# wwPDB NMR Structure Validation Summary Report

Apr 20, 2024 – 09:02 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 2KY8 |
| BMRB ID | : | 16936 |
| Title | : | Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target methylated DNA sequence |
| Authors | : | Williams Jr., D.C.; Scarsdale Jr., J.N. |
| Deposited on | : | 2010-05-18 |

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) 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 (1)) were used in the production of this report:

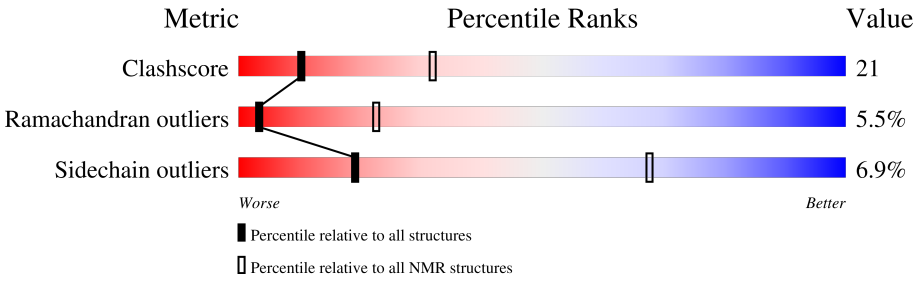
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| wwPDB-RCI | : | v_1n_11_5_13_A (Berjanski et al., 2005) |
| PANAV | : | Wang et al. (2010) |
| wwPDB-ShiftChecker | : | v1.2 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.36.2 |

# 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 64%.

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 | 158937 | 12864 |
| Ramachandran outliers | 154571 | 11451 |
| Sidechain outliers | 154315 | 11428 |

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 | 72 | 65% 24% · 6% · |
| 2 | B | 11 | 64% 36% |
| 3 | C | 11 | 45% 55% |

## 2 Ensemble composition and analysis

This entry contains 20 models. Model 13 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

| Well-defined (core) protein residues | | | |
|---|---|---|---|
| **Well-defined core** | **Residue range (total)** | **Backbone RMSD (Å)** | **Medoid model** |
| 1 | A:6-A:71 (66) | 0.82 | 13 |

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 3 clusters and 2 single-model clusters were found.

| Cluster number | Models |
|---|---|
| 1 | 1, 2, 7, 9, 11, 13, 14, 16, 17 |
| 2 | 3, 4, 5, 6, 15, 19 |
| 3 | 12, 18, 20 |
| Single-model clusters | 8; 10 |

# 3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2081 atoms, of which 923 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Methyl-CpG-binding domain protein 2.

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 1 | A | 70 | Total | C | H | N | O | S | 0 |
| | | | 1104 | 349 | 555 | 98 | 98 | 4 | |

There are 2 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 1 | GLY | - | expression tag | UNP Q5EFL0 |
| A | 2 | SER | - | expression tag | UNP Q5EFL0 |

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

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 2 | B | 11 | Total | C | H | N | O | P | 0 |
| | | | 487 | 150 | 184 | 55 | 88 | 10 | |

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

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 3 | C | 11 | Total | C | H | N | O | P | 0 |
| | | | 484 | 149 | 184 | 53 | 88 | 10 | |

- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

| Mol | Chain | Residues | Atoms | |
|---|---|---|---|---|
| 4 | B | 1 | Total | Mn |
| | | | 3 | 3 |
| 4 | C | 1 | Total | Mn |
| | | | 3 | 3 |

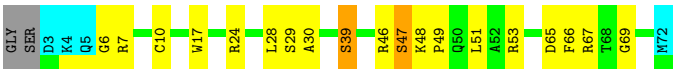
# 4 Residue-property plots

## 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.

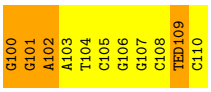
• Molecule 1: Methyl-CpG-binding domain protein 2

Chain A:


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

Chain B:


• Molecule 3: DNA (5'-D(*GP*AP*GP*CP*(5CM)P*GP*AP*TP*(TED)P*CP*C)-3')

Chain C:

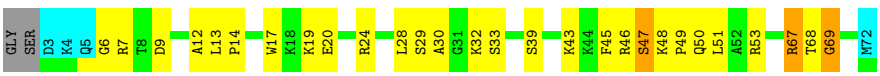


## 4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 13. Colouring as in section 4.1 above.

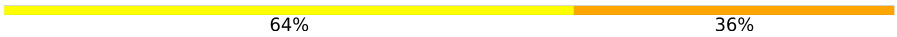
• Molecule 1: Methyl-CpG-binding domain protein 2

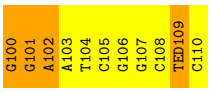
Chain A:


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

Chain B:




• Molecule 3: DNA (5'-D(*GP*AP*GP*CP*(5CM)P*GP*AP*TP*(TED)P*CP*C)-3')

Chain C:


# 5 Refinement protocol and experimental data overview

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

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *all calculated structures submitted*.

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

| Software name | Classification | Version |
|---|---|---|
| X-PLOR NIH | 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) | working_cs.cif |
|---|---|
| Number of chemical shift lists | 1 |
| Total number of shifts | 959 |
| Number of shifts mapped to atoms | 852 |
| Number of unparsed shifts | 0 |
| Number of shifts with mapping errors | 107 |
| Number of shifts with mapping warnings | 0 |
| Assignment completeness (well-defined parts) | 64% |

# 6 Model quality

## 6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: TED, MN, 5CM

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 | 1.02±0.00 | 0±0/530 ( 0.0± 0.0%) | 0.89±0.00 | 0±0/711 ( 0.0± 0.0%) |
| 2 | B | 1.51±0.04 | 7±1/206 ( 3.5± 0.4%) | 2.13±0.01 | 11±0/312 ( 3.6± 0.1%) |
| 3 | C | 1.52±0.03 | 7±1/202 ( 3.5± 0.5%) | 2.13±0.01 | 11±0/305 ( 3.6± 0.1%) |
| All | All | 1.26 | 287/18760 ( 1.5%) | 1.59 | 442/26560 ( 1.7%) |

5 of 19 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 | 110 | DC | O4'-C1' | 7.45 | 1.51 | 1.42 | 15 | 20 |
| 3 | C | 118 | DT | O4'-C1' | 7.19 | 1.50 | 1.42 | 8 | 20 |
| 3 | C | 120 | DC | O4'-C1' | 6.93 | 1.50 | 1.42 | 8 | 20 |
| 2 | B | 100 | DG | O4'-C1' | 6.90 | 1.50 | 1.42 | 6 | 15 |
| 3 | C | 111 | DG | O4'-C1' | 6.83 | 1.50 | 1.42 | 12 | 20 |

5 of 24 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($^o$) | Ideal($^o$) | Models | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Worst | Total |
| 3 | C | 120 | DC | C4'-C3'-C2' | 8.81 | 111.03 | 103.10 | 12 | 20 |
| 3 | C | 117 | DA | C4'-C3'-C2' | 8.79 | 111.01 | 103.10 | 18 | 20 |
| 3 | C | 111 | DG | C4'-C3'-C2' | 8.79 | 111.01 | 103.10 | 17 | 20 |
| 2 | B | 100 | DG | C4'-C3'-C2' | 8.78 | 111.00 | 103.10 | 2 | 20 |
| 2 | B | 107 | DG | C4'-C3'-C2' | 8.78 | 111.00 | 103.10 | 19 | 20 |

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 | 516 | 521 | 521 | 27±5 |
| 2 | B | 303 | 184 | 175 | 9±3 |
| 3 | C | 300 | 184 | 175 | 10±3 |
| All | All | 22500 | 17780 | 17420 | 849 |

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 21.

5 of 275 unique clashes are listed below, sorted by their clash magnitude.

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 3:C:111:DG:HO5' | 3:C:111:DG:H8 | 0.96 | 0.99 | 18 | 20 |
| 2:B:100:DG:HO5' | 2:B:100:DG:H8 | 0.95 | 0.95 | 11 | 20 |
| 1:A:28:LEU:O | 1:A:28:LEU:HD13 | 0.83 | 1.73 | 7 | 2 |
| 1:A:26:SER:O | 1:A:27:GLY:O | 0.81 | 1.99 | 18 | 2 |
| 1:A:48:LYS:N | 1:A:49:PRO:HD2 | 0.80 | 1.92 | 8 | 20 |

## 6.3 Torsion angles

### 6.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. 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 | 66/72 (92%) | 58±2 (88±2%) | 4±2 (7±3%) | 4±1 (5±2%) | 3 | 23 |
| All | All | 1320/1440 (92%) | 1161 (88%) | 87 (7%) | 72 (5%) | 3 | 23 |

5 of 12 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

| Mol | Chain | Res | Type | Models (Total) |
|---|---|---|---|---|
| 1 | A | 47 | SER | 19 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type | Models (Total) |
|---|---|---|---|---|
| 1 | A | 29 | SER | 18 |
| 1 | A | 6 | GLY | 9 |
| 1 | A | 14 | PRO | 5 |
| 1 | A | 58 | ALA | 4 |

### 6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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 | 55/60 (92%) | 51±1 (93±2%) | 4±1 (7±2%) | 19 | 68 |
| All | All | 1100/1200 (92%) | 1024 (93%) | 76 (7%) | 19 | 68 |

5 of 19 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

| Mol | Chain | Res | Type | Models (Total) |
|---|---|---|---|---|
| 1 | A | 39 | SER | 17 |
| 1 | A | 10 | CYS | 9 |
| 1 | A | 47 | SER | 7 |
| 1 | A | 67 | ARG | 7 |
| 1 | A | 48 | LYS | 6 |

### 6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

8 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 |
| 3 | TED | C | 119[C] | 4 | 44,47,48 | 0.96±0.02 | 4±0 (9±0%) |
| 3 | TED | C | 119[B] | 4 | 44,47,48 | 0.96±0.02 | 4±0 (9±0%) |
| 2 | TED | B | 109[B] | 4 | 44,47,48 | 0.95±0.01 | 4±0 (9±0%) |
| 3 | TED | C | 119[A] | 4 | 44,47,48 | 0.96±0.02 | 4±0 (9±0%) |
| 2 | TED | B | 109[A] | 4 | 44,47,48 | 0.95±0.01 | 4±0 (9±0%) |
| 3 | 5CM | C | 115 | 2,3 | 17,21,22 | 0.75±0.03 | 1±0 (5±0%) |
| 2 | TED | B | 109[C] | 4 | 44,47,48 | 0.95±0.01 | 4±0 (9±0%) |
| 2 | 5CM | B | 105 | 2,3 | 17,21,22 | 0.79±0.03 | 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 | TED | C | 119[C] | 4 | 56,63,66 | 1.44±0.02 | 11±1 (19±0%) |
| 3 | TED | C | 119[B] | 4 | 56,63,66 | 1.42±0.01 | 10±0 (18±0%) |
| 2 | TED | B | 109[B] | 4 | 56,63,66 | 1.42±0.01 | 10±1 (18±1%) |
| 3 | TED | C | 119[A] | 4 | 56,63,66 | 1.43±0.02 | 10±1 (18±1%) |
| 2 | TED | B | 109[A] | 4 | 56,63,66 | 1.41±0.01 | 10±1 (17±1%) |
| 3 | 5CM | C | 115 | 2,3 | 24,30,33 | 1.14±0.01 | 2±0 (8±0%) |
| 2 | TED | B | 109[C] | 4 | 56,63,66 | 1.42±0.02 | 10±1 (18±0%) |
| 2 | 5CM | B | 105 | 2,3 | 24,30,33 | 1.14±0.01 | 2±0 (8±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 | 5CM | C | 115 | 2,3 | - | 0±0,7,21,22 | 0±0,2,2,2 |
| 2 | TED | B | 109[B] | 4 | - | 0±0,40,54,55 | 0±0,2,2,2 |
| 3 | TED | C | 119[B] | 4 | - | 0±0,40,54,55 | 0±0,2,2,2 |
| 2 | TED | B | 109[C] | 4 | - | 0±0,40,54,55 | 0±0,2,2,2 |
| 3 | TED | C | 119[C] | 4 | - | 0±0,40,54,55 | 0±0,2,2,2 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Type | Chain | Res | Link | Chirals | Torsions | Rings |
|---|---|---|---|---|---|---|---|
| 2 | 5CM | B | 105 | 2,3 | - | 0±0,7,21,22 | 0±0,2,2,2 |
| 2 | TED | B | 109[A] | 4 | - | 0±0,40,54,55 | 0±0,2,2,2 |
| 3 | TED | C | 119[A] | 4 | - | 0±0,40,54,55 | 0±0,2,2,2 |

5 of 26 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 |
| 3 | C | 119[A] | TED | O4'-C1' | 3.27 | 1.49 | 1.42 | 14 | 20 |
| 3 | C | 119[B] | TED | O4'-C1' | 3.27 | 1.49 | 1.42 | 14 | 20 |
| 3 | C | 119[C] | TED | O4'-C1' | 3.27 | 1.49 | 1.42 | 14 | 20 |
| 2 | B | 109[A] | TED | O4'-C1' | 3.13 | 1.49 | 1.42 | 15 | 20 |
| 2 | B | 109[B] | TED | O4'-C1' | 3.13 | 1.49 | 1.42 | 15 | 20 |

5 of 72 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($^o$) | Ideal($^o$) | Models | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Worst | Total |
| 2 | B | 105 | 5CM | C2'-C3'-C4' | 3.95 | 110.99 | 102.76 | 15 | 20 |
| 3 | C | 115 | 5CM | C2'-C3'-C4' | 3.95 | 110.98 | 102.76 | 18 | 20 |
| 2 | B | 109[A] | TED | C2'-C3'-C4' | 3.94 | 110.97 | 102.76 | 6 | 20 |
| 2 | B | 109[B] | TED | C2'-C3'-C4' | 3.94 | 110.97 | 102.76 | 6 | 20 |
| 2 | B | 109[C] | TED | C2'-C3'-C4' | 3.94 | 110.97 | 102.76 | 6 | 20 |

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

## 6.6 Ligand geometry (i)

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

## 6.7 Other polymers (i)

There are no such molecules in this entry.

## 6.8 Polymer linkage issues (i)

The following chains have linkage breaks:

| Mol | Chain | Number of breaks |
|---|---|---|
| 3 | 10-C | 3 |
| 3 | 8-C | 3 |
| 3 | 13-C | 3 |
| 3 | 17-C | 3 |
| 3 | 11-C | 3 |
| 3 | 6-C | 3 |
| 3 | 14-C | 3 |
| 3 | 18-C | 3 |
| 3 | 16-C | 3 |
| 3 | 4-C | 3 |
| 3 | 15-C | 3 |
| 3 | 12-C | 3 |
| 2 | 17-B | 3 |
| 2 | 13-B | 3 |
| 3 | 7-C | 3 |
| 3 | 5-C | 3 |
| 3 | 3-C | 3 |
| 2 | 3-B | 3 |
| 2 | 2-B | 3 |
| 3 | 20-C | 3 |
| 3 | 1-C | 3 |
| 2 | 8-B | 3 |
| 3 | 19-C | 3 |
| 3 | 2-C | 3 |
| 2 | 18-B | 3 |
| 2 | 10-B | 3 |
| 2 | 9-B | 3 |
| 2 | 11-B | 3 |
| 2 | 1-B | 3 |
| 2 | 6-B | 3 |
| 3 | 9-C | 3 |
| 2 | 4-B | 3 |
| 2 | 19-B | 3 |
| 2 | 7-B | 3 |
| 2 | 5-B | 3 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Number of breaks |
|---|---|---|
| 2 | 20-B | 3 |
| 2 | 12-B | 3 |
| 2 | 16-B | 3 |
| 2 | 14-B | 3 |
| 2 | 15-B | 3 |

The worst 5 of 120 chain breaks are listed below:

| Model | Chain | Residue-1 | Atom-1 | Residue-2 | Atom-2 | Distance (Å) |
|---|---|---|---|---|---|---|
| 10 | C | 119[C]:TED | OY6 | 120:DC | P | 20.80 |
| 8 | C | 119[B]:TED | OY4 | 120:DC | P | 19.91 |
| 13 | C | 119[C]:TED | OY6 | 120:DC | P | 19.06 |
| 17 | C | 119[B]:TED | OY6 | 120:DC | P | 18.52 |
| 11 | C | 119[C]:TED | OY6 | 120:DC | P | 18.51 |

# 7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 64% for the well-defined parts and 65% for the entire structure.

## 7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

### 7.1.1 Bookkeeping

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 | 959 |
| Number of shifts mapped to atoms | 852 |
| Number of unparsed shifts | 0 |
| Number of shifts with mapping errors | 107 |
| Number of shifts with mapping warnings | 0 |
| Number of shift outliers (ShiftChecker) | 2 |

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 107) occurrences are reported below.

| List ID | Chain | Res | Type | Atom | Shift Data | | |
|---|---|---|---|---|---|---|---|
| | | | | | Value | Uncertainty | Ambiguity |
| 1 | A | 2 | SER | HA | 4.5403 | . | 1 |
| 1 | A | 2 | SER | HB2 | 3.8951 | . | 2 |
| 1 | A | 2 | SER | HB3 | 3.8951 | . | 2 |
| 1 | A | 2 | SER | C | 174.2497 | . | 1 |
| 1 | A | 2 | SER | CA | 58.3359 | . | 1 |
| 1 | A | 2 | SER | CB | 63.9052 | . | 1 |
| 1 | UNMAPPED | 4 | DA | H2 | 7.826 | . | 1 |
| 1 | UNMAPPED | 4 | DA | H2' | 2.887 | . | 2 |
| 1 | UNMAPPED | 4 | DA | H2" | 2.555 | . | 2 |
| 1 | UNMAPPED | 4 | DA | H8 | 7.883 | . | 1 |
| 1 | UNMAPPED | 4 | DA | H62 | 6.472 | . | 2 |
| 1 | UNMAPPED | 5 | DT | H1' | 5.913 | . | 1 |
| 1 | UNMAPPED | 5 | DT | H2' | 2.054 | . | 2 |
| 1 | UNMAPPED | 5 | DT | H2" | 2.536 | . | 2 |

*Continued on next page...*

*Continued from previous page...*

| List ID | Chain | Res | Type | Atom | Shift Data | | |
|---|---|---|---|---|---|---|---|
| | | | | | Value | Uncertainty | Ambiguity |
| 1 | UNMAPPED | 5 | DT | H3 | 13.984 | . | 1 |
| 1 | UNMAPPED | 5 | DT | H5' | 4.196 | . | 2 |
| 1 | UNMAPPED | 5 | DT | H5" | 4.196 | . | 2 |
| 1 | UNMAPPED | 5 | DT | H6 | 7.264 | . | 1 |
| 1 | UNMAPPED | 5 | DT | H71 | 1.206 | . | 1 |
| 1 | UNMAPPED | 5 | DT | H72 | 1.206 | . | 1 |
| 1 | UNMAPPED | 5 | DT | H73 | 1.206 | . | 1 |
| 1 | UNMAPPED | 6 | 5CM | H5A1 | 1.585 | . | 1 |
| 1 | UNMAPPED | 6 | 5CM | H5A2 | 1.585 | . | 1 |
| 1 | UNMAPPED | 6 | 5CM | H5A3 | 1.585 | . | 1 |
| 1 | UNMAPPED | 6 | 5CM | H6 | 7.208 | . | 1 |
| 1 | UNMAPPED | 6 | 5CM | H2' | 2.09 | . | 2 |
| 1 | UNMAPPED | 6 | 5CM | H2" | 2.602 | . | 2 |
| 1 | UNMAPPED | 6 | 5CM | HN41 | 6.759 | . | 2 |
| 1 | UNMAPPED | 6 | 5CM | HN42 | 9.212 | . | 2 |
| 1 | UNMAPPED | 7 | DG | H1 | 13.024 | . | 1 |
| 1 | UNMAPPED | 7 | DG | H8 | 7.82 | . | 1 |
| 1 | UNMAPPED | 8 | DG | H1 | 12.694 | . | 1 |
| 1 | UNMAPPED | 8 | DG | H2' | 2.612 | . | 2 |
| 1 | UNMAPPED | 8 | DG | H2" | 2.531 | . | 2 |
| 1 | UNMAPPED | 9 | DC | H2' | 2.528 | . | 2 |
| 1 | UNMAPPED | 9 | DC | H2" | 2.226 | . | 2 |
| 1 | UNMAPPED | 9 | DC | H5 | 5.425 | . | 1 |
| 1 | UNMAPPED | 9 | DC | H6 | 7.562 | . | 1 |
| 1 | UNMAPPED | 9 | DC | H41 | 7.888 | . | 2 |
| 1 | UNMAPPED | 9 | DC | H42 | 6.456 | . | 2 |
| 1 | UNMAPPED | 10 | THY | H6 | 7.542 | . | 1 |
| 1 | UNMAPPED | 10 | THY | H71 | 1.755 | . | 1 |
| 1 | UNMAPPED | 10 | THY | H72 | 1.755 | . | 1 |
| 1 | UNMAPPED | 10 | THY | H73 | 1.755 | . | 1 |
| 1 | UNMAPPED | 10 | THY | H3 | 13.883 | . | 1 |
| 1 | UNMAPPED | 10 | THY | H1' | 6.152 | . | 1 |
| 1 | UNMAPPED | 10 | THY | H2' | 2.226 | . | 2 |
| 1 | UNMAPPED | 10 | THY | H2" | 2.528 | . | 2 |
| 1 | UNMAPPED | 10 | THY | H3' | 4.888 | . | 1 |
| 1 | UNMAPPED | 11 | DC | H2' | 2.287 | . | 2 |
| 1 | UNMAPPED | 11 | DC | H2" | 2.623 | . | 2 |
| 1 | UNMAPPED | 11 | DC | H5 | 5.889 | . | 1 |
| 1 | UNMAPPED | 11 | DC | H5' | 4.181 | . | 2 |
| 1 | UNMAPPED | 11 | DC | H5" | 4.054 | . | 2 |
| 1 | UNMAPPED | 11 | DC | H6 | 7.692 | . | 1 |

*Continued on next page...*

*Continued from previous page...*

| List ID | Chain | Res | Type | Atom | Shift Data | | |
|---|---|---|---|---|---|---|---|
| | | | | | Value | Uncertainty | Ambiguity |
| 1 | UNMAPPED | 1 | DG | H1' | 5.966 | . | 1 |
| 1 | UNMAPPED | 2 | DA | H1' | 5.565 | . | 1 |
| 1 | UNMAPPED | 2 | DA | H2 | 7.875 | . | 1 |
| 1 | UNMAPPED | 2 | DA | H2' | 2.79 | . | 2 |
| 1 | UNMAPPED | 2 | DA | H2" | 2.459 | . | 2 |
| 1 | UNMAPPED | 2 | DA | H8 | 8.228 | . | 1 |
| 1 | UNMAPPED | 3 | DG | H1 | 12.901 | . | 1 |
| 1 | UNMAPPED | 3 | DG | H1' | 5.938 | . | 1 |
| 1 | UNMAPPED | 3 | DG | H2' | 2.457 | . | 1 |
| 1 | UNMAPPED | 3 | DG | H5' | 3.66 | . | 1 |
| 1 | UNMAPPED | 3 | DG | H8 | 7.844 | . | 1 |
| 1 | UNMAPPED | 4 | DC | H1' | 5.428 | . | 1 |
| 1 | UNMAPPED | 4 | DC | H2' | 2.772 | . | 2 |
| 1 | UNMAPPED | 4 | DC | H2" | 2.478 | . | 2 |
| 1 | UNMAPPED | 4 | DC | H3' | 4.792 | . | 1 |
| 1 | UNMAPPED | 4 | DC | H5 | 5.412 | . | 1 |
| 1 | UNMAPPED | 4 | DC | H5' | 4.094 | . | 2 |
| 1 | UNMAPPED | 4 | DC | H5" | 4.094 | . | 2 |
| 1 | UNMAPPED | 4 | DC | H6 | 7.367 | . | 1 |
| 1 | UNMAPPED | 4 | DC | H41 | 6.957 | . | 2 |
| 1 | UNMAPPED | 4 | DC | H42 | 8.171 | . | 2 |
| 1 | UNMAPPED | 5 | 5CM | H5A1 | 1.637 | . | 1 |
| 1 | UNMAPPED | 5 | 5CM | H5A2 | 1.637 | . | 1 |
| 1 | UNMAPPED | 5 | 5CM | H5A3 | 1.637 | . | 1 |
| 1 | UNMAPPED | 5 | 5CM | H6 | 7.382 | . | 1 |
| 1 | UNMAPPED | 5 | 5CM | H1' | 5.586 | . | 1 |
| 1 | UNMAPPED | 5 | 5CM | H2' | 2.342 | . | 2 |
| 1 | UNMAPPED | 5 | 5CM | H2" | 2.611 | . | 2 |
| 1 | UNMAPPED | 5 | 5CM | H3' | 4.86 | . | 1 |
| 1 | UNMAPPED | 5 | 5CM | H5' | 4.094 | . | 2 |
| 1 | UNMAPPED | 5 | 5CM | H5" | 4.094 | . | 2 |
| 1 | UNMAPPED | 5 | 5CM | H41 | 6.814 | . | 2 |
| 1 | UNMAPPED | 5 | 5CM | H42 | 8.686 | . | 2 |
| 1 | UNMAPPED | 6 | DG | H1 | 12.681 | . | 1 |
| 1 | UNMAPPED | 6 | DG | H1' | 5.83 | . | 1 |
| 1 | UNMAPPED | 6 | DG | H2' | 2.464 | . | 2 |
| 1 | UNMAPPED | 6 | DG | H2" | 2.668 | . | 2 |
| 1 | UNMAPPED | 6 | DG | H8 | 7.855 | . | 1 |
| 1 | UNMAPPED | 7 | DA | H1' | 6.229 | . | 1 |
| 1 | UNMAPPED | 7 | DA | H2 | 7.897 | . | 1 |
| 1 | UNMAPPED | 7 | DA | H2' | 2.625 | . | 2 |

*Continued on next page...*

*Continued from previous page...*

| List ID | Chain | Res | Type | Atom | Shift Data | | |
|---|---|---|---|---|---|---|---|
| | | | | | Value | Uncertainty | Ambiguity |
| 1 | UNMAPPED | 7 | DA | H2" | 2.928 | . | 2 |
| 1 | UNMAPPED | 7 | DA | H3' | 5.056 | . | 1 |
| 1 | UNMAPPED | 7 | DA | H8 | 8.255 | . | 1 |
| 1 | UNMAPPED | 7 | DA | H62 | 7.902 | . | 2 |
| 1 | UNMAPPED | 9 | THY | H6 | 7.248 | . | 1 |
| 1 | UNMAPPED | 9 | THY | H71 | 1.248 | . | 1 |
| 1 | UNMAPPED | 9 | THY | H72 | 1.248 | . | 1 |
| 1 | UNMAPPED | 9 | THY | H73 | 1.248 | . | 1 |
| 1 | UNMAPPED | 9 | THY | H3 | 13.445 | . | 1 |
| 1 | UNMAPPED | 9 | THY | H2' | 2.115 | . | 2 |
| 1 | UNMAPPED | 9 | THY | H2" | 2.468 | . | 2 |

### 7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

| Nucleus | # values | Correction ± precision, *ppm* | Suggested action |
|---|---|---|---|
| $^{13}C_{\alpha}$ | 71 | -0.28 ± 0.28 | None needed (< 0.5 ppm) |
| $^{13}C_{\beta}$ | 62 | -0.42 ± 0.22 | None needed (< 0.5 ppm) |
| $^{13}C'$ | 61 | -0.12 ± 0.34 | None needed (< 0.5 ppm) |
| $^{15}N$ | 61 | -0.40 ± 0.37 | None needed (< 0.5 ppm) |

### 7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 64%, i.e. 799 atoms were assigned a chemical shift out of a possible 1252. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

| | Total | $^{1}H$ | $^{13}C$ | $^{15}N$ |
|---|---|---|---|---|
| Backbone | 309/327 (94%) | 129/134 (96%) | 123/132 (93%) | 57/61 (93%) |
| Sidechain | 424/483 (88%) | 285/311 (92%) | 132/147 (90%) | 7/25 (28%) |
| Aromatic | 66/79 (84%) | 33/38 (87%) | 32/40 (80%) | 1/1 (100%) |
| Sugar | 0/216 (0%) | 0/126 (0%) | 0/90 (0%) | 0/0 (—%) |
| Base | 0/147 (0%) | 0/93 (0%) | 0/29 (0%) | 0/25 (0%) |
| Overall | 799/1252 (64%) | 447/702 (64%) | 287/438 (66%) | 65/112 (58%) |

### 7.1.4 Statistically unusual chemical shifts (i)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

| List Id | Chain | Res | Type | Atom | Shift, *ppm* | Expected range, *ppm* | Z-score |
|---|---|---|---|---|---|---|---|
| 1 | A | 39 | SER | HA | 2.00 | 2.50 – 6.44 | -6.2 |
| 1 | A | 40 | PRO | HD2 | 1.63 | 1.93 – 5.38 | -5.8 |

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

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: