# wwPDB NMR Structure Validation Summary Report (i)

Feb 15, 2024 – 10:01 PM EST

| | | |
|---|---|---|
| PDB ID | : | 1K09 |
| Title | : | Solution structure of BetaCore, A Designed Water Soluble Four-Stranded Antiparallel b-sheet Protein |
| Authors | : | Carulla, N.; Woodward, C.; Barany, G. |
| Deposited on | : | 2001-09-18 |

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

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

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

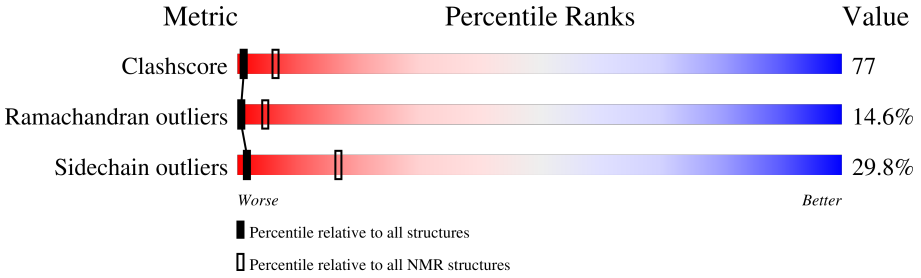
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| buster-report | : | 1.1.7 (2018) |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| wwPDB-RCI | : | v_1n_11_5_13_A (Berjanski et al., 2005) |
| PANAV | : | Wang et al. (2010) |
| wwPDB-ShiftChecker | : | v1.2 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.36 |

# 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 | 26 | 58% 12% 31% |
| 2 | B | 26 | 50% 23% 27% |

# 2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 20 is the overall representative, medoid model (most similar to other models). The authors have identified model 8 as representative, based on the following criterion: *closest to the average*.

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:16, A:18-A:23, B:4-B:16, B:18-B:23 (37) | 2.42 | 20 |

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. No single-model clusters were found.

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

# 3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 824 atoms, of which 407 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Core Module I.

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | S | |
| 1 | A | 26 | 397 | 131 | 198 | 33 | 33 | 2 | 1 |

There are 5 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 1 | MPT | CYS | engineered mutation | UNP P00974 |
| A | 4 | LYS | ARG | engineered mutation | UNP P00974 |
| A | 14 | ASP | ALA | engineered mutation | UNP P00974 |
| A | 17 | ABA | CYS | engineered mutation | UNP P00974 |
| A | 26 | NH2 | - | amidation | UNP P00974 |

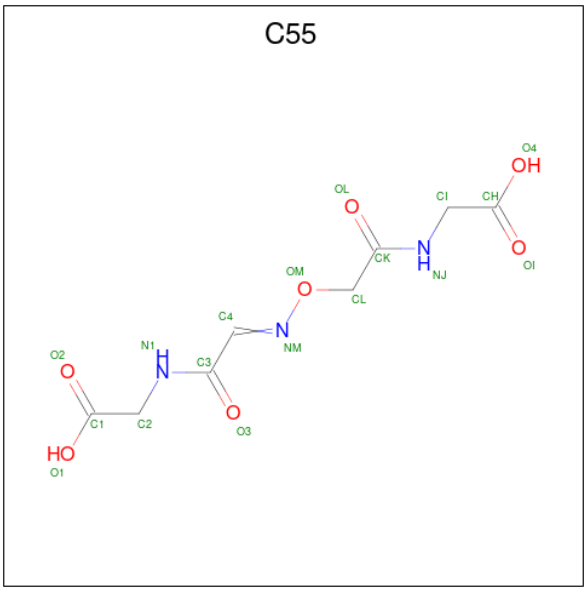
- Molecule 2 is a protein called Core Module II.

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | S | |
| 2 | B | 26 | 402 | 131 | 200 | 36 | 33 | 2 | 1 |

There are 5 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| B | 1 | MPT | CYS | engineered mutation | UNP P00974 |
| B | 14 | ASP | ALA | engineered mutation | UNP P00974 |
| B | 16 | LYS | LEU | engineered mutation | UNP P00974 |
| B | 17 | ABA | CYS | engineered mutation | UNP P00974 |
| B | 26 | NH2 | - | amidation | UNP P00974 |

- Molecule 3 is (7E)-4,9-dioxo-6-oxa-3,7,10-triazadodec-7-ene-1,12-dioic acid (three-letter code: C55) (formula: $C_8H_{11}N_3O_7$).




| Mol | Chain | Residues | Atoms | | | | |
|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O |
| 3 | A | 1 | 25 | 8 | 9 | 3 | 5 |

# 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: Core Module I

Chain A: 58% 12% 31%

• Molecule 2: Core Module II

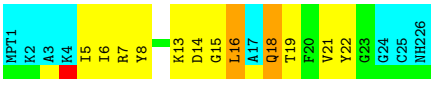
Chain B: 50% 23% 27%

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

The representative model is number 20. Colouring as in section 4.1 above.

• Molecule 1: Core Module I

Chain A: 23% 38% 8% 31%




• Molecule 2: Core Module II

Chain B: 19% 38% 12% • 27%

# 5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *The submitted conformers are those with no constraint violations greater than 0.5 angstroms for NOEs and 5 degrees for dihedrals. They are also the ones with best covalent geometry.*.

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

| Software name | Classification | Version |
|---|---|---|
| X-PLOR | structure solution | 3.851 |
| X-PLOR | refinement | 3.851 |

No chemical shift data was provided.

# 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: C55, ABA, MPT, NH2

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 |
|---|---|---|---|
| 1 | A | 0.0±0.0 | 1.0±0.0 |
| 2 | B | 0.0±0.0 | 2.0±0.0 |
| All | All | 0 | 60 |

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) |
|---|---|---|---|---|---|
| 1 | A | 7 | ARG | Sidechain | 20 |
| 2 | B | 4 | ARG | Sidechain | 20 |
| 2 | B | 7 | ARG | Sidechain | 20 |

## 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 | 154 | 149 | 149 | 29±12 |
| 2 | B | 166 | 162 | 162 | 30±11 |
| 3 | A | 16 | 9 | 0 | 1±2 |
| All | All | 6720 | 6400 | 6215 | 991 |

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 77.

5 of 716 unique clashes are listed below, sorted by their clash magnitude.

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 1:A:21:VAL:HG13 | 2:B:6:ILE:HG23 | 1.01 | 1.25 | 7 | 1 |
| 2:B:6:ILE:HD12 | 2:B:21:VAL:HG21 | 1.00 | 1.34 | 5 | 1 |
| 1:A:16:LEU:HD11 | 2:B:9:PHE:CZ | 0.97 | 1.94 | 18 | 1 |
| 1:A:5:ILE:HG22 | 1:A:22:TYR:O | 0.87 | 1.69 | 1 | 1 |
| 1:A:16:LEU:HD23 | 1:A:16:LEU:O | 0.87 | 1.69 | 15 | 2 |

## 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 | 18/26 (69%) | 11±2 (62±10%) | 4±2 (24±10%) | 2±2 (13±9%) | 1 | 5 |
| 2 | B | 19/26 (73%) | 10±3 (53±18%) | 6±3 (31±15%) | 3±2 (16±9%) | 0 | 4 |
| All | All | 740/1040 (71%) | 426 (58%) | 206 (28%) | 108 (15%) | 0 | 4 |

5 of 25 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 | 13 | LYS | 14 |
| 2 | B | 18 | GLN | 9 |
| 2 | B | 11 | ASN | 9 |
| 2 | B | 13 | LYS | 9 |
| 1 | A | 5 | ILE | 8 |

### 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 | 15/18 (83%) | 11±1 (73±9%) | 4±1 (27±9%) | 2 22 |
| 2 | B | 16/18 (89%) | 11±2 (67±12%) | 5±2 (33±12%) | 1 12 |
| All | All | 620/720 (86%) | 435 (70%) | 185 (30%) | 1 17 |

5 of 28 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 | 16 | LYS | 15 |
| 2 | B | 7 | ARG | 12 |
| 1 | A | 7 | ARG | 12 |
| 2 | B | 13 | LYS | 12 |
| 1 | A | 13 | LYS | 11 |

### 6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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 | ABA | B | 17 | 2 | 4,5,6 | 0.43±0.02 | 0±0 (0±0%) |
| 1 | ABA | A | 17 | 1 | 4,5,6 | 0.43±0.03 | 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 | ABA | B | 17 | 2 | 1,5,7 | 1.67±0.10 | 0±0 (0±0%) |
| 1 | ABA | A | 17 | 1 | 1,5,7 | 1.71±0.07 | 0±0 (0±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 | ABA | B | 17 | 2 | - | 0±0,3,4,6 | - |
| 1 | ABA | A | 17 | 1 | - | 0±0,3,4,6 | - |

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

## 6.6 Ligand geometry (i)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

| Mol | Type | Chain | Res | Link | Bond lengths | | |
|---|---|---|---|---|---|---|---|
| | | | | | Counts | RMSZ | #Z>2 |
| 3 | C55 | A | 27 | 1,2 | 14,15,17 | 1.09±0.01 | 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 |
| 3 | C55 | A | 27 | 1,2 | 13,16,20 | 2.85±0.01 | 1±0 (7±0%) |

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

| Mol | Type | Chain | Res | Link | Chirals | Torsions | Rings |
|---|---|---|---|---|---|---|---|
| 3 | C55 | A | 27 | 1,2 | - | 0±0,13,15,17 | - |

There are no bond-length outliers.

All unique angle outliers are listed below.

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) | Models | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Worst | Total |
| 3 | A | 27 | C55 | OM-NM-C4 | 9.49 | 120.09 | 110.25 | 17 | 20 |

There are no chirality outliers.

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

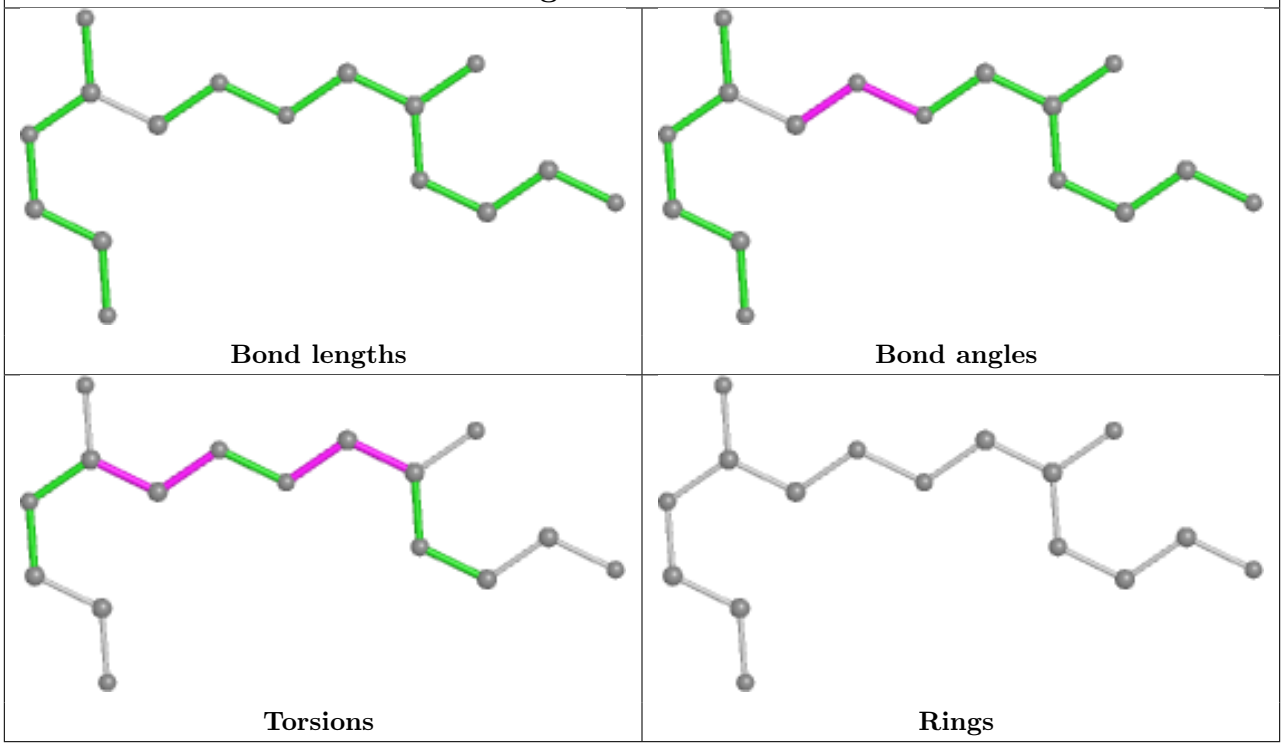
| Mol | Chain | Res | Type | Atoms | Models (Total) |
|---|---|---|---|---|---|
| 3 | A | 27 | C55 | C3-C4-NM-OM | 6 |
| 3 | A | 27 | C55 | C4-NM-OM-CL | 2 |

There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and

any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

**Ligand C55 A 27**




## 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