# wwPDB NMR Structure Validation Summary Report (i)

Feb 7, 2022 – 06:03 PM EST

| | | |
|---|---|---|
| PDB ID | : | 1AB2 |
| Title | : | THREE-DIMENSIONAL SOLUTION STRUCTURE OF THE SRC HOMOLOGY 2 DOMAIN OF C-ABL |
| Authors | : | Overduin, M.; Rios, C.B.; Mayer, B.J.; Baltimore, D.; Cowburn, D. |
| Deposited on | : | 1993-07-19 |

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:

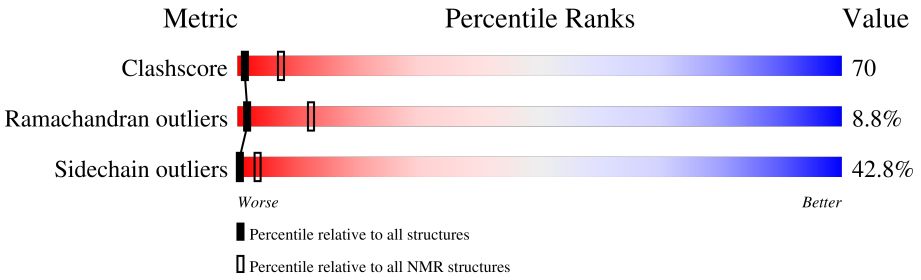
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| RCI | : | v_1n_11_5_13_A (Berjanski et al., 2005) |
| PANAV | : | Wang et al. (2010) |
| ShiftChecker | : | 2.26 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.26 |

# 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 | 109 | 12% 57% 17% • 13% |

# 2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 7 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:6-A:100 (95) | 0.83 | 7 |

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 and 1 single-model cluster was found.

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

# 3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1684 atoms, of which 824 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called C-ABL TYROSINE KINASE SH2 DOMAIN.

| Mol | Chain | Residues | Atoms | | | | | Trace |
|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | |
| 1 | A | 109 | 1684 | 533 | 824 | 162 | 165 | 0 |

# 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: C-ABL TYROSINE KINASE SH2 DOMAIN

Chain A: 12% 57% 17% • 13%

G1 S2 G3 N4 S5 L6 E7 K8 H9 S10 W11 Y12 H13 V16 S17 R18 N19 A20 A21 E22 Y23 L24 L25 S26 S27 G28 I29 S32 F33 L34 V35 R36 E37 S38 E39 S40 S41 Q44 R45 S46 I47 S48 L49 R50 Y51 E52 G53 R54 V55 Y56 H57 Y58 R59 I60 N61 T62 A63

S64 D65 G66 K67 L68 Y69 V70 S71 S72 E73 S74 R75 F76 N77 T78 L79 A80 E81 L82 V83 H84 H85 H86 S87 T88 V89 A90 D91 G92 L93 I94 T95 T96 L97 H98 Y99 P100 A101 P102 K103 R104 G105 I106 H107 R108 D109

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

The representative model is number 7. Colouring as in section 4.1 above.

• Molecule 1: C-ABL TYROSINE KINASE SH2 DOMAIN

Chain A: 16% 50% 20% • 13%

G1 S2 G3 N4 S5 L6 E7 K8 H9 S10 W11 Y12 H13 V16 S17 R18 N19 A20 A21 E22 Y23 L24 L25 S26 S27 G28 I29 N30 G31 S32 F33 L34 V35 R36 E37 S38 E39 S40 S41 Q44 R45 S46 I47 S48 L49 R50 Y51 E52 V55 Y56 H57 Y58 R59 I60 N61 T62 A63

S64 D65 G66 K67 L68 Y69 V70 S71 S72 E73 S74 R75 F76 N77 T78 L79 A80 E81 L82 V83 H84 H85 H86 S87 T88 V89 A90 D91 G92 L93 I94 T95 T96 L97 H98 Y99 P100 A101 P102 K103 R104 G105 I106 H107 R108 D109

# 5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 20 were deposited, based on the following criterion: *?*.

The authors did not provide any information on software used for structure solution, optimization or 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 | 758 | 724 | 724 | 104±14 |
| All | All | 15160 | 14480 | 14480 | 2082 |

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 70.

5 of 1021 unique clashes are listed below, sorted by their clash magnitude.

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 1:A:12:TYR:CE2 | 1:A:79:LEU:HD11 | 1.11 | 1.79 | 18 | 6 |
| 1:A:99:TYR:CD1 | 1:A:99:TYR:O | 1.10 | 2.04 | 17 | 2 |
| 1:A:35:VAL:HG21 | 1:A:79:LEU:HD21 | 1.06 | 1.22 | 11 | 1 |
| 1:A:70:VAL:CG2 | 1:A:82:LEU:HD21 | 1.06 | 1.81 | 6 | 4 |
| 1:A:60:ILE:HG23 | 1:A:70:VAL:HG22 | 1.05 | 1.24 | 5 | 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 | 95/109 (87%) | 63±4 (66±4%) | 24±4 (25±4%) | 8±2 (9±2%) | 1 12 |
| All | All | 1900/2180 (87%) | 1262 (66%) | 470 (25%) | 168 (9%) | 1 12 |

5 of 27 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 | 8 | LYS | 20 |
| 1 | A | 94 | ILE | 18 |
| 1 | A | 92 | GLY | 14 |
| 1 | A | 27 | SER | 12 |
| 1 | A | 97 | LEU | 11 |

### 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 | 83/93 (89%) | 47±3 (57±4%) | 36±3 (43±4%) | 0 3 |
| All | All | 1660/1860 (89%) | 949 (57%) | 711 (43%) | 0 3 |

5 of 71 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 | 47 | ILE | 20 |
| 1 | A | 84 | HIS | 20 |
| 1 | A | 58 | TYR | 19 |
| 1 | A | 22 | GLU | 18 |
| 1 | A | 33 | PHE | 18 |

### 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