# wwPDB NMR Structure Validation Summary Report (i)

Oct 24, 2016 – 10:57 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 5KIH |
| Title | : | Structural impact of single ribonucleotides in DNA |
| Authors | : | Evich, M.; Spring-Connell, A.M.; Storici, F.; Germann, M.W. |
| Deposited on | : | 2016-06-16 |

This is a wwPDB NMR 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
http://wwpdb.org/validation/2016/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

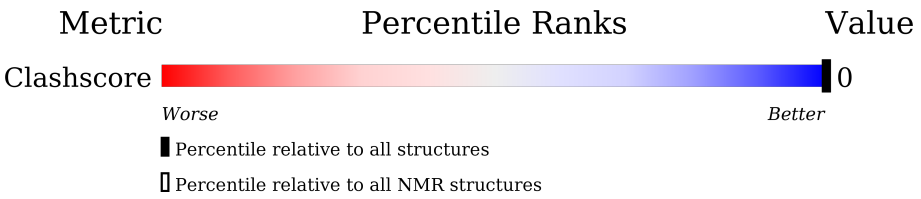
| | | |
|---|---|---|
| Cyrange | : | Kirchner and Güntert (2011) |
| NmrClust | : | Kelley et al. (1996) |
| MolProbity | : | 4.02b-467 |
| Mogul | : | unknown |
| Percentile statistics | : | 20151230.v01 (using entries in the PDB archive December 30th 2015) |
| RCI | : | v_1n_11_5_13_A (Berjanski et al., 2005) |
| PANAV | : | Wang et al. (2010) |
| ShiftChecker | : | rb-20027939 |
| 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 (i)

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

The overall completeness of chemical shifts assignment is 33%.

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) | NMR archive (#Entries) |
|---|---|---|
| Clashscore | 114402 | 11133 |

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 | 9 | 67% 33% |
| 2 | B | 9 | 78% 22% |

## 2 Ensemble composition and analysis (i)

This entry contains 2 models. Identification of well-defined residues and clustering analysis are not possible.

# 3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 570 atoms, of which 206 are hydrogens and 0 are deuteriums.

- Molecule 1 is DNA/RNA hybrid called DNA/RNA (5'-D(*TP*TP*AP*G)-R(P*G)-D(P*CP*CP*TP*G)-3').

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | P | |
| 1 | A | 9 | 287 | 88 | 104 | 32 | 55 | 8 | 0 |

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

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | P | |
| 2 | B | 9 | 283 | 87 | 102 | 36 | 50 | 8 | 0 |

# 4 Residue-property plots (i)

## 4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: DNA/RNA (5'-D(*TP*TP*AP*G)-R(P*G)-D(P*CP*CP*TP*G)-3')

Chain A: 67% 33%

T1 T2 A3 G4 G5 C6 C7 T8 G9

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

Chain B: 78% 22%

C10 A11 G12 G13 C14 C15 T16 A17 A18

## 4.2 Residue scores for the representative (author defined) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: DNA/RNA (5'-D(*TP*TP*AP*G)-R(P*G)-D(P*CP*CP*TP*G)-3')

Chain A: 56% 44%

T1 T2 A3 G4 G5 C6 C7 T8 G9

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

Chain B: 89% 11%

C10 A11 G12 G13 C14 C15 T16 A17 A18

# 5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *matrix relaxation, molecular dynamics.*

Of the 6 calculated structures, 2 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 |
|---|---|---|
| CORMA | refinement | |
| MARDIGRAS | refinement | |

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

| | |
|---|---|
| Chemical shift file(s) | 5kih_cs.cif |
| Number of chemical shift lists | 1 |
| Total number of shifts | 136 |
| Number of shifts mapped to atoms | 136 |
| Number of unparsed shifts | 0 |
| Number of shifts with mapping errors | 0 |
| Number of shifts with mapping warnings | 0 |
| Assignment completeness (well-defined parts) | 33% |

No validations of the models with respect to experimental NMR restraints is performed at this time.

# 6 Model quality (i)

## 6.1 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 (average) 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 | 2.42±0.06 | 12±2/204 (5.9±1.0%) | 3.01±0.07 | 35±2/314 (11.0±0.5%) |
| 2 | B | 3.68±0.76 | 27±12/203 (13.1±5.7%) | 3.27±0.02 | 44±1/311 (14.1±0.3%) |
| All | All | 3.16 | 77/814 (9.5%) | 3.15 | 157/1250 (12.6%) |

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 | Planarity |
|---|---|---|---|
| 1 | A | 0.0±0.0 | 2.5±1.5 |
| 2 | B | 0.0±0.0 | 1.5±0.5 |
| All | All | 0 | 8 |

5 of 69 unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) | Models | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Worst | Total |
| 2 | B | 18 | DA | C6-N1 | -17.72 | 1.23 | 1.35 | 1 | 2 |
| 2 | B | 18 | DA | C3'-C2' | -16.41 | 1.32 | 1.52 | 1 | 1 |
| 2 | B | 10 | DC | N1-C6 | -15.88 | 1.27 | 1.37 | 2 | 1 |
| 2 | B | 16 | DT | C3'-C2' | -14.56 | 1.34 | 1.52 | 1 | 1 |
| 2 | B | 17 | DA | N9-C8 | -13.37 | 1.27 | 1.37 | 1 | 1 |

5 of 121 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

| Mol | Chain | Res | Type | Atoms | Z | Observed(°) | Ideal(°) | Models | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Worst | Total |
| 1 | A | 1 | DT | O4'-C4'-C3' | 12.73 | 113.64 | 106.00 | 2 | 1 |
| 2 | B | 11 | DA | C5-C6-N1 | 12.53 | 123.96 | 117.70 | 1 | 2 |
| 2 | B | 18 | DA | N1-C6-N6 | -11.87 | 111.48 | 118.60 | 2 | 2 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) | Models | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Worst | Total |
| 2 | B | 10 | DC | O4'-C1'-N1 | 11.86 | 116.30 | 108.00 | 2 | 2 |
| 2 | B | 11 | DA | N1-C6-N6 | -11.62 | 111.63 | 118.60 | 1 | 2 |

There are no chirality outliers.

5 of 6 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

| Mol | Chain | Res | Type | Group | Models (Total) |
|---|---|---|---|---|---|
| 2 | B | 10 | DC | Sidechain | 2 |
| 1 | A | 5 | G | Sidechain | 2 |
| 1 | A | 4 | DG | Sidechain | 1 |
| 1 | A | 9 | DG | Sidechain | 1 |
| 1 | A | 6 | DC | Sidechain | 1 |

## 6.2 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 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 |
|---|---|---|---|---|---|
| All | All | 728 | 412 | 395 | - |

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 -.

There are no clashes.

## 6.3 Torsion angles (i)

### 6.3.1 Protein backbone (i)

There are no protein molecules in this entry.

### 6.3.2 Protein sidechains (i)

There are no protein molecules in this entry.

#### 6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

### 6.5 Carbohydrates (i)

There are no carbohydrates in this entry.

### 6.6 Ligand geometry (i)

There are no ligands in this entry.

### 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.

# 7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 33% for the well-defined parts and 33% for the entire structure.

## 7.1 Chemical shift list 1

File name: 5kih_cs.cif

Chemical shift list name: *rGGxCshiftsout.str*

### 7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

| | |
|---|---|
| Total number of shifts | 136 |
| Number of shifts mapped to atoms | 136 |
| Number of unparsed shifts | 0 |
| Number of shifts with mapping errors | 0 |
| Number of shifts with mapping warnings | 0 |
| Number of shift outliers (ShiftChecker) | 0 |

### 7.1.2 Chemical shift referencing (i)

No chemical shift referencing corrections were calculated (not enough data).

### 7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 33%, i.e. 116 atoms were assigned a chemical shift out of a possible 356. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

| | Total | $^{1}$H | $^{13}$C | $^{15}$N |
|---|---|---|---|---|
| Backbone | 0/0 (—%) | 0/0 (—%) | 0/0 (—%) | 0/0 (—%) |
| Sidechain | 0/0 (—%) | 0/0 (—%) | 0/0 (—%) | 0/0 (—%) |
| Aromatic | 0/0 (—%) | 0/0 (—%) | 0/0 (—%) | 0/0 (—%) |
| Overall | 116/356 (33%) | 116/212 (55%) | 0/121 (0%) | 0/23 (0%) |

### 7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

### 7.1.5 Random Coil Index (RCI) plots (i)

No *random coil index* (RCI) plot could be generated from the current chemical shift list (rGGxC-shiftsout.str). RCI is only applicable to proteins.