDB ID	:	2KH0
Title	:	2-Hydroxy-7-nitrofluorene covalently linked into a 13mer DNA duplex - solution structure of the face-down orientation
Authors	:	Dallmann, A.; Pfaffe, M.; Muegge, C.; Mahrwald, R.; Kovalenko, S.A.; Ernsting, N.P.
Deposited on	:	2009-03-23

This is a Full wwPDB NMR Structure Validation 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/NMRValidationReportHelp
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:

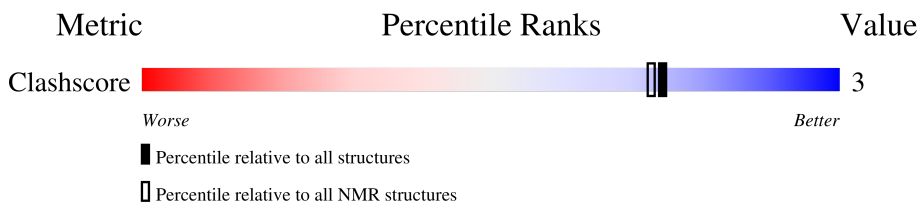|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
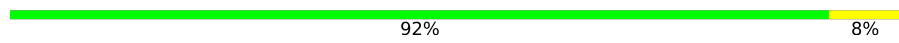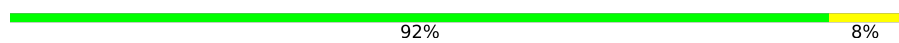*SOLUTION NMR*

The overall completeness of chemical shifts assignment was not calculated.

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
Clashscore	*Worse* *Better*	3

▌ Percentile relative to all structures

▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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	13	92% 8%
2	B	13	92% 8%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2 Ensemble composition and analysis ⓘ

This entry contains 11 models. This entry does not contain polypeptide chains, therefore identification of well-defined residues and clustering analysis are not possible. All residues are included in the validation scores.

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 820 atoms, of which 295 are hydrogens and 0 are deuteriums.
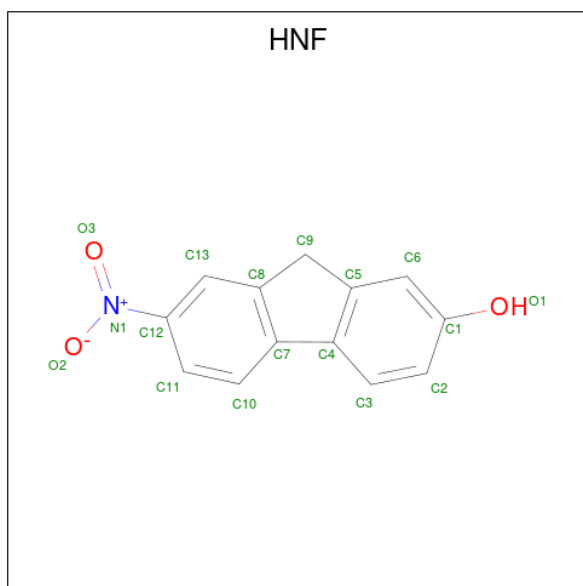
- Molecule 1 is a DNA chain called 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'.

Mol	Chain	Residues	Atoms						Trace
1	A	13	Total	C	H	N	O	P	0
			397	121	143	46	75	12	

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

Mol	Chain	Residues	Atoms						Trace
2	B	13	Total	C	H	N	O	P	0
			398	121	144	46	75	12	

- Molecule 3 is 7-nitro-9H-fluoren-2-ol (three-letter code: HNF) (formula: $C_{13}H_9NO_3$).

Mol	Chain	Residues	Atoms				
3	A	1	Total	C	H	N	O
			25	13	8	1	3

4 Residue-property plots ⓘ

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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 outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

- Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A:

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

Chain B:

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

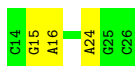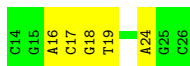### 4.2.1 Score per residue for model 1

- Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A:

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

Chain B:

4.2.2 Score per residue for model 2

• Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A:

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

Chain B:

4.2.3 Score per residue for model 3

• Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A:

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

Chain B:

4.2.4 Score per residue for model 4

• Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A:

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

Chain B:

4.2.5 Score per residue for model 5

• Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A:

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

Chain B:

4.2.6 Score per residue for model 6

• Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A:

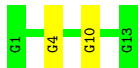• Molecule 2: 5'-D(*CP*GP*AP*CP*GP*TP*(3DR)P*TP*GP*CP*AP*GP*C)-3'

Chain B:

4.2.7 Score per residue for model 7
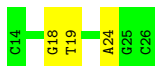
• Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A:

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

Chain B:

4.2.8 Score per residue for model 8

- Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A: 77%   23%

G1 C2 T3 G4 G13

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

Chain B: 92%   8%

C14 A24 G25 C26

4.2.9 Score per residue for model 9

- Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A: 77%   23%

G1 G4 A8 C9 G13

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

Chain B: 92%   8%

C14 A24 G25 C26

4.2.10 Score per residue for model 10

- Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'

Chain A: 77%   23%

G1 G4 C5 A6 G10 G13

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

Chain B: 62%   38%

C14 G15 A16 C17 G18 G22 C23 A24 G25 C26
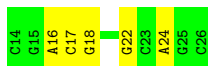
4.2.11 Score per residue for model 11

• Molecule 1: 5'-D(*GP*CP*TP*GP*CP*AP*(3DR)P*AP*CP*GP*TP*CP*G)-3'
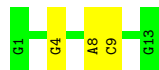
Chain A:

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

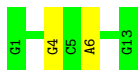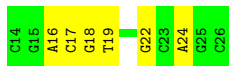Chain B:

5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *simulated annealing, torsion angle dynamics.*

Of the 100 calculated structures, 11 were deposited, based on the following criterion: *structures with the lowest energy.*

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
X-PLOR NIH	geometry optimization	2.20
X-PLOR NIH	refinement	2.20
X-PLOR NIH	structure solution	2.21

No chemical shift data was provided.

6 Model quality ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 3DR, HNF

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	254	143	143	2±1
2	B	254	144	144	2±1
All	All	5775	3245	3245	29

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 3.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:4:DG:C2	2:B:24:DA:C2	0.53	2.97	6	11
1:A:10:DG:C2	2:B:18:DG:C2	0.45	3.04	10	2
2:B:16:DA:C2	2:B:17:DC:C2	0.45	3.05	6	3
2:B:18:DG:C2	2:B:19:DT:C2	0.43	3.07	5	3
1:A:1:DG:C4	1:A:2:DC:C5	0.42	3.08	7	3
1:A:6:DA:C2	2:B:22:DG:C2	0.42	3.07	6	2
1:A:8:DA:C2	1:A:9:DC:C2	0.42	3.07	9	2
2:B:15:DG:C6	2:B:16:DA:C6	0.41	3.09	11	2
1:A:5:DC:N4	1:A:6:DA:N6	0.40	2.70	7	1

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

There are no protein molecules in this entry.

6.3.2 Protein sidechains ⓘ

There are no protein molecules in this entry.

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds 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 is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
1	3DR	A	7	1,3	8,11,12	0.31±0.02	0±0 (0±0%)
2	3DR	B	20	2	8,11,12	0.33±0.01	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of 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 angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
1	3DR	A	7	1,3	9,14,17	0.72±0.04	0±0 (0±0%)
2	3DR	B	20	2	9,14,17	0.70±0.04	0±0 (0±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
2	3DR	B	20	2	-	0±0,3,15,16	0±0,1,1,1
1	3DR	A	7	1,3	-	0±0,3,15,16	0±0,1,1,1

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.

6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds 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 is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
3	HNF	A	27	1	18,19,19	0.94±0.01	1±0 (5±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of 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 angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
3	HNF	A	27	1	26,28,28	0.76±0.01	0±0 (0±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
3	HNF	A	27	1	-	0±0,2,12,12	0±0,3,3,3

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
3	A	27	HNF	O1-C1	3.53	1.45	1.37	1	11

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

7 Chemical shift validation (i)

No chemical shift data were provided