# WORLDWIDE PDB

## PROTEIN DATA BANK

# Full wwPDB NMR Structure Validation Report (i)

May 15, 2023 – 08:11 PM JST

|  |  |  |
|---|---|---|
| PDB ID | : | 8HEQ |
| BMRB ID | : | 25358 |
| Title | : | Solution structure of the periplasmic domain of the anti-sigma factor RsgI2 from Clostridium thermocellum |
| Authors | : | Chen, C.; Feng, Y. |
| Deposited on | : | 2022-11-08 |

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

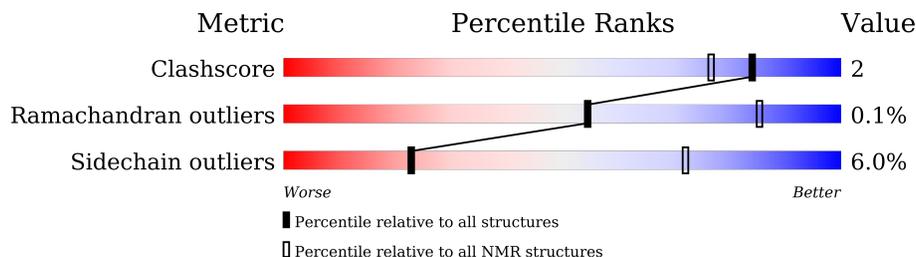|  |  |  |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| wwPDB-RCI | : | v_1n_11_5_13_A (Berjanski et al., 2005) |
| PANAV | : | Wang et al. (2010) |
| wwPDB-ShiftChecker | : | v1.2 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.32.2 |

# 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 95%.

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 | | 2 |
| Ramachandran outliers | | 0.1% |
| Sidechain outliers | | 6.0% |

*Worse* | | *Better*

▮ Percentile relative to all structures
▯ Percentile relative to all NMR structures

| 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 | 10 | 100% |
| 2 | B | 159 | 83% 6% 11% |

# 2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 14 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:1-A:10, B:11-B:78, B:89-B:162 (152) | 0.30 | 14 |

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 4 single-model clusters were found.

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

## 3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 2647 atoms, of which 1329 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Anti-sigma-I factor RsgI2.

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | | | Total | C | H | N | O | S | |
| 1 | A | 10 | 164 | 55 | 79 | 11 | 18 | 1 | 0 |

There is a discrepancy between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|:---:|:---:|:---:|:---:|:---:|:---:|
| A | 1 | MET | - | initiating methionine | UNP A3DC27 |

- Molecule 2 is a protein called Anti-sigma-I factor RsgI2.

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | | | Total | C | H | N | O | S | |
| 2 | B | 159 | 2483 | 770 | 1250 | 214 | 247 | 2 | 0 |

There are 8 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|:---:|:---:|:---:|:---:|:---:|:---:|
| B | 162 | LEU | - | expression tag | UNP A3DC27 |
| B | 163 | GLU | - | expression tag | UNP A3DC27 |
| B | 164 | HIS | - | expression tag | UNP A3DC27 |
| B | 165 | HIS | - | expression tag | UNP A3DC27 |
| B | 166 | HIS | - | expression tag | UNP A3DC27 |
| B | 167 | HIS | - | expression tag | UNP A3DC27 |
| B | 168 | HIS | - | expression tag | UNP A3DC27 |
| B | 169 | HIS | - | expression tag | UNP A3DC27 |

# 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: Anti-sigma-I factor RsgI2

Chain A:

100%

There are no outlier residues in this chain.

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:

83% 6% 11%



## 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: Anti-sigma-I factor RsgI2

Chain A:

100%

There are no outlier residues in this chain.

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:

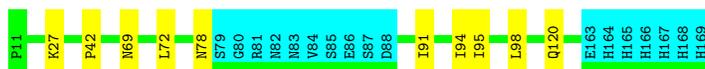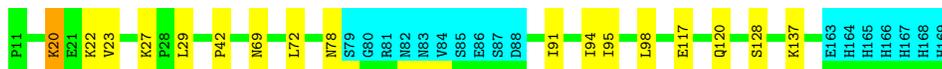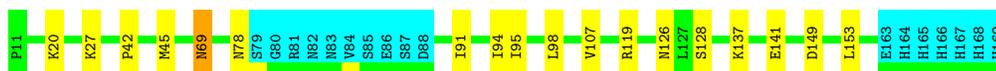79% 10% · 11%

### 4.2.2 Score per residue for model 2

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:

100%

There are no outlier residues in this chain.

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:

78%          11%    ·    11%

P11 K20 K27 P42 M45 N69 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I94 I95 L98 V107 R119 N126 L127 S128 K137 E141 D149 L153 E163 H164 H165 H166 H167 H168 H169

### 4.2.3 Score per residue for model 3

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:

80%                    20%

M1 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:

79%             9%    ·    11%

P11 K27 P28 P42 Y48 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I94 I95 L98 V101 V107 S128 D149 S154 E163 H164 H165 H166 H167 H168 H169

### 4.2.4 Score per residue for model 4

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:

70%                    30%

M1 Y2 A3 D6 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:

75%          14%         11%

P11 I17 K20 D32 P42 M45 Y48 K59 N69 L72 F73 S74 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I94 I95 L98 V101 E110 Q120 G130 K137 E141 E157 E163 H164 H165 H166 H167

H168
H169

### 4.2.5 Score per residue for model 5

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:

80%                                    20%

M1    I9 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:

81%                          8%      11%

P11 K27 P28 I64 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I94 I95 L98 E117 N126 S154 E163 H164 H165 H166 H167 H168 H169

### 4.2.6 Score per residue for model 6

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:

100%

There are no outlier residues in this chain.

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:

79%                          10%  ·  11%

P11 K20 D32 P42 M45 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I95 E110 R119 Q120 E148 D149 A150 R151 N152 L153 E163 H164 H165 H166 H167 H168 H169

### 4.2.7 Score per residue for model 7

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:

90%                                    10%

M1    D6 N10

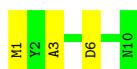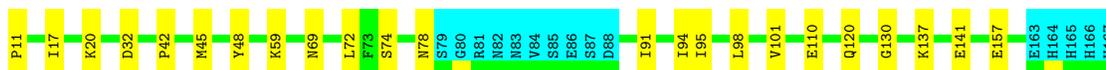• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:

75%                          14%  ·  11%

P11 K27 D32 I36 P42 K43 D44 M45 Y63 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I94 I95 L98 K103 E110 R119 Q120 D149 L153 S154 I158 E163 H164 H165 H166 H167 H168 H169

## 4.2.8 Score per residue for model 8

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A: 90% 10%

M1 D6 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B: 74% 15% 11%

P11 K20 K27 D32 N69 L72 F73 S74 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 K89 G90 I91 I94 I95 L98 I112 E116 R119 Q120 N126 L127 S128 K137 R151 E157 K161 L162 E163 H164 H165 H166 H167 H168 H169

## 4.2.9 Score per residue for model 9

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A: 90% 10%

M1 Y2 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B: 81% 8% 11%

P11 K20 K27 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I94 I95 L98 D104 V107 N126 D149 L153 E163 H164 H165 H166 H167 H168 H169
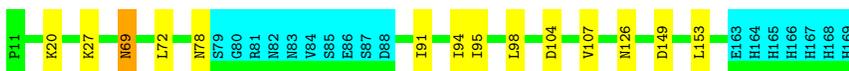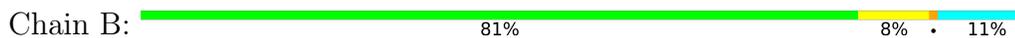
## 4.2.10 Score per residue for model 10

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A: 90% 10%

M1 D6 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B: 79% 9% 11%

P11 Y48 N69 L72 F73 S74 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 K89 G90 I91 I95 V101 I112 R119 Q120 S128 D149 L153 E157 K161 L162 E163 H164 H165 H166 H167 H168 H169

## 4.2.11 Score per residue for model 11

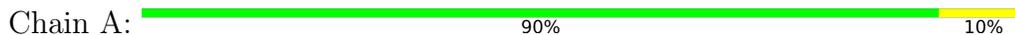• Molecule 1: Anti-sigma-I factor RsgI2

Chain A: 80% 20%

M1 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B: 77% 11% · 11%

P11 V23 K27 P42 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I94 I95 L98 K103 V107 R119 Q120 Q125 S128 K137 E141 S154 E163 H164 H165 H166 H167 H168 H169

## 4.2.12 Score per residue for model 12

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A: 90% 10%

M1 D6 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B: 80% 9% 11%

P11 K27 D56 K59 N69 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I94 I95 L98 V107 N126 K137 E141 D149 A150 R151 E163 H164 H165 H166 H167 H168 H169

## 4.2.13 Score per residue for model 13

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A: 80% 20%

M1 D6 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B: 79% 10% 11%

P11 I17 K20 E21 P42 M45 L51 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I95 R119 Q120 N126 K137 E141 E163 H164 H165 H166 H167 H168 H169

### 4.2.14 Score per residue for model 14 (medoid)

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:                                100%

There are no outlier residues in this chain.

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:                          78%                    11%    ·    11%

P11 D18 K27 P42 K43 D44 M45 I54 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 K89 G90 I91 I94 I95 L98 K99 K103 V107 S128 E163 H164 H165 H166 H167 H168 H169

### 4.2.15 Score per residue for model 15

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:                                90%                        10%
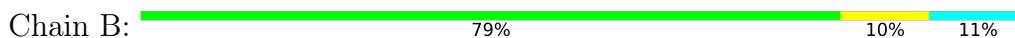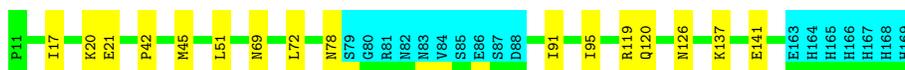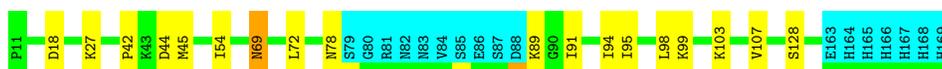
M1 D6 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:                            79%                    11%    11%

P11 K27 D32 P42 M45 N69 L72 F73 S74 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I95 I112 R119 Q120 S154 E157 I158 E163 H164 H165 H166 H167 H168 H169

### 4.2.16 Score per residue for model 16

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:                                90%                        10%

M1 D6 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:                              81%                    9%    11%

P11 D32 I36 Y63 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I94 I95 L98 E116 Q120 N126 E163 H164 H165 H166 H167 H168 H169

#### 4.2.17 Score per residue for model 17

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:
90%                                                                    10%

M1 D8 I9 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:
77%                                                  13%          11%

P11 K27 P42 M45 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 K89 G90 I91 I94 I95 L98 Q120 K121 Q125 S128 R131 K137 E141 D149 L153 E163 H164 H165 H166 H167 H168 H169

#### 4.2.18 Score per residue for model 18

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:
100%

There are no outlier residues in this chain.
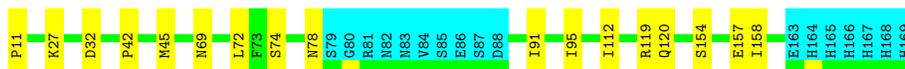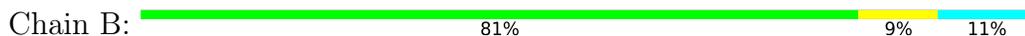
• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:
80%                                                     9%      11%

P11 D32 K53 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 K89 G90 I91 I95 E116 K121 Q125 S128 E157 K161 L162 E163 H164 H165 H166 H167 H168 H169
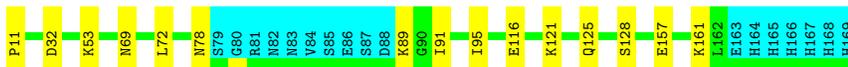
#### 4.2.19 Score per residue for model 19

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:
100%

There are no outlier residues in this chain.

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:
77%                                                  12%    ·    11%

P11 K20 D32 K59 N65 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 I94 I95 L98 V107 Q120 K121 Q125 N126 K137 E141 R151 E163 H164 H165 H166 H167 H168 H169

## 4.2.20 Score per residue for model 20

• Molecule 1: Anti-sigma-I factor RsgI2

Chain A:

90%                                                          10%

M1 I9 N10

• Molecule 2: Anti-sigma-I factor RsgI2

Chain B:

76%                              13%        •      11%

P11 K27 D32 I54 N69 L72 N78 S79 G80 R81 N82 N83 V84 S85 E86 S87 D88 I91 Q92 I95 K103 V107 E110 R119 K137 E141 E148 D149 L153 E157 K161 L162 E163 H164 H165 H166 H167 H168 H169

# 5 Refinement protocol and experimental data overview ⓘ

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

Of the 100 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 |
|---|---|---|
| CNS | refinement | |
| CNS | structure calculation | |

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 | 2128 |
| Number of shifts mapped to atoms | 2128 |
| Number of unparsed shifts | 0 |
| Number of shifts with mapping errors | 0 |
| Number of shifts with mapping warnings | 0 |
| Assignment completeness (well-defined parts) | 95% |

# 6 Model quality ⓘ

## 6.1 Standard geometry ⓘ

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

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

| Mol | Chain | Chirality | Planarity |
|-----|-------|-----------|-----------|
| 2 | B | 0.0±0.0 | 0.2±0.4 |
| All | All | 0 | 4 |

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

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

| Mol | Chain | Res | Type | Group | Models (Total) |
|-----|-------|-----|------|-------|----------------|
| 2 | B | 119 | ARG | Sidechain | 2 |
| 2 | B | 151 | ARG | Sidechain | 2 |

## 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 | 85 | 79 | 79 | 0±1 |
| 2 | B | 1090 | 1134 | 1134 | 6±1 |
| All | All | 23500 | 24260 | 24260 | 112 |

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 2.

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

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 2:B:149:ASP:O | 2:B:153:LEU:HG | 0.55 | 2.02 | 2 | 7 |
| 1:A:9:ILE:HG23 | 2:B:64:ILE:HG23 | 0.54 | 1.80 | 5 | 1 |
| 1:A:8:ASP:OD2 | 2:B:131:ARG:HB2 | 0.53 | 2.04 | 17 | 1 |
| 2:B:91:ILE:O | 2:B:95:ILE:HG12 | 0.53 | 2.03 | 19 | 20 |
| 2:B:69:ASN:ND2 | 2:B:107:VAL:HA | 0.52 | 2.20 | 19 | 8 |
| 2:B:20:LYS:HA | 2:B:20:LYS:NZ | 0.51 | 2.21 | 1 | 1 |
| 2:B:42:PRO:HA | 2:B:45:MET:SD | 0.51 | 2.46 | 2 | 8 |
| 2:B:121:LYS:O | 2:B:125:GLN:HG2 | 0.50 | 2.07 | 19 | 3 |
| 2:B:11:PRO:HA | 2:B:32:ASP:OD1 | 0.50 | 2.06 | 8 | 7 |
| 2:B:157:GLU:O | 2:B:161:LYS:HG2 | 0.50 | 2.07 | 10 | 3 |
| 2:B:48:TYR:CD1 | 2:B:101:VAL:HG22 | 0.49 | 2.42 | 3 | 1 |
| 2:B:137:LYS:O | 2:B:141:GLU:HG2 | 0.48 | 2.08 | 12 | 8 |
| 2:B:74:SER:HA | 2:B:112:ILE:O | 0.47 | 2.09 | 15 | 3 |
| 2:B:154:SER:O | 2:B:158:ILE:HG12 | 0.46 | 2.10 | 15 | 2 |
| 2:B:94:ILE:O | 2:B:98:LEU:HG | 0.45 | 2.11 | 3 | 14 |
| 2:B:36:ILE:HD11 | 2:B:63:TYR:CE1 | 0.45 | 2.47 | 16 | 1 |
| 2:B:17:ILE:HD12 | 2:B:21:GLU:HA | 0.45 | 1.89 | 13 | 1 |
| 2:B:23:VAL:HB | 2:B:42:PRO:O | 0.45 | 2.12 | 11 | 2 |
| 2:B:65:ASN:O | 2:B:69:ASN:HB3 | 0.44 | 2.12 | 19 | 1 |
| 2:B:11:PRO:HA | 2:B:32:ASP:OD2 | 0.44 | 2.12 | 16 | 2 |
| 1:A:6:ASP:OD1 | 2:B:128:SER:HB2 | 0.44 | 2.13 | 8 | 2 |
| 1:A:10:ASN:O | 2:B:154:SER:HA | 0.43 | 2.13 | 5 | 2 |
| 1:A:3:ALA:HB3 | 2:B:17:ILE:HD11 | 0.43 | 1.90 | 4 | 1 |
| 2:B:48:TYR:CD2 | 2:B:101:VAL:HG22 | 0.42 | 2.50 | 10 | 2 |
| 2:B:126:ASN:HB3 | 2:B:151:ARG:NH2 | 0.42 | 2.28 | 19 | 2 |
| 2:B:27:LYS:HZ2 | 2:B:29:LEU:HD21 | 0.42 | 1.75 | 1 | 1 |
| 2:B:89:LYS:N | 2:B:89:LYS:HD2 | 0.42 | 2.30 | 8 | 1 |
| 2:B:74:SER:OG | 2:B:130:GLY:HA2 | 0.41 | 2.15 | 4 | 1 |
| 2:B:99:LYS:O | 2:B:103:LYS:HG3 | 0.41 | 2.16 | 14 | 1 |
| 2:B:56:ASP:O | 2:B:59:LYS:HG2 | 0.41 | 2.16 | 12 | 1 |
| 2:B:157:GLU:O | 2:B:161:LYS:HG3 | 0.41 | 2.16 | 18 | 1 |
| 2:B:36:ILE:HD11 | 2:B:63:TYR:CE2 | 0.40 | 2.51 | 7 | 1 |
| 1:A:10:ASN:OXT | 2:B:154:SER:HA | 0.40 | 2.16 | 11 | 1 |
| 1:A:9:ILE:HD11 | 2:B:54:ILE:HG23 | 0.40 | 1.93 | 20 | 1 |

## 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 | 8/10 (80%) | 7±0 (88±3%) | 1±0 (11±4%) | 0±0 (1±3%) | 29 | 74 |
| 2 | B | 141/159 (89%) | 139±1 (98±1%) | 2±1 (2±1%) | 0±0 (0±0%) | 54 | 85 |
| All | All | 2980/3380 (88%) | 2913 (98%) | 64 (2%) | 3 (0%) | 54 | 85 |

All 2 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

| Mol | Chain | Res | Type | Models (Total) |
|-----|-------|-----|------|----------------|
| 2 | B | 44 | ASP | 2 |
| 1 | A | 2 | TYR | 1 |

### 6.3.2　Protein sidechains　ⓘ

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 | 9/9 (100%) | 9±1 (96±6%) | 0±1 (4±6%) | 32 | 81 |
| 2 | B | 122/138 (88%) | 115±1 (94±1%) | 7±1 (6±1%) | 22 | 71 |
| All | All | 2620/2940 (89%) | 2463 (94%) | 157 (6%) | 23 | 72 |

All 33 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) |
|-----|-------|-----|------|----------------|
| 2 | B | 78 | ASN | 19 |
| 2 | B | 69 | ASN | 18 |
| 2 | B | 72 | LEU | 18 |
| 2 | B | 120 | GLN | 12 |
| 2 | B | 27 | LYS | 12 |
| 2 | B | 20 | LYS | 8 |
| 2 | B | 128 | SER | 8 |
| 2 | B | 119 | ARG | 8 |
| 2 | B | 126 | ASN | 7 |
| 1 | A | 6 | ASP | 6 |
| 2 | B | 110 | GLU | 4 |
| 2 | B | 89 | LYS | 4 |

*Continued on next page...*

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

*Continued from previous page...*

| Mol | Chain | Res | Type | Models (Total) |
|-----|-------|-----|------|----------------|
| 2 | B | 103 | LYS | 3 |
| 2 | B | 116 | GLU | 3 |
| 2 | B | 117 | GLU | 2 |
| 2 | B | 137 | LYS | 2 |
| 2 | B | 28 | PRO | 2 |
| 2 | B | 149 | ASP | 2 |
| 2 | B | 59 | LYS | 2 |
| 2 | B | 157 | GLU | 2 |
| 2 | B | 148 | GLU | 2 |
| 1 | A | 1 | MET | 2 |
| 2 | B | 22 | LYS | 1 |
| 2 | B | 42 | PRO | 1 |
| 2 | B | 151 | ARG | 1 |
| 2 | B | 104 | ASP | 1 |
| 2 | B | 161 | LYS | 1 |
| 2 | B | 125 | GLN | 1 |
| 2 | B | 51 | LEU | 1 |
| 2 | B | 18 | ASP | 1 |
| 2 | B | 54 | ILE | 1 |
| 2 | B | 53 | LYS | 1 |
| 2 | B | 92 | GLN | 1 |

### 6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 6.6 Ligand geometry ⓘ

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 95% for the well-defined parts and 93% 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 | 2128 |
| Number of shifts mapped to atoms | 2128 |
| Number of unparsed shifts | 0 |
| Number of shifts with mapping errors | 0 |
| Number of shifts with mapping warnings | 0 |
| Number of shift outliers (ShiftChecker) | 2 |

### 7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

| Nucleus | # values | Correction ± precision, *ppm* | Suggested action |
|---|---|---|---|
| $^{13}C_\alpha$ | 168 | -0.31 ± 0.23 | None needed (< 0.5 ppm) |
| $^{13}C_\beta$ | 159 | 0.09 ± 0.12 | None needed (< 0.5 ppm) |
| $^{13}C'$ | 168 | 0.38 ± 0.12 | None needed (< 0.5 ppm) |
| $^{15}N$ | 161 | 0.27 ± 0.33 | 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 95%, i.e. 1965 atoms were assigned a chemical shift out of a possible 2066. 0 out of 25 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

| | Total | $^1H$ | $^{13}C$ | $^{15}N$ |
|---|---|---|---|---|
| Backbone | 748/755 (99%) | 302/305 (99%) | 302/304 (99%) | 144/146 (99%) |
| Sidechain | 1149/1237 (93%) | 785/805 (98%) | 348/393 (89%) | 16/39 (41%) |

*Continued on next page...*

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

*Continued from previous page...*

|  | **Total** | **$^1$H** | **$^{13}$C** | **$^{15}$N** |
|---|---|---|---|---|
| Aromatic | 68/74 (92%) | 34/34 (100%) | 34/40 (85%) | 0/0 (—%) |
| Overall | 1965/2066 (95%) | 1121/1144 (98%) | 684/737 (93%) | 160/185 (86%) |

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 93%, i.e. 2128 atoms were assigned a chemical shift out of a possible 2287. 0 out of 26 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

|  | **Total** | **$^1$H** | **$^{13}$C** | **$^{15}$N** |
|---|---|---|---|---|
| Backbone | 834/841 (99%) | 337/340 (99%) | 336/338 (99%) | 161/163 (99%) |
| Sidechain | 1224/1324 (92%) | 835/859 (97%) | 370/421 (88%) | 19/44 (43%) |
| Aromatic | 70/122 (57%) | 35/58 (60%) | 35/52 (67%) | 0/12 (0%) |
| Overall | 2128/2287 (93%) | 1207/1257 (96%) | 741/811 (91%) | 180/219 (82%) |

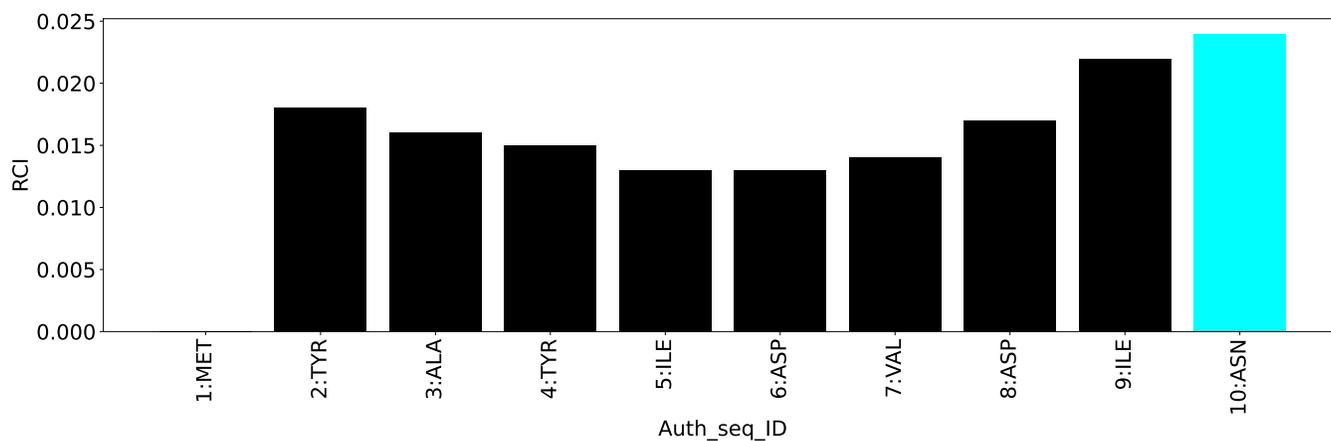### 7.1.4　Statistically unusual chemical shifts ⓘ

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 | B | 129 | MET | H | 11.13 | 5.39 – 11.10 | 5.1 |
| 1 | A | 6 | ASP | HA | 6.13 | 3.04 – 6.12 | 5.0 |

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

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:

Random coil index (RCI) for chain B: