# wwPDB NMR Structure Validation Summary Report (i)

Mar 5, 2022 – 10:01 AM EST

| | | |
|---|---|---|
| PDB ID | : | 2K0S |
| Title | : | Solution NMR structure of protein BC066483 |
| Authors | : | Cornilescu, C.C.; Loushin-Newman, C.L.; Vinarov, D.A.; Markley, J.L.; Center for Eukaryotic Structural Genomics (CESG) |
| Deposited on | : | 2008-02-14 |

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

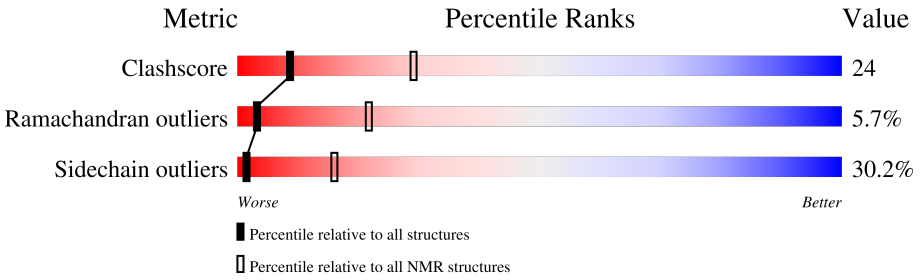
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| RCI | : | v_1n_11_5_13_A (Berjanski et al., 2005) |
| PANAV | : | Wang et al. (2010) |
| ShiftChecker | : | 2.27 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.27 |

# 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 | 130 | 41% 45% 7% 8% |

# 2 Ensemble composition and analysis (i)

This entry contains 5 models. Model 1 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:11-A:130 (120) | 1.00 | 1 |

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 1 clusters and 2 single-model clusters were found.

| Cluster number | Models |
|---|---|
| 1 | 1, 2, 4 |
| Single-model clusters | 3; 5 |

# 3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1927 atoms, of which 833 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ADP-ribosylation factor-like protein 2-binding protein.

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | S | |
| 1 | A | 130 | 1927 | 695 | 833 | 176 | 217 | 6 | 0 |

There are 9 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 1 | GLY | - | expression tag | UNP Q7SYL1 |
| A | 2 | HIS | - | expression tag | UNP Q7SYL1 |
| A | 3 | HIS | - | expression tag | UNP Q7SYL1 |
| A | 4 | HIS | - | expression tag | UNP Q7SYL1 |
| A | 5 | HIS | - | expression tag | UNP Q7SYL1 |
| A | 6 | HIS | - | expression tag | UNP Q7SYL1 |
| A | 7 | HIS | - | expression tag | UNP Q7SYL1 |
| A | 8 | LEU | - | expression tag | UNP Q7SYL1 |
| A | 9 | GLU | - | expression tag | UNP Q7SYL1 |

# 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: ADP-ribosylation factor-like protein 2-binding protein

Chain A: 41% 45% 7% 8%

G1 H2 H3 H4 H5 H6 H7 L8 E9 D10 A11 D12 A13 E14 N21 I22 E23 D24 I25 I26 M27 E28 D29 E30 F31 Q32 H33 L34 S37 F38 M39 F46 S49 E50 E51 N52 K53 L54 S55 Y56 T57 P58 L67 E68 K69 H70 L71 Q74 L75 R78 I79 P80 G81

F82 N83 M84 D85 A86 F87 T88 H89 S90 L91 K92 K95 V98 S99 G100 D101 I102 L103 D104 M105 L106 L107 T108 F109 T110 D111 F112 M113 K116 E117 M118 F119 T120 A124 E125 K126 R129 G130

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

The representative model is number 1. Colouring as in section 4.1 above.

• Molecule 1: ADP-ribosylation factor-like protein 2-binding protein

Chain A: 48% 34% 11% 8%

G1 H2 H3 H4 H5 H6 H7 L8 E9 D10 A11 N21 I22 E23 D24 I25 I26 M27 E28 D29 E30 F31 Q32 H33 L34 M39 Y42 D48 E51 N52 K53 L54 S55 Y56 T57 P58 Y63 I66 L67 E68 K69 H70 L71 E72 Q73 Q74 L75 V76 E77 R78 I79 P80 G81

F82 N83 M84 D85 A86 F87 T88 K95 V98 S99 G100 D101 D104 M105 L106 L107 T108 F109 T110 D111 F112 M113 K116 E117 M118 F119 T120 A124 R129 G130

# 5 Refinement protocol and experimental data overview (i)

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

Of the 20 calculated structures, 5 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 |
|---|---|---|
| CYANA | structure solution | |
| CYANA | refinement | |

No chemical shift data was provided.

# 6 Model quality (i)

## 6.1 Standard geometry (i)

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

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

## 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 | 1005 | 777 | 939 | 46±5 |
| All | All | 5025 | 3885 | 4695 | 230 |

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 24.

5 of 161 unique clashes are listed below, sorted by their clash magnitude.

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 1:A:25:ILE:HD13 | 1:A:86:ALA:HB3 | 1.07 | 1.20 | 2 | 3 |
| 1:A:23:GLU:O | 1:A:26:ILE:HD12 | 0.86 | 1.70 | 3 | 3 |
| 1:A:21:ASN:O | 1:A:25:ILE:HG23 | 0.84 | 1.71 | 5 | 4 |
| 1:A:29:ASP:HA | 1:A:86:ALA:HB2 | 0.80 | 1.55 | 2 | 3 |
| 1:A:82:PHE:O | 1:A:84:MET:N | 0.77 | 2.18 | 2 | 4 |

## 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 | 119/130 (92%) | 90±4 (75±3%) | 22±3 (19±3%) | 7±2 (6±2%) | 3 22 |
| All | All | 595/650 (92%) | 449 (75%) | 112 (19%) | 34 (6%) | 3 22 |

5 of 18 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 | 83 | ASN | 4 |
| 1 | A | 111 | ASP | 4 |
| 1 | A | 98 | VAL | 3 |
| 1 | A | 53 | LYS | 2 |
| 1 | A | 108 | THR | 2 |

### 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 | 110/119 (92%) | 77±2 (70±2%) | 33±2 (30±2%) | 1 16 |
| All | All | 550/595 (92%) | 384 (70%) | 166 (30%) | 1 16 |

5 of 77 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 | 95 | LYS | 5 |
| 1 | A | 112 | PHE | 5 |
| 1 | A | 116 | LYS | 5 |
| 1 | A | 34 | LEU | 4 |
| 1 | A | 69 | LYS | 4 |

### 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