# WORLDWIDE wwPDB PROTEIN DATA BANK

# Full wwPDB NMR Structure Validation Report (i)

Dec 24, 2024 – 10:55 PM EST

| | | |
|---|---|---|
| PDB ID | : | 2MOE |
| BMRB ID | : | 19939 |
| Title | : | Solution structure of MBD4 methyl-cytosine binding domain bound to methylated DNA |
| Authors | : | Williams, D.C.; Walavalkar, N.M. |
| Deposited on | : | 2014-04-24 |

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

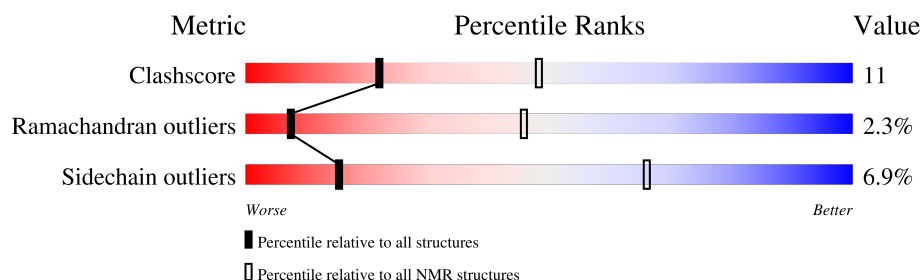| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 2022.3.0, CSD as543be (2022) |
| Percentile statistics | : | 20231227.v01 (using entries in the PDB archive December 27th 2023) |
| wwPDB-RCI | : | v_1n_11_5_13_A (Berjanski et al., 2005) |
| PANAV | : | Wang et al. (2010) |
| wwPDB-ShiftChecker | : | v1.2 |
| BMRB Restraints Analysis | : | v1.2 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.40 |

# 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 is 63%.

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 | | 11 |
| Ramachandran outliers | | 2.3% |
| Sidechain outliers | | 6.9% |

*Worse* *Better*

■ Percentile relative to all structures

▯ Percentile relative to all NMR structures

| Metric | Whole archive (#Entries) | NMR archive (#Entries) |
|--------|--------------------------|------------------------|
| Clashscore | 210492 | 14027 |
| Ramachandran outliers | 207382 | 12486 |
| Sidechain outliers | 206894 | 12463 |

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 | 71 | 59% 21% 20% |
| 2 | B | 10 | 60% 40% |
| 3 | C | 10 | 80% 20% |

WORLDWIDE
wwPDB
PROTEIN DATA BANK

# 2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 4 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:87-A:143 (57) | 0.87 | 4 |

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 4 clusters. No single-model clusters were found.

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

# 3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1779 atoms, of which 805 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Methyl-CpG-binding domain protein 4.

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|------|-------|----------|-------|-----|-----|-----|-----|---|-------|
| | | | Total | C | H | N | O | S | |
| 1 | A | 71 | 1142 | 363 | 574 | 101 | 102 | 2 | 0 |

There are 2 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|-------|---------|----------|--------|----------------|------------|
| A | 78 | GLY | - | expression tag | UNP O95243 |
| A | 79 | SER | - | expression tag | UNP O95243 |

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

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|------|-------|----------|-------|----|-----|----|----|---|-------|
| | | | Total | C | H | N | O | P | |
| 2 | B | 10 | 320 | 98 | 116 | 38 | 59 | 9 | 0 |

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

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|------|-------|----------|-------|----|-----|----|----|---|-------|
| | | | Total | C | H | N | O | P | |
| 3 | C | 10 | 317 | 97 | 115 | 39 | 57 | 9 | 0 |

# 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.
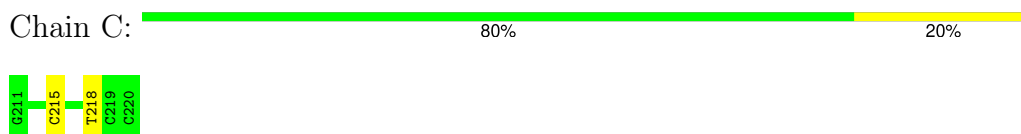
- Molecule 1: Methyl-CpG-binding domain protein 4
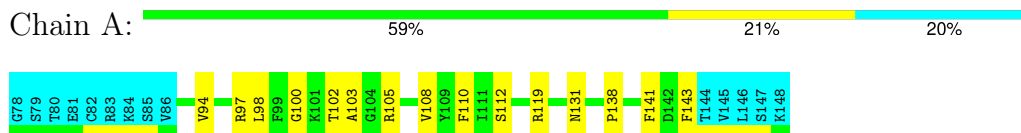
Chain A:



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

Chain B:



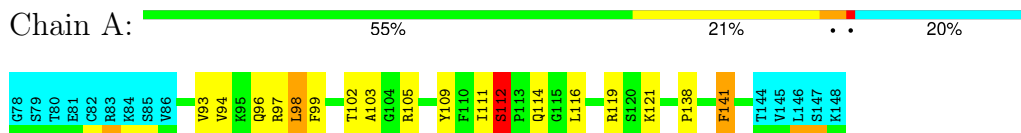- Molecule 3: DNA (5'-D(*GP*AP*GP*CP*(5CM)P*GP*AP*TP*CP*C)-3')

Chain C:



## 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: Methyl-CpG-binding domain protein 4

Chain A:

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

Chain B:



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

Chain C:



### 4.2.2 Score per residue for model 2

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A:



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

Chain B:



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

Chain C:



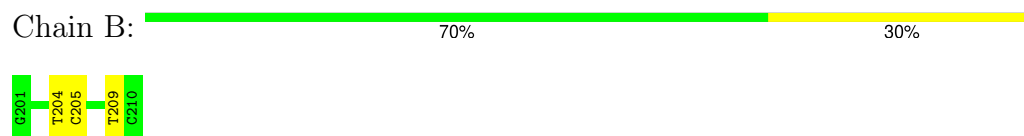### 4.2.3 Score per residue for model 3
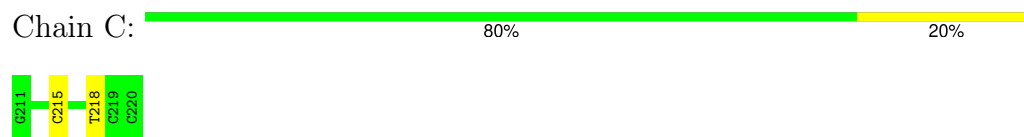
• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A:



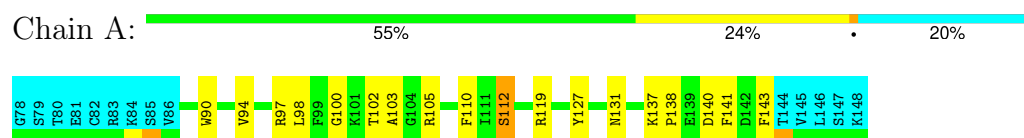• Molecule 2: DNA (5'-D(*GP*GP*AP*TP*(5CM)P*GP*GP*CP*TP*C)-3')

Chain B:    70%      30%

G201 T204 C205 T209 C210

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

Chain C:    80%      20%

G211 C215 T218 C219 C220

### 4.2.4   Score per residue for model 4 (medoid)

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A:    55%     24%   •   20%

G78 S79 T80 E81 C82 R83 K84 S85 V86 W90 V94 R97 L98 F99 G100 K101 T102 A103 G104 R105 F110 I111 S112 R119 Y127 N131 K137 P138 E139 D140 F141 D142 F143 T144 V145 L146 S147 K148

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

Chain B:    60%      40%

G201 T204 C205 G206 T209 C210

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

Chain C:    80%      20%

G211 C215 T218 C219 C220

### 4.2.5   Score per residue for model 5

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A:    54%     25%   •   20%

G78 S79 T80 E81 C82 R83 K84 S85 V86 R92 V93 V94 R97 L98 T102 R105 F110 I111 S112 P113 Q114 G115 L116 R119 L124 Y127 N131 P138 F141 D142 F143 T144 V145 L146 S147 K148

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

Chain B:    70%      30%

G201 T204 C205 T209 C210

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

Chain C: 70% 30%

G211 A212 G213 C214 C215 T218 C219 C220

### 4.2.6 Score per residue for model 6

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A: 55% 24% • 20%

G78 S79 T80 E81 C82 R83 K84 S85 V86 R92 V93 V94 K95 Q96 R97 L98 T102 A103 F106 D107 V108 Y109 F110 I111 L116 R119 E133 T134 S135 P138 F141 T144 V145 L146 S147 K148

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

Chain B: 70% 30%

G201 T204 C205 T209 C210

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

Chain C: 80% 20%

G211 C215 T218 C219 C220

### 4.2.7 Score per residue for model 7

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A: 58% 17% 6% 20%

G78 S79 T80 E81 C82 R83 K84 S85 V86 E91 R97 L98 F99 G100 K101 T102 A103 G104 R105 V108 Y109 F110 I111 S112 N131 K137 P138 E139 D140 F141 T144 V145 L146 S147 K148

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

Chain B: 60% 40%

G201 T204 C205 G206 T209 C210

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

Chain C:



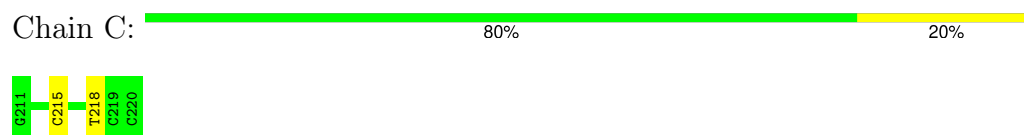## 4.2.8 Score per residue for model 8

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A:



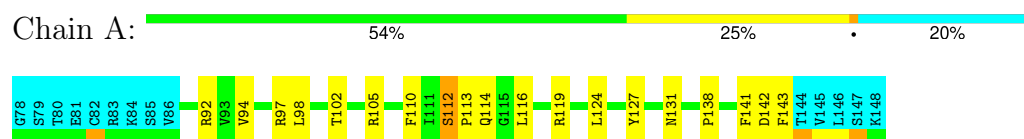• Molecule 2: DNA (5'-D(*GP*GP*AP*TP*(5CM)P*GP*GP*CP*TP*C)-3')

Chain B:



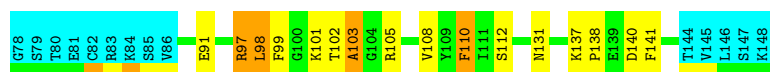• Molecule 3: DNA (5'-D(*GP*AP*GP*CP*(5CM)P*GP*AP*TP*CP*C)-3')

Chain C:



## 4.2.9 Score per residue for model 9

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A:
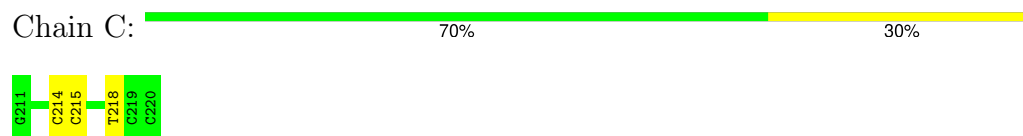


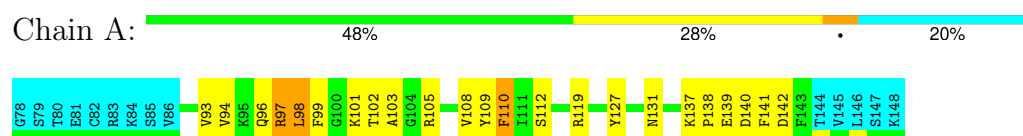• Molecule 2: DNA (5'-D(*GP*GP*AP*TP*(5CM)P*GP*GP*CP*TP*C)-3')

Chain B:



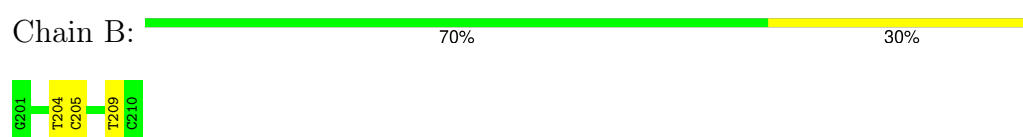• Molecule 3: DNA (5'-D(*GP*AP*GP*CP*(5CM)P*GP*AP*TP*CP*C)-3')

Chain C: 80% 20%

G211 C215 T218 C219 C220

### 4.2.10 Score per residue for model 10

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A: 58% 17% 6% 20%

G78 S79 T80 E81 C82 R83 K84 S85 V86 V94 L98 F99 T102 A103 G104 R105 V108 Y109 F110 I111 S112 L116 K117 F118 Y127 P138 F141 D142 F143 T144 V145 L146 S147 K148

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

Chain B: 70% 30%

G201 T204 C205 T209 C210

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

Chain C: 70% 30%

G211 C214 C215 T218 C219 C220

### 4.2.11 Score per residue for model 11

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A: 59% 20% • 20%

G78 S79 T80 E81 C82 R83 K84 S85 V86 W90 Q96 R97 L98 F106 F110 I111 S112 K117 F118 R119 S120 K121 L136 K137 P138 F141 T144 V145 L146 S147 K148

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

Chain B: 70% 30%

G201 T204 C205 T209 C210

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

Chain C: 80% 20%

### 4.2.12 Score per residue for model 12

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A:





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

Chain B:





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

Chain C:
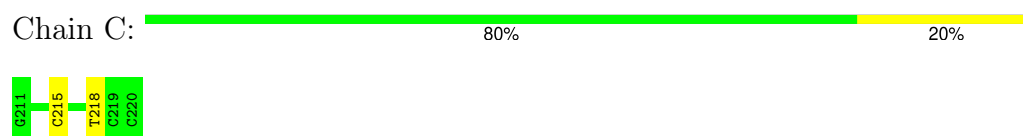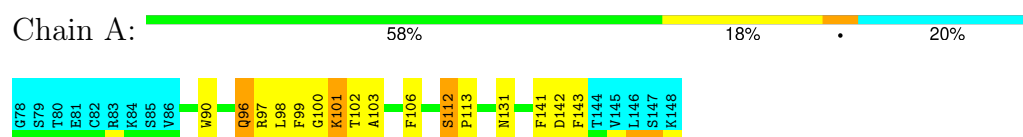

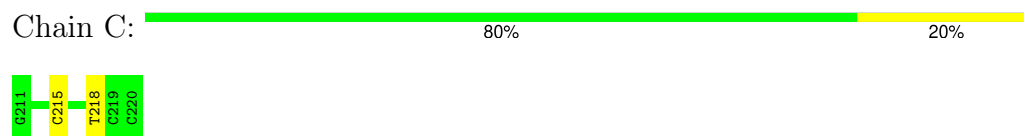


### 4.2.13 Score per residue for model 13

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A:





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

Chain B:





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

Chain C:

### 4.2.14 Score per residue for model 14

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A:



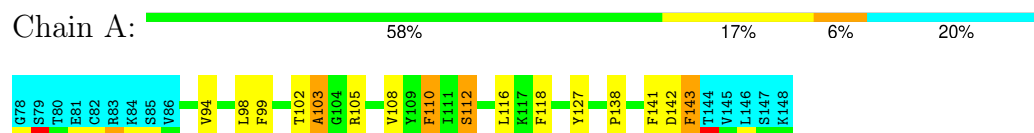• Molecule 2: DNA (5'-D(*GP*GP*AP*TP*(5CM)P*GP*GP*CP*TP*C)-3')

Chain B:



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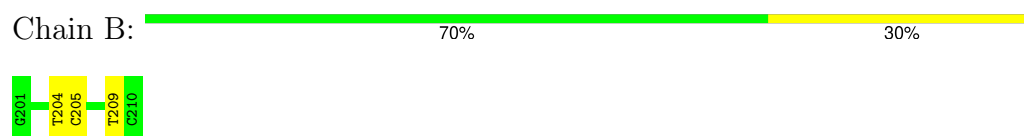
Chain C:
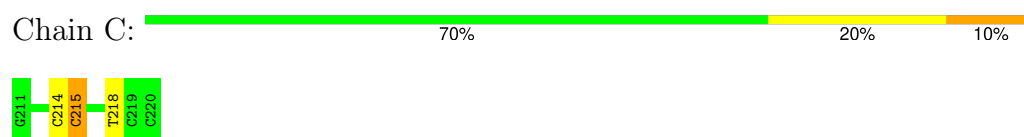


### 4.2.15 Score per residue for model 15

• Molecule 1: Methyl-CpG-binding domain protein 4

Chain A:



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

Chain B:



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

Chain C:
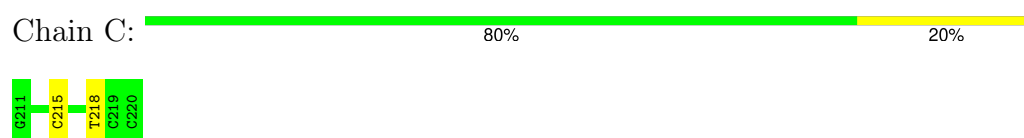
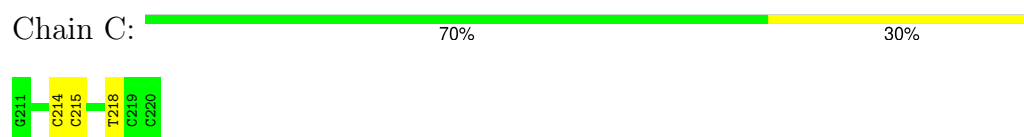### 4.2.16 Score per residue for model 16

• Molecule 1: Methyl-CpG-binding domain protein 4
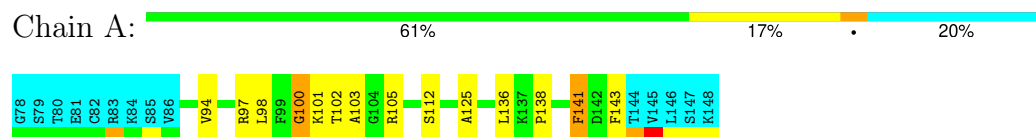
Chain A:



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

Chain B:



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

Chain C:
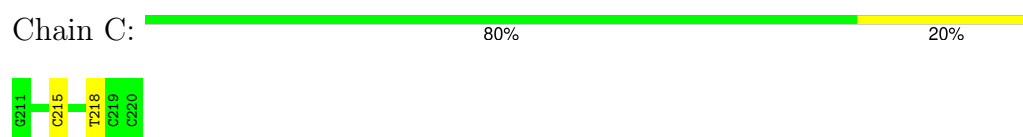


### 4.2.17 Score per residue for model 17

• Molecule 1: Methyl-CpG-binding domain protein 4
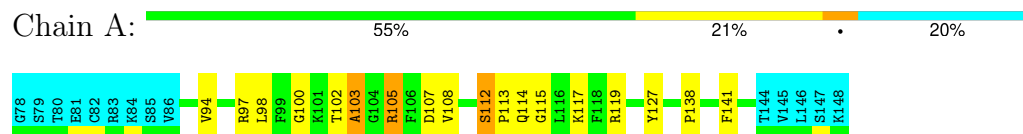
Chain A:



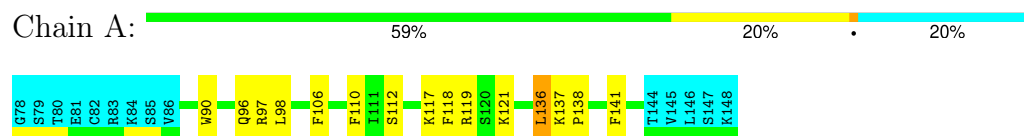• Molecule 2: DNA (5'-D(*GP*GP*AP*TP*(5CM)P*GP*GP*CP*TP*C)-3')

Chain B:



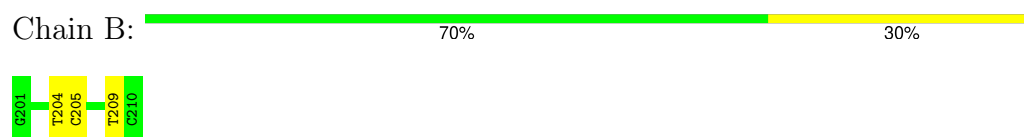• Molecule 3: DNA (5'-D(*GP*AP*GP*CP*(5CM)P*GP*AP*TP*CP*C)-3')

Chain C:
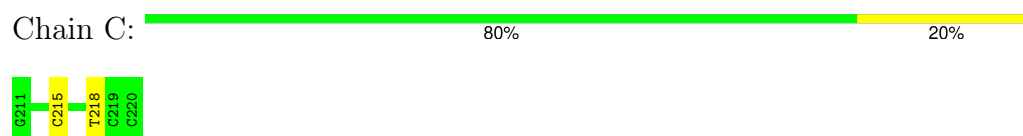
### 4.2.18 Score per residue for model 18

• Molecule 1: Methyl-CpG-binding domain protein 4
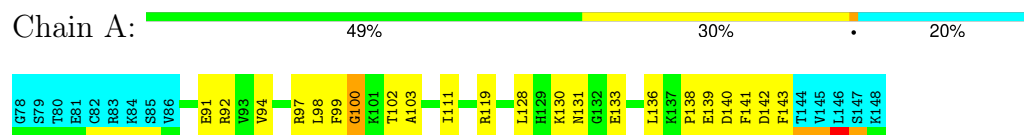
Chain A:



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

Chain B:



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

Chain C:
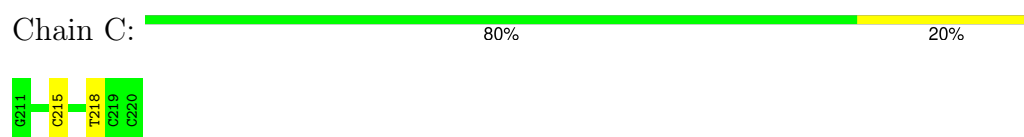


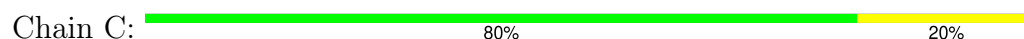### 4.2.19 Score per residue for model 19

• Molecule 1: Methyl-CpG-binding domain protein 4
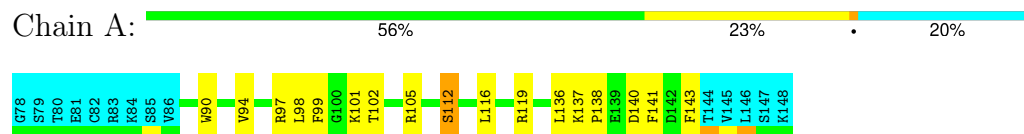
Chain A:



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

Chain B:



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

Chain C:
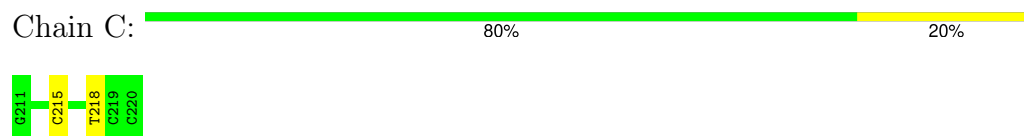
#### 4.2.20　Score per residue for model 20
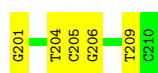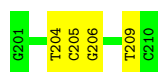
• Molecule 1: Methyl-CpG-binding domain protein 4
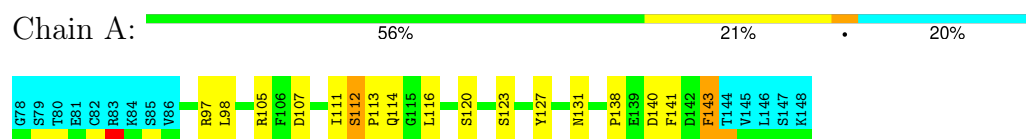
Chain A:
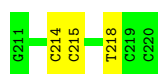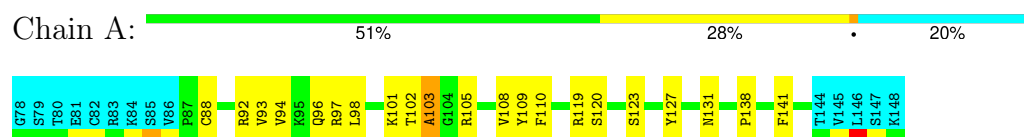


• Molecule 2: DNA (5'-D(*GP*GP*AP*TP*(5CM)P*GP*GP*CP*TP*C)-3')

Chain B:



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

Chain C:
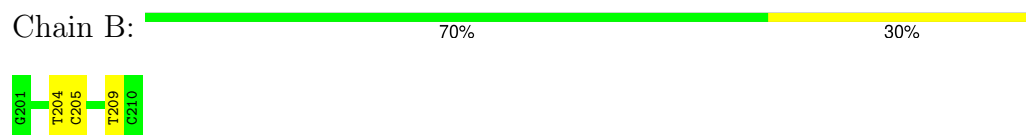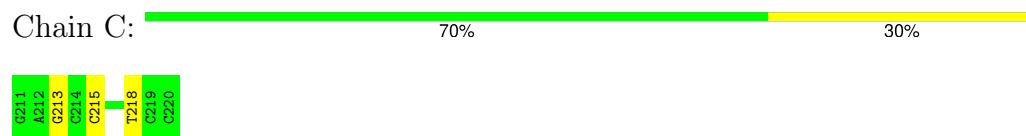
# 5   Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *simulated annealing.*

Of the 50 calculated structures, 20 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 | 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 | 3 |
| Total number of shifts | 892 |
| Number of shifts mapped to atoms | 892 |
| Number of unparsed shifts | 0 |
| Number of shifts with mapping errors | 0 |
| Number of shifts with mapping warnings | 0 |
| Assignment completeness (well-defined parts) | 63% |

# 6 Model quality (i)

## 6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: 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 | 0.66±0.00 | 0±0/481 ( 0.0± 0.0%) | 0.77±0.00 | 0±0/645 ( 0.0± 0.0%) |
| 2 | B | 0.96±0.01 | 1±1/205 ( 0.7± 0.3%) | 1.38±0.02 | 4±0/313 ( 1.3± 0.1%) |
| 3 | C | 0.90±0.01 | 1±0/203 ( 0.3± 0.2%) | 1.30±0.01 | 2±0/309 ( 0.6± 0.0%) |
| All | All | 0.79 | 43/17780 ( 0.2%) | 1.09 | 119/25340 ( 0.5%) |

All 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 | 209 | DT | C5-C7 | 5.42 | 1.53 | 1.50 | 5 | 15 |
| 2 | B | 204 | DT | C5-C7 | 5.37 | 1.53 | 1.50 | 16 | 14 |
| 3 | C | 218 | DT | C5-C7 | 5.30 | 1.53 | 1.50 | 16 | 14 |

All 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 | 218 | DT | C6-C5-C7 | -6.33 | 119.10 | 122.90 | 14 | 20 |
| 2 | B | 204 | DT | C6-C5-C7 | -6.16 | 119.20 | 122.90 | 13 | 20 |
| 2 | B | 209 | DT | C6-C5-C7 | -6.05 | 119.27 | 122.90 | 18 | 20 |
| 3 | C | 218 | DT | C4-C5-C6 | 5.59 | 121.35 | 118.00 | 19 | 20 |
| 2 | B | 204 | DT | C4-C5-C6 | 5.49 | 121.29 | 118.00 | 20 | 20 |
| 2 | B | 209 | DT | C4-C5-C6 | 5.37 | 121.22 | 118.00 | 17 | 19 |

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 | 467 | 461 | 460 | 18±4 |
| 2 | B | 204 | 116 | 116 | 1±1 |
| 3 | C | 202 | 115 | 115 | 1±1 |
| All | All | 17460 | 13840 | 13820 | 355 |

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 11.

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

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models Worst | Total |
|--------|--------|----------|-------------|--------------|-------|
| 1:A:98:LEU:N | 1:A:98:LEU:HD22 | 0.80 | 1.91 | 19 | 4 |
| 1:A:102:THR:HG21 | 2:B:206:DG:N7 | 0.74 | 1.97 | 7 | 8 |
| 1:A:97:ARG:C | 1:A:98:LEU:HD12 | 0.69 | 2.08 | 13 | 3 |
| 1:A:102:THR:HG21 | 2:B:206:DG:C8 | 0.67 | 2.25 | 9 | 5 |
| 1:A:97:ARG:NH1 | 1:A:102:THR:OG1 | 0.66 | 2.29 | 1 | 3 |
| 1:A:127:TYR:CZ | 1:A:131:ASN:ND2 | 0.64 | 2.65 | 4 | 2 |
| 1:A:98:LEU:O | 1:A:99:PHE:CD1 | 0.64 | 2.50 | 7 | 4 |
| 1:A:97:ARG:C | 1:A:98:LEU:HD22 | 0.64 | 2.12 | 12 | 10 |
| 1:A:143:PHE:O | 1:A:143:PHE:CD2 | 0.64 | 2.51 | 5 | 2 |
| 1:A:128:LEU:O | 1:A:131:ASN:OD1 | 0.63 | 2.17 | 18 | 1 |
| 1:A:90:TRP:CZ2 | 1:A:136:LEU:HD22 | 0.62 | 2.30 | 11 | 1 |
| 1:A:110:PHE:CE1 | 1:A:141:PHE:CD2 | 0.62 | 2.88 | 4 | 4 |
| 1:A:141:PHE:CG | 1:A:141:PHE:O | 0.62 | 2.53 | 11 | 11 |
| 1:A:110:PHE:CE1 | 1:A:141:PHE:CE1 | 0.62 | 2.87 | 8 | 1 |
| 1:A:110:PHE:CE1 | 1:A:141:PHE:CE2 | 0.61 | 2.88 | 4 | 1 |
| 1:A:110:PHE:CZ | 1:A:141:PHE:CE1 | 0.61 | 2.88 | 7 | 3 |
| 1:A:110:PHE:CD1 | 1:A:110:PHE:N | 0.60 | 2.68 | 3 | 9 |
| 1:A:98:LEU:HD22 | 1:A:98:LEU:N | 0.60 | 2.11 | 7 | 12 |
| 1:A:143:PHE:O | 1:A:143:PHE:CG | 0.59 | 2.54 | 5 | 3 |
| 1:A:98:LEU:N | 1:A:98:LEU:CD2 | 0.59 | 2.63 | 19 | 13 |
| 1:A:138:PRO:O | 1:A:141:PHE:CD1 | 0.59 | 2.56 | 2 | 14 |
| 1:A:116:LEU:HD12 | 1:A:116:LEU:C | 0.58 | 2.18 | 10 | 1 |
| 1:A:116:LEU:HD12 | 1:A:116:LEU:O | 0.58 | 1.98 | 10 | 1 |
| 1:A:97:ARG:O | 1:A:103:ALA:HB1 | 0.58 | 1.98 | 18 | 1 |
| 1:A:105:ARG:NH1 | 3:C:214:DC:N4 | 0.58 | 2.50 | 14 | 2 |

*Continued on next page...*

*Continued from previous page...*

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 1:A:116:LEU:HD13 | 1:A:118:PHE:CE2 | 0.58 | 2.33 | 10 | 1 |
| 1:A:140:ASP:OD1 | 1:A:141:PHE:N | 0.58 | 2.37 | 7 | 2 |
| 1:A:127:TYR:CE1 | 1:A:131:ASN:ND2 | 0.57 | 2.73 | 13 | 3 |
| 1:A:96:GLN:NE2 | 1:A:106:PHE:CE2 | 0.57 | 2.72 | 6 | 2 |
| 1:A:105:ARG:HE | 1:A:105:ARG:CA | 0.57 | 2.12 | 17 | 1 |
| 1:A:105:ARG:HH12 | 3:C:214:DC:N4 | 0.56 | 1.98 | 14 | 1 |
| 1:A:105:ARG:CA | 1:A:105:ARG:NE | 0.56 | 2.69 | 17 | 1 |
| 1:A:94:VAL:O | 1:A:94:VAL:HG23 | 0.56 | 2.01 | 1 | 5 |
| 1:A:114:GLN:OE1 | 1:A:116:LEU:HD23 | 0.55 | 2.01 | 1 | 1 |
| 1:A:141:PHE:O | 1:A:141:PHE:CD1 | 0.55 | 2.59 | 16 | 4 |
| 1:A:94:VAL:HG12 | 1:A:108:VAL:HG12 | 0.55 | 1.77 | 2 | 8 |
| 1:A:99:PHE:CD1 | 1:A:99:PHE:N | 0.55 | 2.74 | 19 | 2 |
| 1:A:141:PHE:CD1 | 1:A:141:PHE:O | 0.55 | 2.59 | 1 | 3 |
| 1:A:138:PRO:O | 1:A:141:PHE:CD2 | 0.54 | 2.60 | 3 | 3 |
| 1:A:136:LEU:HD12 | 1:A:140:ASP:OD1 | 0.54 | 2.02 | 19 | 1 |
| 1:A:105:ARG:NH2 | 3:C:213:DG:N7 | 0.54 | 2.55 | 20 | 1 |
| 1:A:131:ASN:OD1 | 1:A:133:GLU:N | 0.54 | 2.38 | 18 | 1 |
| 1:A:102:THR:O | 1:A:103:ALA:C | 0.54 | 2.46 | 12 | 11 |
| 1:A:127:TYR:CE1 | 1:A:131:ASN:OD1 | 0.54 | 2.62 | 14 | 2 |
| 1:A:105:ARG:NH1 | 3:C:213:DG:N7 | 0.53 | 2.56 | 5 | 3 |
| 1:A:114:GLN:NE2 | 1:A:114:GLN:H | 0.53 | 2.02 | 5 | 1 |
| 1:A:105:ARG:CZ | 3:C:213:DG:N7 | 0.53 | 2.71 | 20 | 1 |
| 1:A:94:VAL:HG12 | 1:A:108:VAL:CG1 | 0.53 | 2.33 | 2 | 8 |
| 1:A:105:ARG:NH2 | 3:C:214:DC:C5 | 0.53 | 2.76 | 14 | 1 |
| 1:A:92:ARG:NH2 | 1:A:142:ASP:O | 0.53 | 2.41 | 18 | 2 |
| 1:A:133:GLU:O | 1:A:135:SER:N | 0.53 | 2.41 | 15 | 1 |
| 1:A:141:PHE:O | 1:A:141:PHE:CG | 0.53 | 2.61 | 2 | 6 |
| 1:A:96:GLN:OE1 | 1:A:106:PHE:CE2 | 0.53 | 2.62 | 9 | 1 |
| 1:A:138:PRO:O | 1:A:141:PHE:CE2 | 0.53 | 2.62 | 7 | 3 |
| 1:A:143:PHE:CG | 1:A:143:PHE:O | 0.52 | 2.62 | 12 | 1 |
| 1:A:94:VAL:HG23 | 1:A:94:VAL:O | 0.52 | 2.05 | 18 | 5 |
| 1:A:117:LYS:O | 1:A:118:PHE:CD1 | 0.52 | 2.63 | 11 | 1 |
| 1:A:136:LEU:CD2 | 1:A:136:LEU:N | 0.52 | 2.72 | 19 | 2 |
| 1:A:93:VAL:HG11 | 1:A:109:TYR:OH | 0.52 | 2.05 | 8 | 8 |
| 1:A:114:GLN:NE2 | 1:A:127:TYR:OH | 0.51 | 2.42 | 17 | 2 |
| 1:A:138:PRO:O | 1:A:141:PHE:CE1 | 0.51 | 2.63 | 2 | 4 |
| 1:A:112:SER:OG | 1:A:116:LEU:HD11 | 0.51 | 2.05 | 10 | 1 |
| 1:A:102:THR:O | 1:A:105:ARG:NE | 0.51 | 2.43 | 17 | 1 |
| 1:A:136:LEU:N | 1:A:136:LEU:HD22 | 0.51 | 2.19 | 19 | 4 |
| 1:A:105:ARG:HH12 | 3:C:214:DC:H41 | 0.51 | 1.48 | 7 | 1 |
| 1:A:90:TRP:CZ3 | 1:A:112:SER:OG | 0.51 | 2.63 | 19 | 1 |

*Continued on next page...*

*Continued from previous page...*

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 1:A:96:GLN:NE2 | 1:A:96:GLN:N | 0.50 | 2.58 | 2 | 1 |
| 1:A:98:LEU:CD2 | 1:A:98:LEU:N | 0.50 | 2.74 | 8 | 3 |
| 1:A:125:ALA:HB2 | 1:A:138:PRO:HG3 | 0.50 | 1.83 | 16 | 1 |
| 1:A:133:GLU:C | 1:A:135:SER:H | 0.50 | 2.10 | 6 | 1 |
| 1:A:100:GLY:N | 1:A:103:ALA:HB2 | 0.50 | 2.22 | 12 | 4 |
| 1:A:110:PHE:CZ | 1:A:141:PHE:CD1 | 0.50 | 3.00 | 3 | 2 |
| 1:A:110:PHE:CE2 | 1:A:141:PHE:CD2 | 0.50 | 3.00 | 5 | 1 |
| 1:A:105:ARG:NH1 | 2:B:206:DG:O6 | 0.50 | 2.45 | 14 | 1 |
| 1:A:116:LEU:HD13 | 1:A:118:PHE:CD2 | 0.50 | 2.42 | 10 | 1 |
| 1:A:126:ASN:ND2 | 2:B:201:DG:O5' | 0.50 | 2.45 | 12 | 1 |
| 1:A:101:LYS:HG3 | 1:A:102:THR:HG23 | 0.49 | 1.84 | 9 | 1 |
| 1:A:111:ILE:HG23 | 1:A:116:LEU:O | 0.49 | 2.07 | 20 | 2 |
| 1:A:98:LEU:HD12 | 1:A:98:LEU:N | 0.49 | 2.23 | 10 | 4 |
| 1:A:105:ARG:NH1 | 3:C:214:DC:H41 | 0.49 | 2.05 | 10 | 1 |
| 1:A:110:PHE:CZ | 1:A:141:PHE:CD2 | 0.49 | 3.01 | 2 | 1 |
| 1:A:110:PHE:CE1 | 1:A:141:PHE:CD1 | 0.49 | 3.00 | 8 | 3 |
| 1:A:102:THR:O | 1:A:103:ALA:O | 0.49 | 2.31 | 7 | 5 |
| 1:A:116:LEU:CD2 | 1:A:127:TYR:CE1 | 0.49 | 2.95 | 10 | 1 |
| 1:A:105:ARG:HE | 3:C:214:DC:H41 | 0.49 | 1.50 | 13 | 1 |
| 1:A:102:THR:HG21 | 2:B:206:DG:C5 | 0.49 | 2.42 | 15 | 1 |
| 1:A:141:PHE:O | 1:A:141:PHE:CD2 | 0.48 | 2.66 | 20 | 4 |
| 1:A:105:ARG:HE | 1:A:105:ARG:N | 0.48 | 2.06 | 17 | 1 |
| 1:A:91:GLU:N | 1:A:91:GLU:OE1 | 0.48 | 2.47 | 18 | 4 |
| 1:A:112:SER:CB | 1:A:113:PRO:CD | 0.48 | 2.91 | 17 | 4 |
| 1:A:136:LEU:N | 1:A:136:LEU:CD2 | 0.48 | 2.77 | 18 | 2 |
| 1:A:96:GLN:OE1 | 1:A:106:PHE:CZ | 0.47 | 2.67 | 9 | 1 |
| 1:A:99:PHE:H | 1:A:103:ALA:HB2 | 0.47 | 1.69 | 9 | 1 |
| 1:A:137:LYS:O | 1:A:140:ASP:OD2 | 0.47 | 2.32 | 2 | 2 |
| 1:A:96:GLN:OE1 | 1:A:97:ARG:O | 0.47 | 2.33 | 1 | 2 |
| 1:A:114:GLN:HE22 | 1:A:131:ASN:ND2 | 0.47 | 2.08 | 20 | 1 |
| 1:A:102:THR:CA | 1:A:105:ARG:HH21 | 0.47 | 2.23 | 5 | 1 |
| 1:A:141:PHE:CD2 | 1:A:141:PHE:O | 0.47 | 2.69 | 2 | 3 |
| 1:A:142:ASP:OD1 | 1:A:143:PHE:N | 0.46 | 2.49 | 9 | 1 |
| 1:A:110:PHE:CD1 | 1:A:141:PHE:CD2 | 0.46 | 3.03 | 4 | 1 |
| 1:A:96:GLN:OE1 | 1:A:106:PHE:CD2 | 0.46 | 2.68 | 11 | 2 |
| 1:A:115:GLY:O | 1:A:117:LYS:NZ | 0.46 | 2.49 | 17 | 1 |
| 1:A:137:LYS:N | 1:A:140:ASP:OD2 | 0.45 | 2.49 | 4 | 2 |
| 1:A:116:LEU:N | 1:A:116:LEU:CD2 | 0.45 | 2.78 | 6 | 3 |
| 1:A:98:LEU:O | 1:A:99:PHE:CG | 0.45 | 2.68 | 15 | 1 |
| 1:A:105:ARG:NE | 1:A:105:ARG:CA | 0.45 | 2.78 | 3 | 1 |
| 1:A:96:GLN:O | 1:A:98:LEU:CD1 | 0.45 | 2.65 | 2 | 1 |

*Continued on next page...*

*Continued from previous page...*

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 1:A:139:GLU:N | 1:A:139:GLU:OE1 | 0.45 | 2.50 | 18 | 1 |
| 1:A:141:PHE:CD1 | 1:A:141:PHE:C | 0.45 | 2.90 | 8 | 3 |
| 1:A:139:GLU:OE1 | 1:A:139:GLU:N | 0.45 | 2.48 | 8 | 1 |
| 1:A:112:SER:OG | 1:A:116:LEU:O | 0.45 | 2.32 | 12 | 1 |
| 1:A:99:PHE:O | 1:A:100:GLY:C | 0.45 | 2.56 | 18 | 1 |
| 1:A:131:ASN:OD1 | 1:A:131:ASN:O | 0.45 | 2.35 | 7 | 4 |
| 1:A:116:LEU:N | 1:A:116:LEU:HD22 | 0.44 | 2.27 | 6 | 2 |
| 1:A:142:ASP:O | 1:A:142:ASP:OD1 | 0.44 | 2.35 | 2 | 1 |
| 1:A:105:ARG:NE | 3:C:214:DC:H41 | 0.44 | 2.10 | 13 | 1 |
| 1:A:91:GLU:OE1 | 1:A:111:ILE:O | 0.44 | 2.34 | 18 | 1 |
| 1:A:114:GLN:CD | 1:A:114:GLN:N | 0.43 | 2.71 | 5 | 2 |
| 1:A:116:LEU:HD12 | 1:A:127:TYR:CE1 | 0.43 | 2.47 | 5 | 1 |
| 1:A:96:GLN:NE2 | 1:A:106:PHE:CZ | 0.43 | 2.85 | 9 | 1 |
| 1:A:90:TRP:CZ3 | 1:A:112:SER:CA | 0.43 | 3.01 | 4 | 1 |
| 1:A:97:ARG:NH1 | 1:A:102:THR:CB | 0.43 | 2.82 | 3 | 1 |
| 1:A:111:ILE:O | 1:A:112:SER:O | 0.43 | 2.37 | 1 | 1 |
| 1:A:105:ARG:HE | 3:C:214:DC:N4 | 0.43 | 2.11 | 13 | 1 |
| 1:A:97:ARG:NE | 1:A:107:ASP:OD2 | 0.43 | 2.51 | 17 | 1 |
| 1:A:133:GLU:C | 1:A:135:SER:N | 0.43 | 2.71 | 6 | 1 |
| 1:A:133:GLU:O | 1:A:134:THR:OG1 | 0.43 | 2.33 | 6 | 1 |
| 1:A:105:ARG:NH2 | 3:C:215:5CM:N4 | 0.43 | 2.67 | 14 | 1 |
| 1:A:140:ASP:OD1 | 1:A:140:ASP:O | 0.42 | 2.37 | 18 | 1 |
| 1:A:114:GLN:OE1 | 1:A:114:GLN:N | 0.42 | 2.48 | 2 | 1 |
| 1:A:96:GLN:NE2 | 1:A:97:ARG:O | 0.42 | 2.53 | 8 | 1 |
| 1:A:124:LEU:CD2 | 1:A:141:PHE:CD1 | 0.42 | 3.03 | 5 | 1 |
| 1:A:102:THR:O | 1:A:105:ARG:CZ | 0.42 | 2.68 | 17 | 1 |
| 1:A:124:LEU:HD23 | 1:A:141:PHE:CE1 | 0.42 | 2.50 | 5 | 1 |
| 1:A:96:GLN:O | 1:A:98:LEU:CD2 | 0.42 | 2.68 | 14 | 1 |
| 1:A:98:LEU:N | 1:A:98:LEU:CD1 | 0.41 | 2.83 | 2 | 1 |
| 1:A:120:SER:OG | 1:A:123:SER:OG | 0.41 | 2.37 | 14 | 1 |
| 1:A:99:PHE:CE1 | 2:B:206:DG:OP2 | 0.41 | 2.73 | 15 | 1 |
| 1:A:96:GLN:CD | 1:A:106:PHE:CE2 | 0.41 | 2.94 | 9 | 1 |
| 1:A:114:GLN:OE1 | 1:A:116:LEU:CD2 | 0.41 | 2.68 | 1 | 1 |
| 1:A:90:TRP:CH2 | 1:A:112:SER:OG | 0.41 | 2.62 | 2 | 1 |
| 1:A:142:ASP:O | 1:A:143:PHE:O | 0.41 | 2.39 | 10 | 2 |
| 1:A:137:LYS:O | 1:A:140:ASP:CG | 0.41 | 2.59 | 7 | 1 |
| 1:A:120:SER:O | 1:A:123:SER:OG | 0.41 | 2.37 | 20 | 1 |
| 1:A:105:ARG:CZ | 1:A:105:ARG:N | 0.41 | 2.84 | 3 | 1 |
| 1:A:90:TRP:CD1 | 1:A:90:TRP:N | 0.41 | 2.89 | 9 | 1 |
| 1:A:97:ARG:O | 1:A:98:LEU:HD22 | 0.41 | 2.16 | 12 | 1 |
| 1:A:136:LEU:CD1 | 1:A:140:ASP:OD2 | 0.40 | 2.69 | 18 | 1 |

*Continued on next page...*

*Continued from previous page...*

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models Worst | Models Total |
|---|---|---|---|---|---|
| 1:A:131:ASN:OD1 | 1:A:131:ASN:C | 0.40 | 2.57 | 18 | 1 |

## 6.3 Torsion angles (i)

### 6.3.1 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 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 | 57/71 (80%) | 54±1 (95±2%) | 2±1 (3±2%) | 1±1 (2±1%) | 7 | 46 |
| All | All | 1140/1420 (80%) | 1080 (95%) | 34 (3%) | 26 (2%) | 7 | 46 |

All 7 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 | 103 | ALA | 8 |
| 1 | A | 100 | GLY | 7 |
| 1 | A | 98 | LEU | 4 |
| 1 | A | 143 | PHE | 4 |
| 1 | A | 112 | SER | 1 |
| 1 | A | 99 | PHE | 1 |
| 1 | A | 134 | THR | 1 |

### 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 | 50/63 (79%) | 47±1 (93±3%) | 3±1 (7±3%) | 15 | 66 |
| All | All | 1000/1260 (79%) | 931 (93%) | 69 (7%) | 15 | 66 |

All 18 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 | 112 | SER | 15 |
| 1 | A | 119 | ARG | 12 |
| 1 | A | 105 | ARG | 9 |
| 1 | A | 101 | LYS | 7 |
| 1 | A | 110 | PHE | 5 |
| 1 | A | 121 | LYS | 3 |
| 1 | A | 97 | ARG | 3 |
| 1 | A | 141 | PHE | 2 |
| 1 | A | 96 | GLN | 2 |
| 1 | A | 108 | VAL | 2 |
| 1 | A | 137 | LYS | 2 |
| 1 | A | 95 | LYS | 1 |
| 1 | A | 131 | ASN | 1 |
| 1 | A | 136 | LEU | 1 |
| 1 | A | 92 | ARG | 1 |
| 1 | A | 117 | LYS | 1 |
| 1 | A | 130 | LYS | 1 |
| 1 | A | 107 | ASP | 1 |

### 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 |
| 2 | 5CM | B | 205 | 2,3 | 18,21,22 | 0.42±0.02 | 0±0 (0±0%) |
| 3 | 5CM | C | 215 | 2,3 | 18,21,22 | 0.41±0.02 | 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 |
|---|---|---|---|---|---|---|---|
| 2 | 5CM | B | 205 | 2,3 | 24,30,33 | 0.76±0.01 | 1±0 (4±0%) |
| 3 | 5CM | C | 215 | 2,3 | 24,30,33 | 0.76±0.01 | 1±0 (4±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 | 5CM | B | 205 | 2,3 | - | 0±0,7,21,22 | 0±0,2,2,2 |
| 3 | 5CM | C | 215 | 2,3 | - | 0±0,7,21,22 | 0±0,2,2,2 |

There are no bond-length outliers.

All 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 | 205 | 5CM | C5-C6-N1 | 2.57 | 120.53 | 123.31 | 11 | 20 |
| 3 | C | 215 | 5CM | C5-C6-N1 | 2.55 | 120.54 | 123.31 | 20 | 20 |

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 6.5 Carbohydrates (i)

There are no oligosaccharides in this entry.

## 6.6 Ligand geometry (i)

There are no ligands in this entry.

## 6.7 Other polymers ⓘ

There are no such molecules in this entry.

## 6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

# 7   Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 63% 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: *MBD4mbd_ chemshift*

### 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 | 798 |
| Number of shifts mapped to atoms | 798 |
| Number of unparsed shifts | 0 |
| Number of shifts with mapping errors | 0 |
| Number of shifts with mapping warnings | 0 |
| Number of shift outliers (ShiftChecker) | 1 |

### 7.1.2   Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

| Nucleus | # values | Correction ± precision, *ppm* | Suggested action |
|---|---|---|---|
| $^{13}C_\alpha$ | 70 | -0.44 ± 0.08 | None needed (< 0.5 ppm) |
| $^{13}C_\beta$ | 65 | 0.20 ± 0.14 | None needed (< 0.5 ppm) |
| $^{13}C'$ | 65 | -0.09 ± 0.17 | None needed (< 0.5 ppm) |
| $^{15}N$ | 65 | -0.45 ± 0.51 | 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 57%, i.e. 666 atoms were assigned a chemical shift out of a possible 1163. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

| | Total | $^1$H | $^{13}$C | $^{15}$N |
|---|---|---|---|---|
| Backbone | 280/284 (99%) | 114/116 (98%) | 112/114 (98%) | 54/54 (100%) |
| Sidechain | 342/420 (81%) | 223/269 (83%) | 111/129 (86%) | 8/22 (36%) |

*Continued on next page...*

*Continued from previous page...*

| | Total | $^1$H | $^{13}$C | $^{15}$N |
|---|---|---|---|---|
| Aromatic | 44/98 (45%) | 25/48 (52%) | 18/47 (38%) | 1/3 (33%) |
| Sugar | 0/216 (0%) | 0/126 (0%) | 0/90 (0%) | 0/0 (—%) |
| Base | 0/145 (0%) | 0/91 (0%) | 0/29 (0%) | 0/25 (0%) |
| Overall | 666/1163 (57%) | 362/650 (56%) | 241/409 (59%) | 63/104 (61%) |

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 59%, i.e. 798 atoms were assigned a chemical shift out of a possible 1342. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

| | Total | $^1$H | $^{13}$C | $^{15}$N |
|---|---|---|---|---|
| Backbone | 337/355 (95%) | 137/145 (94%) | 135/142 (95%) | 65/68 (96%) |
| Sidechain | 417/528 (79%) | 270/339 (80%) | 138/162 (85%) | 9/27 (33%) |
| Aromatic | 44/98 (45%) | 25/48 (52%) | 18/47 (38%) | 1/3 (33%) |
| Sugar | 0/216 (0%) | 0/126 (0%) | 0/90 (0%) | 0/0 (—%) |
| Base | 0/145 (0%) | 0/91 (0%) | 0/29 (0%) | 0/25 (0%) |
| Overall | 798/1342 (59%) | 432/749 (58%) | 291/470 (62%) | 75/123 (61%) |

### 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 | 113 | PRO | HD2 | 1.61 | $1.93 - 5.38$ | -5.9 |

### 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:
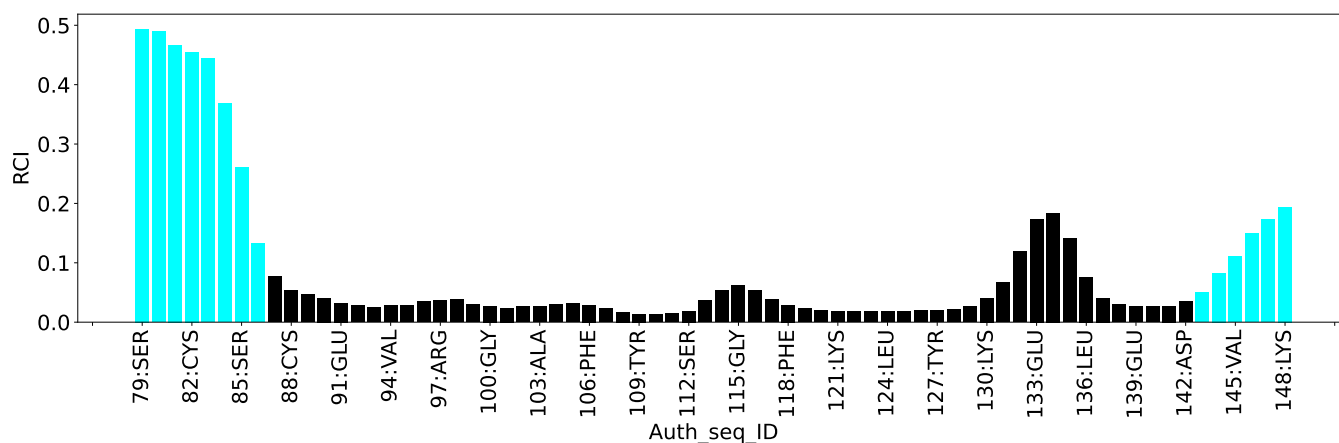
## 7.2 Chemical shift list 2

File name: working_cs.cif

Chemical shift list name: *DNAfor_ chemshift*

### 7.2.1 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 | 50 |
|---|---|
| Number of shifts mapped to atoms | 50 |
| 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 |

### 7.2.2 Chemical shift referencing (i)

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

### 7.2.3 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 3%, i.e. 36 atoms were assigned a chemical shift out of a possible 1163. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

| | Total | $^1$H | $^{13}$C | $^{15}$N |
|---|---|---|---|---|
| Backbone | 0/284 (0%) | 0/116 (0%) | 0/114 (0%) | 0/54 (0%) |

*Continued on next page...*

*Continued from previous page...*

|           | **Total**       | **$^1$H**        | **$^{13}$C**    | **$^{15}$N**    |
|-----------|-----------------|------------------|-----------------|-----------------|
| Sidechain | 0/420 (0%)      | 0/269 (0%)       | 0/129 (0%)      | 0/22 (0%)       |
| Aromatic  | 0/98 (0%)       | 0/48 (0%)        | 0/47 (0%)       | 0/3 (0%)        |
| Sugar     | 19/216 (9%)     | 19/126 (15%)     | 0/90 (0%)       | 0/0 (—%)        |
| Base      | 17/145 (12%)    | 17/91 (19%)      | 0/29 (0%)       | 0/25 (0%)       |
| Overall   | 36/1163 (3%)    | 36/650 (6%)      | 0/409 (0%)      | 0/104 (0%)      |

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 3%, i.e. 36 atoms were assigned a chemical shift out of a possible 1342. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

|           | **Total**       | **$^1$H**        | **$^{13}$C**    | **$^{15}$N**    |
|-----------|-----------------|------------------|-----------------|-----------------|
| Backbone  | 0/355 (0%)      | 0/145 (0%)       | 0/142 (0%)      | 0/68 (0%)       |
| Sidechain | 0/528 (0%)      | 0/339 (0%)       | 0/162 (0%)      | 0/27 (0%)       |
| Aromatic  | 0/98 (0%)       | 0/48 (0%)        | 0/47 (0%)       | 0/3 (0%)        |
| Sugar     | 19/216 (9%)     | 19/126 (15%)     | 0/90 (0%)       | 0/0 (—%)        |
| Base      | 17/145 (12%)    | 17/91 (19%)      | 0/29 (0%)       | 0/25 (0%)       |
| Overall   | 36/1342 (3%)    | 36/749 (5%)      | 0/470 (0%)      | 0/123 (0%)      |

### 7.2.4   Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

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

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

## 7.3   Chemical shift list 3

File name: working_cs.cif

Chemical shift list name: *DNArev_chemshift*

### 7.3.1   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          | 44 |
|---------------------------------|----|
| Number of shifts mapped to atoms | 44 |
| 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 |

### 7.3.2 Chemical shift referencing ⓘ

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

### 7.3.3 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 3%, i.e. 34 atoms were assigned a chemical shift out of a possible 1163. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

| | Total | $^1$H | $^{13}$C | $^{15}$N |
|---|---|---|---|---|
| Backbone | 0/284 (0%) | 0/116 (0%) | 0/114 (0%) | 0/54 (0%) |
| Sidechain | 0/420 (0%) | 0/269 (0%) | 0/129 (0%) | 0/22 (0%) |
| Aromatic | 0/98 (0%) | 0/48 (0%) | 0/47 (0%) | 0/3 (0%) |
| Sugar | 19/216 (9%) | 19/126 (15%) | 0/90 (0%) | 0/0 (—%) |
| Base | 15/145 (10%) | 15/91 (16%) | 0/29 (0%) | 0/25 (0%) |
| Overall | 34/1163 (3%) | 34/650 (5%) | 0/409 (0%) | 0/104 (0%) |

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 3%, i.e. 34 atoms were assigned a chemical shift out of a possible 1342. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

| | Total | $^1$H | $^{13}$C | $^{15}$N |
|---|---|---|---|---|
| Backbone | 0/355 (0%) | 0/145 (0%) | 0/142 (0%) | 0/68 (0%) |
| Sidechain | 0/528 (0%) | 0/339 (0%) | 0/162 (0%) | 0/27 (0%) |
| Aromatic | 0/98 (0%) | 0/48 (0%) | 0/47 (0%) | 0/3 (0%) |
| Sugar | 19/216 (9%) | 19/126 (15%) | 0/90 (0%) | 0/0 (—%) |
| Base | 15/145 (10%) | 15/91 (16%) | 0/29 (0%) | 0/25 (0%) |
| Overall | 34/1342 (3%) | 34/749 (5%) | 0/470 (0%) | 0/123 (0%) |

### 7.3.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

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

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

# 8 NMR restraints analysis ⓘ

## 8.1 Conformationally restricting restraints ⓘ

The following table provides the summary of experimentally observed NMR restraints in different categories. Restraints are classified into different categories based on the sequence separation of the atoms involved.

| Description | Value |
|---|---|
| Total distance restraints | 1039 |
| Intra-residue ($\lvert i\text{-}j \rvert$=0) | 497 |
| Sequential ($\lvert i\text{-}j \rvert$=1) | 269 |
| Medium range ($\lvert i\text{-}j \rvert$>1 and $\lvert i\text{-}j \rvert$<5) | 117 |
| Long range ($\lvert i\text{-}j \rvert$≥5) | 127 |
| Inter-chain | 29 |
| Hydrogen bond restraints | 0 |
| Disulfide bond restraints | 0 |
| Total dihedral-angle restraints | 111 |
| Number of unmapped restraints | 0 |
| Number of restraints per residue | 12.6 |
| Number of long range restraints per residue[1] | 1.4 |

[1]Long range hydrogen bonds and disulfide bonds are counted as long range restraints while calculating the number of long range restraints per residue

## 8.2 Residual restraint violations ⓘ

This section provides the overview of the restraint violations analysis. The violations are binned as small, medium and large violations based on its absolute value. Average number of violations per model is calculated by dividing the total number of violations in each bin by the size of the ensemble.

### 8.2.1 Average number of distance violations per model ⓘ

Distance violations less than 0.1 Å
are not included in the calculation.

| Bins (Å) | Average number of violations per model | Max (Å) |
|---|---|---|
| 0.1-0.2 (Small) | 0.1 | 0.19 |
| 0.2-0.5 (Medium) | 0.2 | 0.5 |
| >0.5 (Large) | None | None |

## 8.2.2 Average number of dihedral-angle violations per model ⓘ

Dihedral-angle violations less than 1° are not included in the calculation.

| Bins (°) | Average number of violations per model | Max (°) |
|---|---|---|
| 1.0-10.0 (Small) | 1.9 | 3.41 |
| 10.0-20.0 (Medium) | None | None |
| >20.0 (Large) | None | None |

# 9   Distance violation analysis (i)

## 9.1   Summary of distance violations (i)

The following table shows the summary of distance violations in different restraint categories based on the sequence separation of the atoms involved. Each category is further sub-divided into three sub-categories based on the atoms involved. Violations less than 0.1 Å
are not included in the statistics.

| Restraints type | Count | $\%^1$ | Violated[3] | | | Consistently Violated[4] | | |
|---|---|---|---|---|---|---|---|---|
| | | | Count | $\%^2$ | $\%^1$ | Count | $\%^2$ | $\%^1$ |
| Intra-residue (\|i-j\|=0) | 497 | 47.8 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Backbone-Backbone | 70 | 6.7 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Backbone-Sidechain | 287 | 27.6 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Sidechain-Sidechain | 140 | 13.5 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Sequential (\|i-j\|=1) | 269 | 25.9 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Backbone-Backbone | 94 | 9.0 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Backbone-Sidechain | 143 | 13.8 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Sidechain-Sidechain | 32 | 3.1 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Medium range (\|i-j\|>1 & \|i-j\|<5) | 117 | 11.3 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Backbone-Backbone | 28 | 2.7 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Backbone-Sidechain | 64 | 6.2 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Sidechain-Sidechain | 25 | 2.4 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Long range (\|i-j\|≥5) | 127 | 12.2 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Backbone-Backbone | 18 | 1.7 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Backbone-Sidechain | 47 | 4.5 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Sidechain-Sidechain | 62 | 6.0 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Inter-chain | 29 | 2.8 | 1 | 3.4 | 0.1 | 0 | 0.0 | 0.0 |
| Backbone-Backbone | 0 | 0.0 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Backbone-Sidechain | 10 | 1.0 | 1 | 10.0 | 0.1 | 0 | 0.0 | 0.0 |
| Sidechain-Sidechain | 19 | 1.8 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Hydrogen bond | 0 | 0.0 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Disulfide bond | 0 | 0.0 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Total | 1039 | 100.0 | 1 | 0.1 | 0.1 | 0 | 0.0 | 0.0 |
| Backbone-Backbone | 210 | 20.2 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |
| Backbone-Sidechain | 551 | 53.0 | 1 | 0.2 | 0.1 | 0 | 0.0 | 0.0 |
| Sidechain-Sidechain | 278 | 26.8 | 0 | 0.0 | 0.0 | 0 | 0.0 | 0.0 |

[1] percentage calculated with respect to the total number of distance restraints, [2] percentage calculated with respect to the number of restraints in a particular restraint category, [3] violated in at least one model, [4] violated in all the models

### 9.1.1 Bar chart : Distribution of distance restraints and violations ⓘ
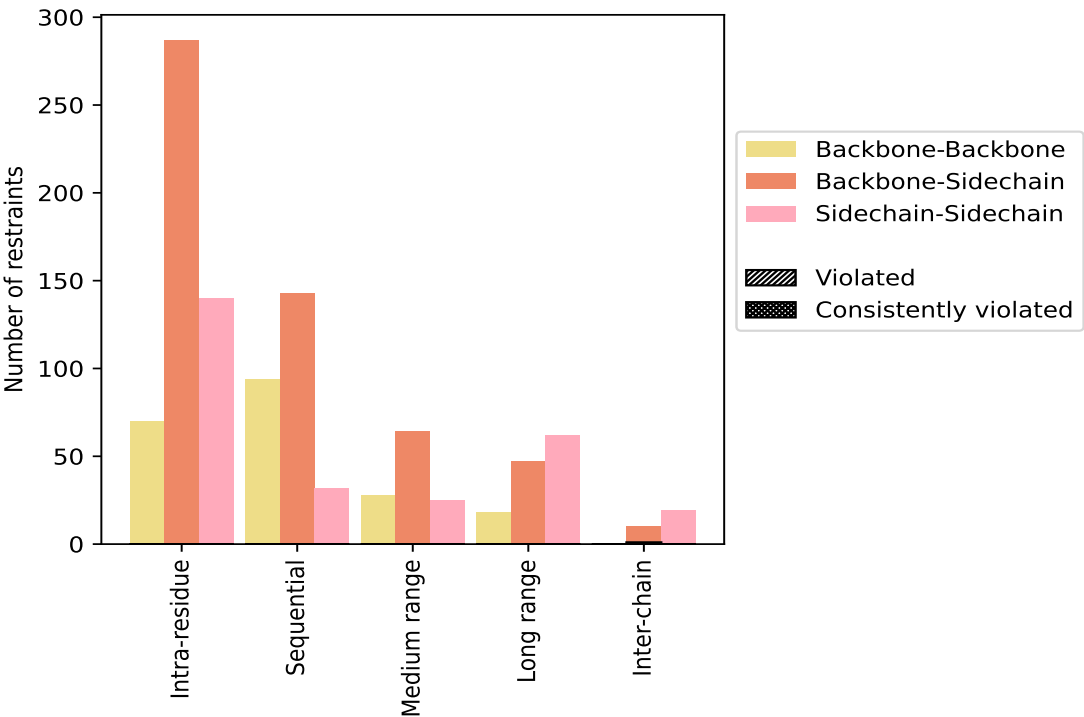


Violated and consistently violated restraints are shown using different hatch patterns in their respective categories. The hydrogen bonds and disulfied bonds are counted in their appropriate category on the x-axis

## 9.2 Distance violation statistics for each model ⓘ

The following table provides the distance violation statistics for each model in the ensemble. Violations less than 0.1 Å
are not included in the statistics.

| Model ID | Number of violations | | | | | | Mean (Å) | Max (Å) | SD[6] (Å) | Median (Å) |
|---|---|---|---|---|---|---|---|---|---|---|
| | IR[1] | SQ[2] | MR[3] | LR[4] | IC[5] | Total | | | | |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 0 | 0 | 0 | 0 | 1 | 1 | 0.48 | 0.48 | 0.0 | 0.48 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 5 | 0 | 0 | 0 | 0 | 1 | 1 | 0.19 | 0.19 | 0.0 | 0.19 |
| 6 | 0 | 0 | 0 | 0 | 1 | 1 | 0.28 | 0.28 | 0.0 | 0.28 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 9 | 0 | 0 | 0 | 0 | 1 | 1 | 0.5 | 0.5 | 0.0 | 0.5 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |

*Continued from previous page...*

| Model ID | Number of violations | | | | | | Mean (Å) | Max (Å) | SD[6] (Å) | Median (Å) |
| | IR[1] | SQ[2] | MR[3] | LR[4] | IC[5] | Total | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 11 | 0 | 0 | 0 | 0 | 1 | 1 | 0.27 | 0.27 | 0.0 | 0.27 |
| 12 | 0 | 0 | 0 | 0 | 1 | 1 | 0.12 | 0.12 | 0.0 | 0.12 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 20 | 0 | 0 | 0 | 0 | 1 | 1 | 0.3 | 0.3 | 0.0 | 0.3 |

[1]Intra-residue restraints, [2]Sequential restraints, [3]Medium range restraints, [4]Long range restraints, [5]Inter-chain restraints, [6]Standard deviation

### 9.2.1　Bar graph : Distance Violation statistics for each model ⓘ
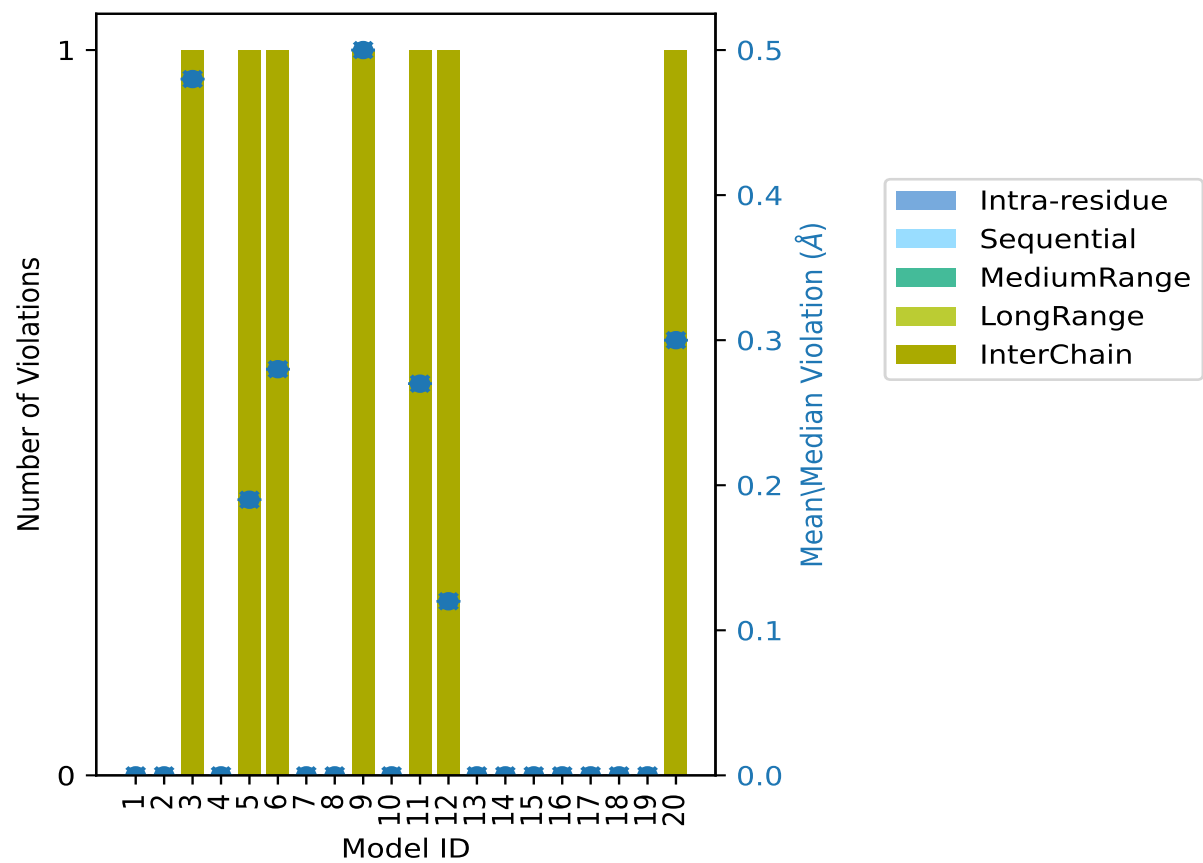


The mean(dot),median(x) and the standard deviation are shown in blue with respect to the y axis on the right

## 9.3   Distance violation statistics for the ensemble (i)

Violation analysis may find that some restraints are violated in few models and some are violated in most of models. The following table provides this information as number of violated restraints for a given fraction of the ensemble. In total, 1038(IR:497, SQ:269, MR:117, LR:127, IC:28) restraints are not violated in the ensemble.

| Number of violated restraints | | | | | | Fraction of the ensemble | |
|---|---|---|---|---|---|---|---|
| IR[1] | SQ[2] | MR[3] | LR[4] | IC[5] | Total | Count[6] | % |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 5.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 2 | 10.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 3 | 15.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 4 | 20.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 5 | 25.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 6 | 30.0 |
| 0 | 0 | 0 | 0 | 1 | 1 | 7 | 35.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 8 | 40.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 9 | 45.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 10 | 50.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 11 | 55.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 12 | 60.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 13 | 65.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 14 | 70.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 15 | 75.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 16 | 80.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 17 | 85.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 18 | 90.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 19 | 95.0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 20 | 100.0 |

[1]Intra-residue restraints, [2]Sequential restraints, [3]Medium range restraints, [4]Long range restraints, [5]Inter-chain restraints, [6] Number of models with violations

### 9.3.1 Bar graph : Distance violation statistics for the ensemble ⓘ
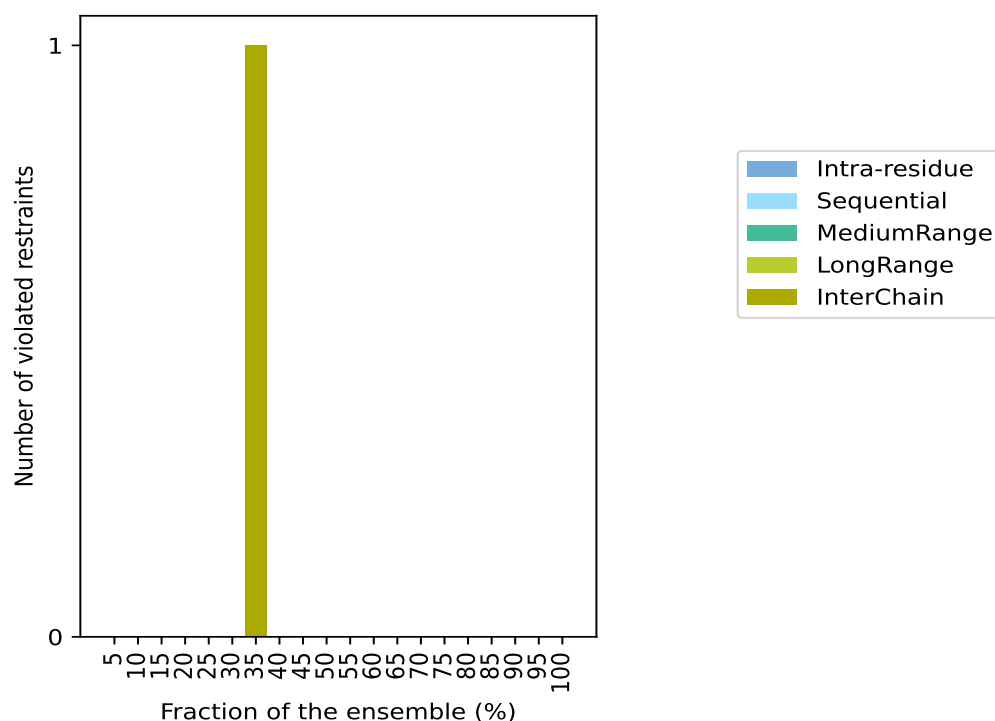


## 9.4 Most violated distance restraints in the ensemble ⓘ

### 9.4.1 Histogram : Distribution of mean distance violations ⓘ

The following histogram shows the distribution of the average value of the violation. The average is calculated for each restraint that is violated in more than one model over all the violated models in the ensemble
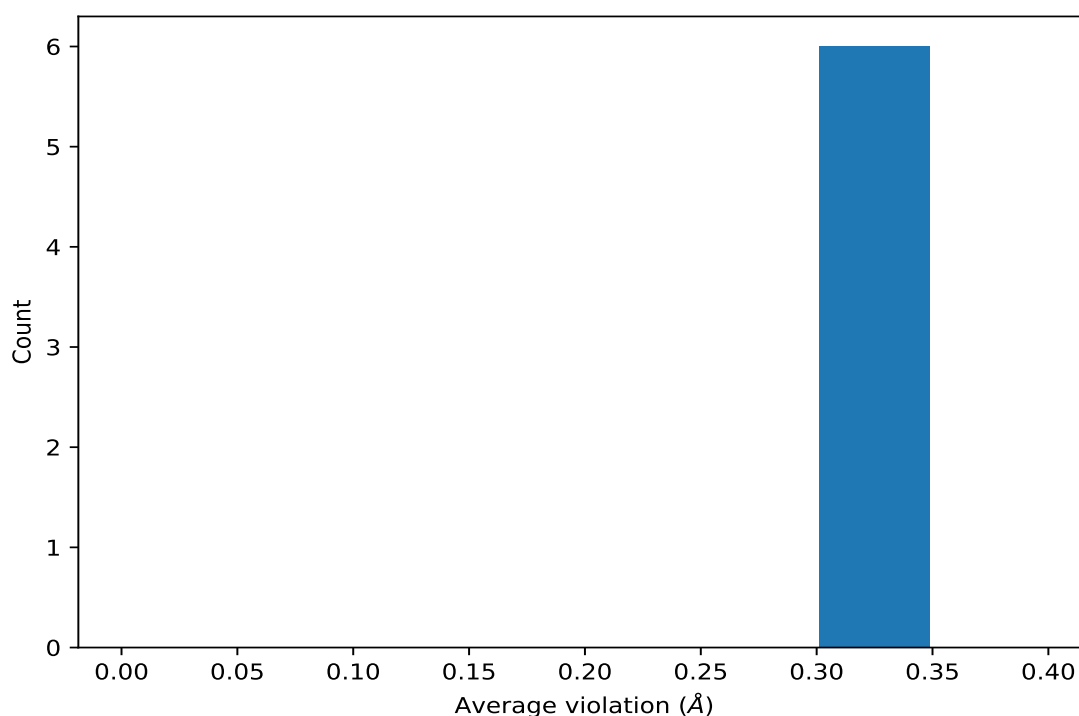
### 9.4.2   Table: Most violated distance restraints ⓘ

The following table provides the mean and the standard deviation of the violation for each restraint sorted by number of violated models and the mean value. The Key (restraint list ID, restraint ID) is the unique identifier for a given restraint. Rows with same key represent combinatorial or ambiguous restraints and are counted as a single restraint.

| Key | Atom-1 | Atom-2 | Models[1] | Mean (Å) | SD[1] (Å) | Median (Å) |
|---|---|---|---|---|---|---|
| (1,919) | 1:119:A:ARG:HB3 | 3:215:C:5CM:H5A1 | 7 | 0.31 | 0.13 | 0.28 |
| (1,919) | 1:119:A:ARG:HB3 | 3:215:C:5CM:H5A2 | 7 | 0.31 | 0.13 | 0.28 |
| (1,919) | 1:119:A:ARG:HB3 | 3:215:C:5CM:H5A3 | 7 | 0.31 | 0.13 | 0.28 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A1 | 7 | 0.31 | 0.13 | 0.28 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A2 | 7 | 0.31 | 0.13 | 0.28 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A3 | 7 | 0.31 | 0.13 | 0.28 |

[1]Number of violated models, [2]Standard deviation

## 9.5   All violated distance restraints ⓘ

### 9.5.1   Histogram : Distribution of distance violations ⓘ

The following histogram shows the distribution of the absolute value of the violation for all violated restraints in the ensemble.

### 9.5.2   Table : All distance violations ⓘ

The following table lists the absolute value of the violation for each restraint in the ensemble sorted by its value. The Key (restraint list ID, restraint ID) is the unique identifier for a given restraint. Rows with same key represent combinatorial or ambiguous restraints and are counted as a single restraint.
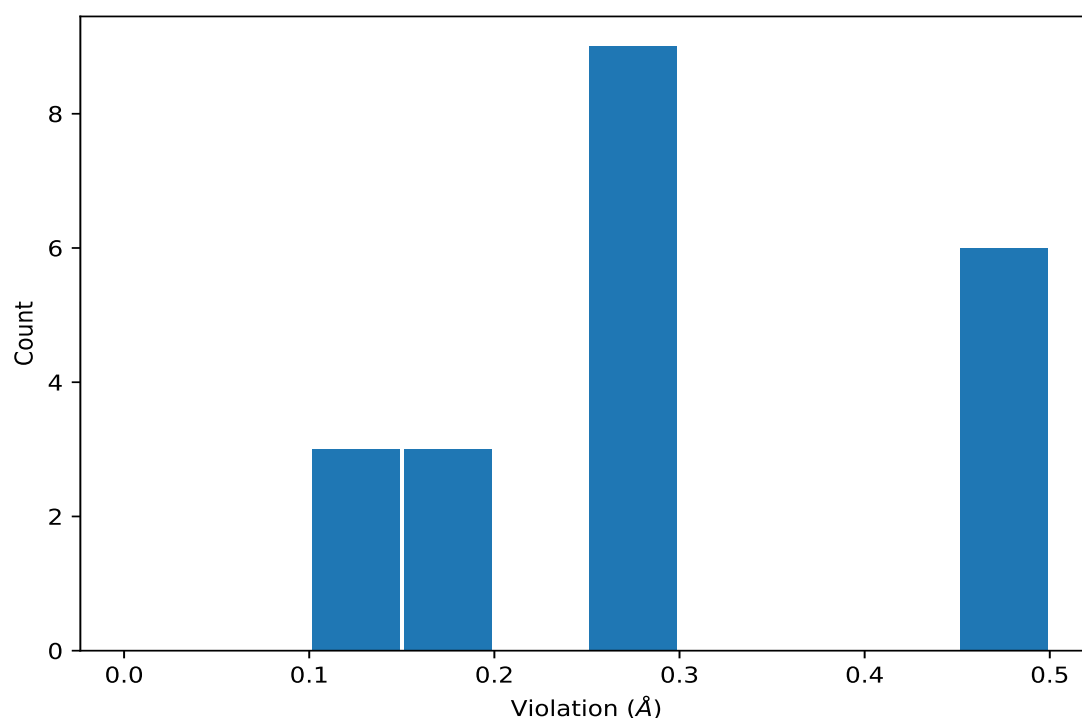
| Key | Atom-1 | Atom-2 | Model ID | Violation (Å) |
|---|---|---|---|---|
| (1,919) | 1:119:A:ARG:HB3 | 3:215:C:5CM:H5A1 | 9 | 0.5 |
| (1,919) | 1:119:A:ARG:HB3 | 3:215:C:5CM:H5A2 | 9 | 0.5 |
| (1,919) | 1:119:A:ARG:HB3 | 3:215:C:5CM:H5A3 | 9 | 0.5 |
| (1,919) | 1:119:A:ARG:HB3 | 3:215:C:5CM:H5A1 | 3 | 0.48 |
| (1,919) | 1:119:A:ARG:HB3 | 3:215:C:5CM:H5A2 | 3 | 0.48 |
| (1,919) | 1:119:A:ARG:HB3 | 3:215:C:5CM:H5A3 | 3 | 0.48 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A1 | 20 | 0.3 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A2 | 20 | 0.3 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A3 | 20 | 0.3 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A1 | 6 | 0.28 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A2 | 6 | 0.28 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A3 | 6 | 0.28 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A1 | 11 | 0.27 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A2 | 11 | 0.27 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A3 | 11 | 0.27 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A1 | 5 | 0.19 |

*Continued on next page...*

*Continued from previous page...*

| Key | Atom-1 | Atom-2 | Model ID | Violation (Å) |
|-----|--------|--------|----------|---------------|
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A2 | 5 | 0.19 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A3 | 5 | 0.19 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A1 | 12 | 0.12 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A2 | 12 | 0.12 |
| (1,919) | 1:119:A:ARG:HB2 | 3:215:C:5CM:H5A3 | 12 | 0.12 |

# 10 Dihedral-angle violation analysis ⓘ

## 10.1 Summary of dihedral-angle violations ⓘ

The following table provides the summary of dihedral-angle violations in different dihedral-angle types. Violations less than 1° are not included in the calculation.

| Angle type | Count | $\%^1$ | Violated[3] | | | Consistently Violated[4] | | |
|---|---|---|---|---|---|---|---|---|
| | | | Count | $\%^2$ | $\%^1$ | Count | $\%^2$ | $\%^1$ |
| CHI1 | 16 | 14.4 | 3 | 18.8 | 2.7 | 0 | 0.0 | 0.0 |
| PHI | 48 | 43.2 | 3 | 6.2 | 2.7 | 0 | 0.0 | 0.0 |
| PSI | 47 | 42.3 | 5 | 10.6 | 4.5 | 0 | 0.0 | 0.0 |
| Total | 111 | 100.0 | 11 | 9.9 | 9.9 | 0 | 0.0 | 0.0 |

[1] percentage calculated with respect to total number of dihedral-angle restraints, [2] percentage calculated with respect to number of restraints in a particular dihedral-angle type, [3] violated in at least one model, [4] violated in all the models

### 10.1.1 Bar chart : Distribution of dihedral-angles and violations ⓘ
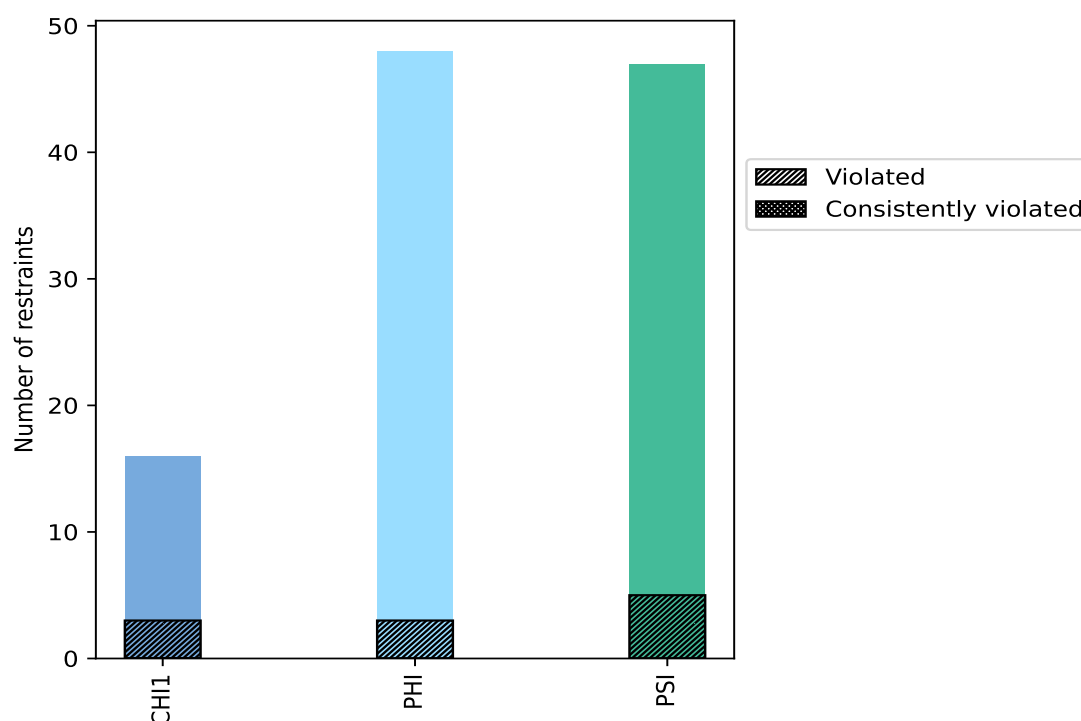


Violated and consistently violated restraints are shown using different hatch patterns in their respective categories

## 10.2 Dihedral-angle violation statistics for each model (i)

The following table provides the dihedral-angle violation statistics for each model in the ensemble. Violations less than 1° are not included in the statistics.

| Model ID | Number of violations | | | | Mean (°) | Max (°) | SD (°) | Median (°) |
|---|---|---|---|---|---|---|---|---|
| | CHI1 | PHI | PSI | Total | | | | |
| 1 | 2 | 0 | 1 | 3 | 1.27 | 1.44 | 0.15 | 1.29 |
| 2 | 1 | 0 | 0 | 1 | 1.02 | 1.02 | 0.0 | 1.02 |
| 3 | 0 | 0 | 1 | 1 | 1.55 | 1.55 | 0.0 | 1.55 |
| 4 | 1 | 0 | 0 | 1 | 1.96 | 1.96 | 0.0 | 1.96 |
| 5 | 1 | 0 | 2 | 3 | 1.38 | 1.69 | 0.25 | 1.39 |
| 6 | 0 | 0 | 1 | 1 | 1.33 | 1.33 | 0.0 | 1.33 |
| 7 | 1 | 1 | 0 | 2 | 1.05 | 1.08 | 0.03 | 1.05 |
| 8 | 0 | 1 | 3 | 4 | 1.93 | 3.41 | 0.85 | 1.45 |
| 9 | 1 | 0 | 1 | 2 | 1.98 | 2.41 | 0.43 | 1.98 |
| 10 | 0 | 0 | 2 | 2 | 1.58 | 1.78 | 0.2 | 1.58 |
| 11 | 2 | 0 | 1 | 3 | 1.55 | 2.32 | 0.55 | 1.2 |
| 12 | 1 | 0 | 1 | 2 | 2.4 | 2.76 | 0.36 | 2.4 |
| 13 | 1 | 0 | 1 | 2 | 1.35 | 1.6 | 0.26 | 1.35 |
| 14 | 0 | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 15 | 1 | 1 | 0 | 2 | 1.17 | 1.34 | 0.17 | 1.17 |
| 16 | 2 | 1 | 0 | 3 | 1.6 | 2.14 | 0.38 | 1.39 |
| 17 | 2 | 0 | 1 | 3 | 1.72 | 2.69 | 0.69 | 1.31 |
| 18 | 2 | 0 | 0 | 2 | 1.58 | 1.98 | 0.4 | 1.58 |
| 19 | 0 | 0 | 1 | 1 | 1.63 | 1.63 | 0.0 | 1.63 |
| 20 | 1 | 0 | 0 | 1 | 1.23 | 1.23 | 0.0 | 1.23 |

### 10.2.1 Bar graph : Dihedral violation statistics for each model ⓘ
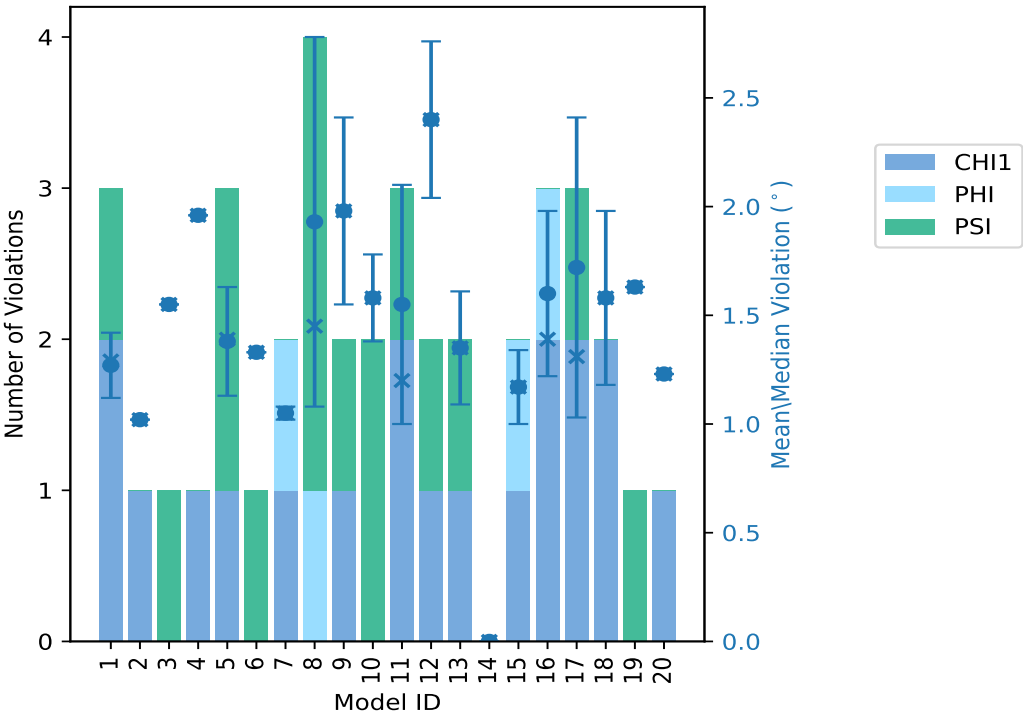


The mean(dot),median(x) and the standard deviation are shown in blue with respect to the y axis on the right

## 10.3 Dihedral-angle violation statistics for the ensemble ⓘ

Violation analysis may find that some restraints are violated in very few models and some are violated in most of models. The following table provides this information as number of violated restraints for a given fraction of ensemble.

| Number of violated restraints | | | | Fraction of the ensemble | |
|---|---|---|---|---|---|
| CHI1 | PHI | PSI | Total | Count[1] | % |
| 1 | 2 | 3 | 6 | 1 | 5.0 |
| 0 | 1 | 1 | 2 | 2 | 10.0 |
| 0 | 0 | 0 | 0 | 3 | 15.0 |
| 0 | 0 | 0 | 0 | 4 | 20.0 |
| 0 | 0 | 0 | 0 | 5 | 25.0 |
| 1 | 0 | 0 | 1 | 6 | 30.0 |
| 0 | 0 | 0 | 0 | 7 | 35.0 |
| 0 | 0 | 0 | 0 | 8 | 40.0 |
| 0 | 0 | 0 | 0 | 9 | 45.0 |
| 0 | 0 | 0 | 0 | 10 | 50.0 |
| 0 | 0 | 1 | 1 | 11 | 55.0 |

*Continued on next page...*

*Continued from previous page...*

| Number of violated restraints | | | | Fraction of the ensemble | |
| CHI1 | PHI | PSI | Total | Count[1] | % |
| --- | --- | --- | --- | --- | --- |
| 1 | 0 | 0 | 1 | 12 | 60.0 |
| 0 | 0 | 0 | 0 | 13 | 65.0 |
| 0 | 0 | 0 | 0 | 14 | 70.0 |
| 0 | 0 | 0 | 0 | 15 | 75.0 |
| 0 | 0 | 0 | 0 | 16 | 80.0 |
| 0 | 0 | 0 | 0 | 17 | 85.0 |
| 0 | 0 | 0 | 0 | 18 | 90.0 |
| 0 | 0 | 0 | 0 | 19 | 95.0 |
| 0 | 0 | 0 | 0 | 20 | 100.0 |

[1] Number of models with violations

### 10.3.1   Bar graph : Dihedral-angle Violation statistics for the ensemble ⓘ
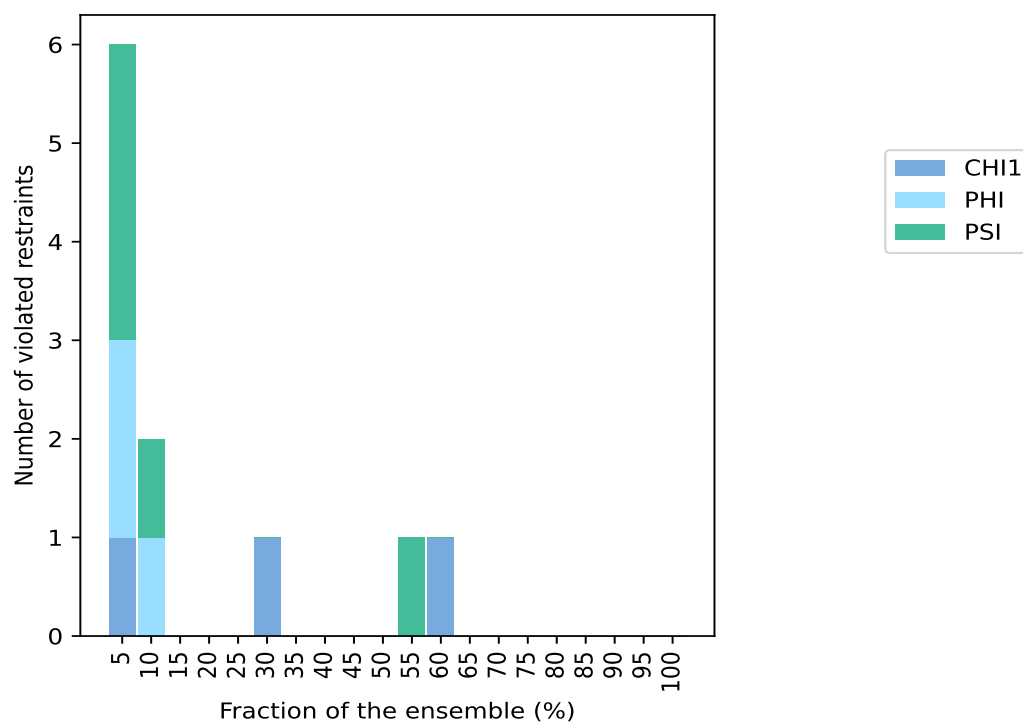


## 10.4   Most violated dihedral-angle restraints in the ensemble ⓘ

### 10.4.1   Histogram : Distribution of mean dihedral-angle violations ⓘ

The following histogram shows the distribution of the average value of the violation. The average is calculated for each restraint that is violated in more than one model over all the violated models

in the ensemble
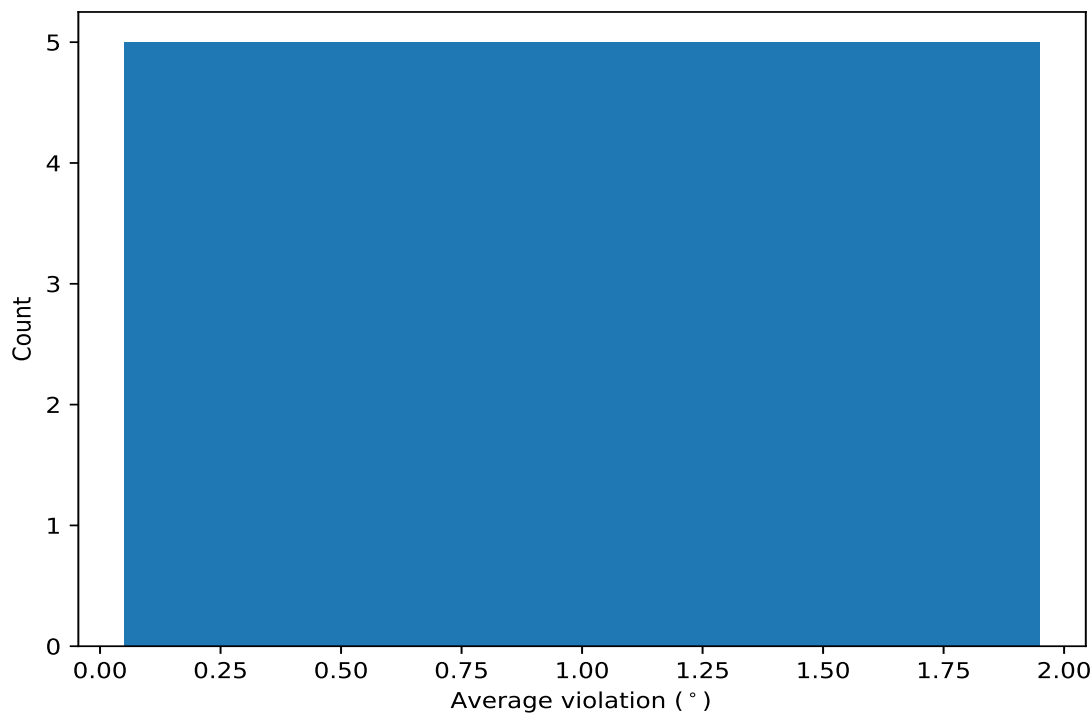


### 10.4.2 Table: Most violated dihedral-angle restraints ⓘ

The following table provides the mean and the standard deviation of the violation for each restraint sorted by number of violated models and the mean value. The Key (restraint list ID, restraint ID) is the unique identifier for a given restraint.

| Key | Atom-1 | Atom-2 | Atom-3 | Atom-4 | Models[1] | Mean | SD[2] | Median |
|------|--------------|---------------|---------------|---------------|------|------|------|--------|
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 12 | 1.76 | 0.63 | 1.82 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 11 | 1.65 | 0.29 | 1.55 |
| (1,110) | 1:141:A:PHE:N | 1:141:A:PHE:CA | 1:141:A:PHE:CB | 1:141:A:PHE:CG | 6 | 1.21 | 0.11 | 1.19 |
| (1,38) | 1:108:A:VAL:C | 1:109:A:TYR:N | 1:109:A:TYR:CA | 1:109:A:TYR:C | 2 | 1.31 | 0.03 | 1.31 |
| (1,37) | 1:108:A:VAL:N | 1:108:A:VAL:CA | 1:108:A:VAL:C | 1:109:A:TYR:N | 2 | 1.25 | 0.18 | 1.25 |

[1] Number of violated models, [2]Standard deviation, All angle values are in degree (°)

## 10.5 All violated dihedral-angle restraints ⓘ

### 10.5.1 Histogram : Distribution of violations ⓘ

The following histogram shows the distribution of the absolute value of the violation for all violated restraints in the ensemble.
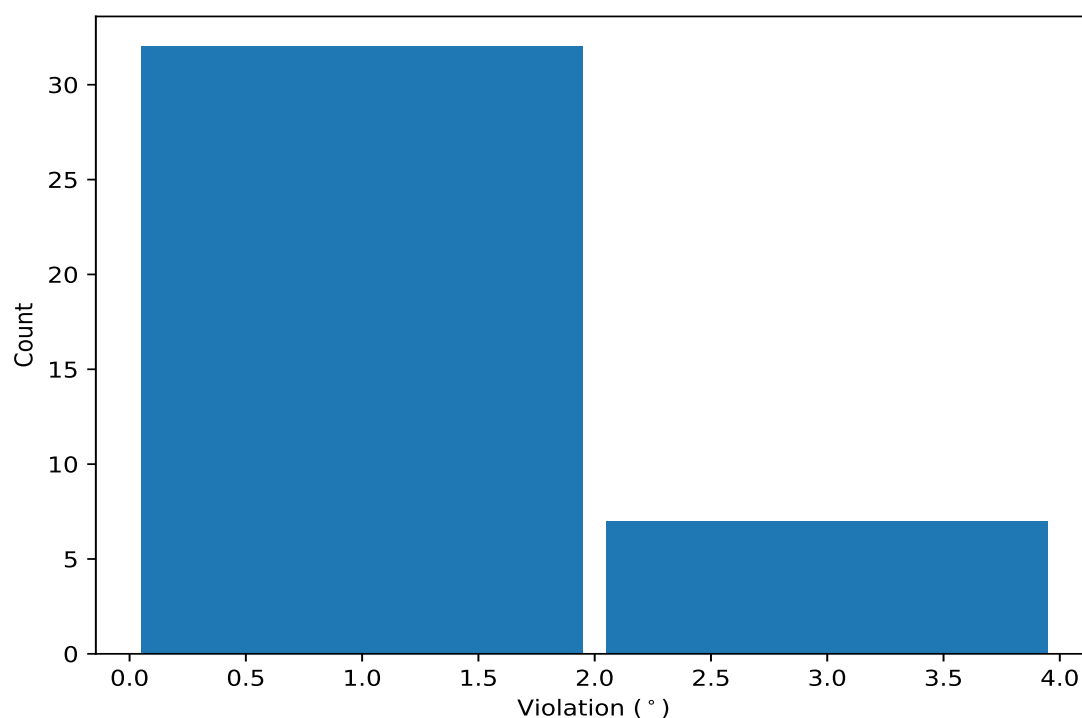
## 10.5.2 Table: All violated dihedral-angle restraints ⓘ

The following table lists the absolute value of the violation for each restraint in the ensemble sorted by its value. The Key (restraint list ID, restraint ID) is the unique identifier for a given restraint.

| Key | Atom-1 | Atom-2 | Atom-3 | Atom-4 | Model ID | Violation (°) |
|---|---|---|---|---|---|---|
| (1,5) | 1:89:A:GLY:C | 1:90:A:TRP:N | 1:90:A:TRP:CA | 1:90:A:TRP:C | 8 | 3.41 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 12 | 2.76 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 17 | 2.69 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 9 | 2.41 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 11 | 2.32 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 16 | 2.14 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 12 | 2.05 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 18 | 1.98 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 4 | 1.96 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 10 | 1.78 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 5 | 1.69 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 19 | 1.63 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 13 | 1.6 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 3 | 1.55 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 9 | 1.55 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 8 | 1.46 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 1 | 1.44 |
| (1,4) | 1:89:A:GLY:N | 1:89:A:GLY:CA | 1:89:A:GLY:C | 1:90:A:TRP:N | 8 | 1.44 |
| (1,37) | 1:108:A:VAL:N | 1:108:A:VAL:CA | 1:108:A:VAL:C | 1:109:A:TYR:N | 8 | 1.43 |
| (1,110) | 1:141:A:PHE:N | 1:141:A:PHE:CA | 1:141:A:PHE:CB | 1:141:A:PHE:CG | 16 | 1.39 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 5 | 1.39 |

*Continued on next page...*

*Continued from previous page...*

| Key | Atom-1 | Atom-2 | Atom-3 | Atom-4 | Model ID | Violation (°) |
|-----|--------|--------|--------|--------|----------|---------------|
| (1,27) | 1:103:A:ALA:N | 1:103:A:ALA:CA | 1:103:A:ALA:C | 1:104:A:GLY:N | 10 | 1.38 |
| (1,38) | 1:108:A:VAL:C | 1:109:A:TYR:N | 1:109:A:TYR:CA | 1:109:A:TYR:C | 15 | 1.34 |
| (1,91) | 1:140:A:ASP:N | 1:140:A:ASP:CA | 1:140:A:ASP:C | 1:141:A:PHE:N | 6 | 1.33 |
| (1,89) | 1:139:A:GLU:N | 1:139:A:GLU:CA | 1:139:A:GLU:C | 1:140:A:ASP:N | 17 | 1.31 |
| (1,110) | 1:141:A:PHE:N | 1:141:A:PHE:CA | 1:141:A:PHE:CB | 1:141:A:PHE:CG | 1 | 1.29 |
| (1,38) | 1:108:A:VAL:C | 1:109:A:TYR:N | 1:109:A:TYR:CA | 1:109:A:TYR:C | 16 | 1.28 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 20 | 1.23 |
| (1,110) | 1:141:A:PHE:N | 1:141:A:PHE:CA | 1:141:A:PHE:CB | 1:141:A:PHE:CG | 11 | 1.2 |
| (1,110) | 1:141:A:PHE:N | 1:141:A:PHE:CA | 1:141:A:PHE:CB | 1:141:A:PHE:CG | 18 | 1.19 |
| (1,110) | 1:141:A:PHE:N | 1:141:A:PHE:CA | 1:141:A:PHE:CB | 1:141:A:PHE:CG | 17 | 1.17 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 11 | 1.12 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 13 | 1.09 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 1 | 1.08 |
| (1,82) | 1:135:A:SER:C | 1:136:A:LEU:N | 1:136:A:LEU:CA | 1:136:A:LEU:C | 7 | 1.08 |
| (1,37) | 1:108:A:VAL:N | 1:108:A:VAL:CA | 1:108:A:VAL:C | 1:109:A:TYR:N | 5 | 1.07 |
| (1,110) | 1:141:A:PHE:N | 1:141:A:PHE:CA | 1:141:A:PHE:CB | 1:141:A:PHE:CG | 2 | 1.02 |
| (1,97) | 1:90:A:TRP:N | 1:90:A:TRP:CA | 1:90:A:TRP:CB | 1:90:A:TRP:CG | 7 | 1.02 |
| (1,99) | 1:98:A:LEU:N | 1:98:A:LEU:CA | 1:98:A:LEU:CB | 1:98:A:LEU:CG | 15 | 1.0 |