# Full wwPDB NMR Structure Validation Report (i)

Oct 23, 2021 – 05:33 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 1AZH |
| Title | : | THREE-DIMENSIONAL STRUCTURES OF THREE ENGINEERED CELLULOSE-BINDING DOMAINS OF CELLOBIOHYDROLASE I FROM TRICHODERMA REESEI, NMR, 14 STRUCTURES |
| Authors | : | Mattinen, M.-L. |
| Deposited on | : | 1997-11-18 |

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

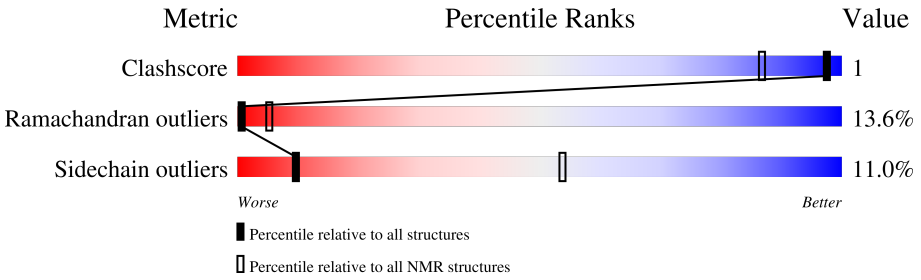
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| RCI | : | v_1n_11_5_13_A (Berjanski et al., 2005) |
| PANAV | : | Wang et al. (2010) |
| ShiftChecker | : | 2.23.2 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.23.2 |

# 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 was not calculated.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.




| Metric | 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 | 36 | 72% 17% 11% |

# 2 Ensemble composition and analysis (i)

This entry contains 14 models. Model 14 is the overall representative, medoid model (most similar to other models).

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:5-A:36 (32) | 0.70 | 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 2 clusters and 4 single-model clusters were found.

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

# 3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 484 atoms, of which 231 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called CELLOBIOHYDROLASE I.

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | S | |
| 1 | A | 36 | 484 | 153 | 231 | 43 | 53 | 4 | 0 |

There is a discrepancy between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 5 | ALA | TYR | engineered mutation | UNP P62694 |

# 4 Residue-property plots (i)

## 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: CELLOBIOHYDROLASE I

Chain A: 72% 17% 11%

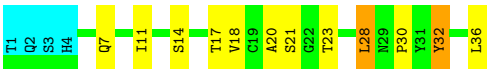
## 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

### 4.2.1 Score per residue for model 1

• Molecule 1: CELLOBIOHYDROLASE I

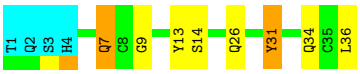
Chain A: 56% 28% 6% 11%




### 4.2.2 Score per residue for model 2

• Molecule 1: CELLOBIOHYDROLASE I

Chain A: 67% 17% 6% 11%



### 4.2.3 Score per residue for model 3

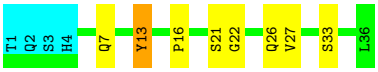
- Molecule 1: CELLOBIOHYDROLASE I

Chain A: 61% 25% • 11%

### 4.2.4 Score per residue for model 4

- Molecule 1: CELLOBIOHYDROLASE I

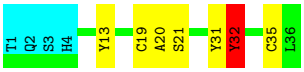
Chain A: 67% 19% • 11%




### 4.2.5 Score per residue for model 5

- Molecule 1: CELLOBIOHYDROLASE I

Chain A: 69% 17% • 11%




### 4.2.6 Score per residue for model 6

- Molecule 1: CELLOBIOHYDROLASE I

Chain A: 67% 19% • 11%

### 4.2.7 Score per residue for model 7

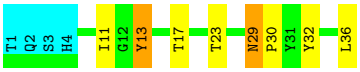
- Molecule 1: CELLOBIOHYDROLASE I

Chain A: 64% 22% • 11%

### 4.2.8 Score per residue for model 8

- Molecule 1: CELLOBIOHYDROLASE I

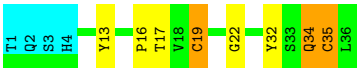
Chain A: 67% 17% 6% 11%




### 4.2.9 Score per residue for model 9

- Molecule 1: CELLOBIOHYDROLASE I

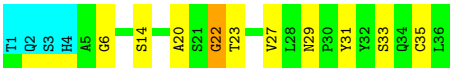
Chain A: 67% 14% 8% 11%




### 4.2.10 Score per residue for model 10

- Molecule 1: CELLOBIOHYDROLASE I

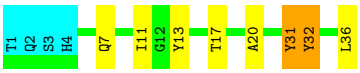
Chain A: 61% 25% • 11%




### 4.2.11 Score per residue for model 11

- Molecule 1: CELLOBIOHYDROLASE I

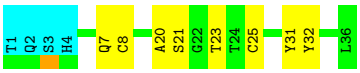
Chain A: 67% 17% 6% 11%




### 4.2.12 Score per residue for model 12

- Molecule 1: CELLOBIOHYDROLASE I

Chain A: 67% 22% 11%



### 4.2.13 Score per residue for model 13

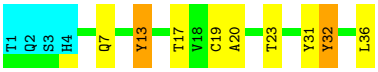
- Molecule 1: CELLOBIOHYDROLASE I

Chain A: 75% 8% 6% 11%

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

- Molecule 1: CELLOBIOHYDROLASE I

Chain A: 64% 19% 6% 11%



# 5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *distance geometry*.

Of the 60 calculated structures, 14 were deposited, based on the following criterion: *LEAST RESTRAINT VIOLATION*.

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

| Software name | Classification | Version |
|---|---|---|
| SA/MD | refinement | |
| BIOSYM TECHNOLOGIES | structure solution | TECHNOLOGIES |
| Felix | structure solution | |
| Insight II | structure solution | |

No chemical shift data was provided.

# 6 Model quality (i)

## 6.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

| Mol | Chain | Bond lengths | | Bond angles | |
|---|---|---|---|---|---|
| | | RMSZ | #Z>5 | RMSZ | #Z>5 |
| 1 | A | 0.92±0.02 | 0±0/225 ( 0.0± 0.0%) | 1.20±0.10 | 0±1/305 ( 0.2± 0.2%) |
| All | All | 0.92 | 0/3150 ( 0.0%) | 1.20 | 7/4270 ( 0.2%) |

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

| Mol | Chain | Chirality | Planarity |
|---|---|---|---|
| 1 | A | 0.0±0.0 | 2.6±1.0 |
| All | All | 0 | 36 |

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 |
| 1 | A | 32 | TYR | CB-CG-CD1 | -8.82 | 115.71 | 121.00 | 11 | 2 |
| 1 | A | 32 | TYR | CB-CG-CD2 | -8.05 | 116.17 | 121.00 | 13 | 4 |
| 1 | A | 31 | TYR | CB-CG-CD2 | -5.28 | 117.83 | 121.00 | 2 | 1 |

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) |
|---|---|---|---|---|---|
| 1 | A | 32 | TYR | Sidechain,Peptide | 9 |
| 1 | A | 31 | TYR | Peptide,Sidechain | 8 |
| 1 | A | 17 | THR | Peptide | 5 |
| 1 | A | 13 | TYR | Sidechain | 5 |
| 1 | A | 29 | ASN | Peptide | 2 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type | Group | Models (Total) |
|---|---|---|---|---|---|
| 1 | A | 7 | GLN | Peptide | 1 |
| 1 | A | 18 | VAL | Peptide | 1 |
| 1 | A | 28 | LEU | Peptide | 1 |
| 1 | A | 33 | SER | Peptide | 1 |
| 1 | A | 19 | CYS | Peptide | 1 |
| 1 | A | 22 | GLY | Peptide | 1 |
| 1 | A | 5 | ALA | Peptide | 1 |

## 6.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

| Mol | Chain | Non-H | H(model) | H(added) | Clashes |
|---|---|---|---|---|---|
| 1 | A | 221 | 202 | 202 | 0±1 |
| All | All | 3094 | 2828 | 2828 | 3 |

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 1.

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

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 1:A:28:LEU:HD13 | 1:A:32:TYR:CD2 | 0.56 | 2.36 | 1 | 1 |
| 1:A:28:LEU:N | 1:A:28:LEU:HD12 | 0.42 | 2.29 | 1 | 1 |
| 1:A:17:THR:HG23 | 1:A:17:THR:O | 0.41 | 2.16 | 9 | 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 | 31/36 (86%) | 17±2 (54±8%) | 10±2 (33±8%) | 4±1 (14±4%) | 1 5 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Analysed | Favoured | Allowed | Outliers | Percentiles |
|---|---|---|---|---|---|---|
| All | All | 434/504 (86%) | 233 (54%) | 142 (33%) | 59 (14%) | 1 5 |

All 20 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 | 20 | ALA | 8 |
| 1 | A | 7 | GLN | 7 |
| 1 | A | 11 | ILE | 5 |
| 1 | A | 23 | THR | 5 |
| 1 | A | 35 | CYS | 5 |
| 1 | A | 22 | GLY | 4 |
| 1 | A | 14 | SER | 3 |
| 1 | A | 30 | PRO | 3 |
| 1 | A | 17 | THR | 3 |
| 1 | A | 21 | SER | 2 |
| 1 | A | 9 | GLY | 2 |
| 1 | A | 8 | CYS | 2 |
| 1 | A | 16 | PRO | 2 |
| 1 | A | 19 | CYS | 2 |
| 1 | A | 29 | ASN | 1 |
| 1 | A | 34 | GLN | 1 |
| 1 | A | 6 | GLY | 1 |
| 1 | A | 33 | SER | 1 |
| 1 | A | 31 | TYR | 1 |
| 1 | A | 5 | ALA | 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 | 24/28 (86%) | 21±1 (89±5%) | 3±1 (11±5%) | 10 54 |
| All | All | 336/392 (86%) | 299 (89%) | 37 (11%) | 10 54 |

All 13 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 | 13 | TYR | 10 |
| 1 | A | 36 | LEU | 7 |
| 1 | A | 32 | TYR | 5 |
| 1 | A | 26 | GLN | 2 |
| 1 | A | 34 | GLN | 2 |
| 1 | A | 25 | CYS | 2 |
| 1 | A | 21 | SER | 2 |
| 1 | A | 27 | VAL | 2 |
| 1 | A | 23 | THR | 1 |
| 1 | A | 28 | LEU | 1 |
| 1 | A | 7 | GLN | 1 |
| 1 | A | 19 | CYS | 1 |
| 1 | A | 35 | CYS | 1 |

#### 6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

### 6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

### 6.6 Ligand geometry (i)

There are no ligands in this entry.

### 6.7 Other polymers (i)

There are no such molecules in this entry.

### 6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

# 7 Chemical shift validation (i)

No chemical shift data were provided