# Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2024 – 06:34 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 2N41 |
| BMRB ID | : | 25664 |
| Title | : | Solution NMR Structure of DE NOVO DESIGNED PROTEIN Top7NNSTYCC, Northeast Structural Genomics Consortium (NESG) Target OR34 |
| Authors | : | Liu, G.; Chan, K.; Basanta, B.; Xiao, R.; Janjua, H.; Kogan, S.; Maglaqui, M.; Ciccosanti, C.; Acton, T.B.; Kornhaber, G.; Everett, J.K.; Baker, D.; Montelione, G.T.; Northeast Structural Genomics Consortium (NESG) |
| Deposited on | : | 2015-06-16 |

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 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:

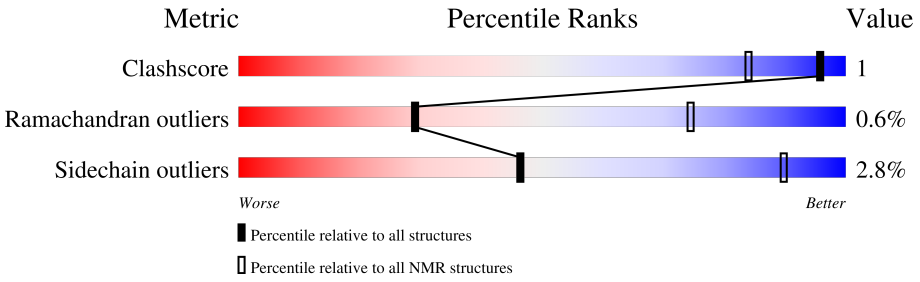
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| wwPDB-RCI | : | v_1n_11_5_13_A (Berjanski et al., 2005) |
| PANAV | : | Wang et al. (2010) |
| wwPDB-ShiftChecker | : | v1.2 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.36.2 |

# 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 | 106 | 75% · 25% |

# 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 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:11-A:19, A:27-A:83, A:87-A:100 (80) | 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 3 clusters and 6 single-model clusters were found.

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

# 3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1689 atoms, of which 836 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called OR34.

| Mol | Chain | Residues | Atoms | | | | | | Trace |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | S | |
| 1 | A | 106 | 1689 | 527 | 836 | 153 | 170 | 3 | 0 |

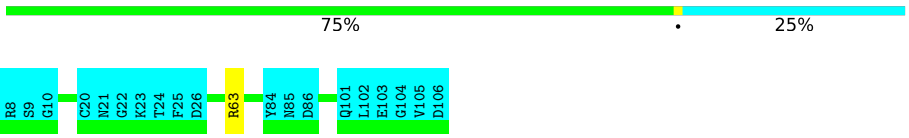
# 4 Residue-property plots (i)

## 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: OR34

Chain A:



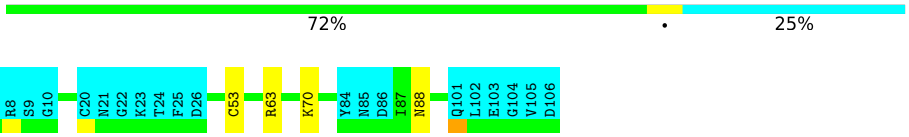

## 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

### 4.2.1 Score per residue for model 1

• Molecule 1: OR34

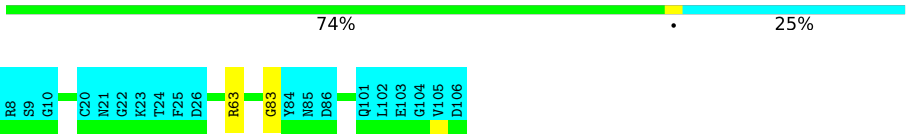
Chain A:




### 4.2.2 Score per residue for model 2

• Molecule 1: OR34

Chain A:



### 4.2.3 Score per residue for model 3

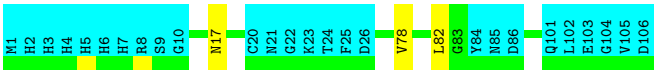
- Molecule 1: OR34

Chain A: 67% 8% 25%

### 4.2.4 Score per residue for model 4

- Molecule 1: OR34

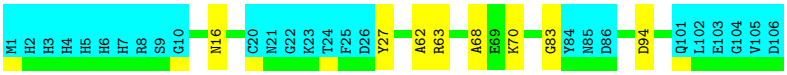
Chain A: 73% • 25%




### 4.2.5 Score per residue for model 5

- Molecule 1: OR34

Chain A: 68% 8% 25%




### 4.2.6 Score per residue for model 6

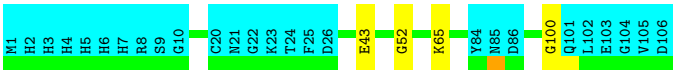
- Molecule 1: OR34

Chain A: 73% • 25%

### 4.2.7 Score per residue for model 7 (medoid)

- Molecule 1: OR34

Chain A: 72% • 25%



### 4.2.8 Score per residue for model 8

- Molecule 1: OR34

Chain A: 70% 5% • 25%

M1 H2 H3 H4 H5 H6 H7 R8 S9 G10 N18 T19 C20 N21 G22 K23 T24 F25 D26 C53 R63 A72 K77 Y84 N85 D86 V96 Q101 L102 E103 G104 V105 D106

### 4.2.9 Score per residue for model 9

- Molecule 1: OR34

Chain A: 73% • • 25%

M1 H2 H3 H4 H5 H6 H7 R8 S9 G10 C20 N21 G22 K23 T24 F25 D26 T58 R63 L75 Y84 N85 D86 Q101 L102 E103 G104 V105 D106

### 4.2.10 Score per residue for model 10

- Molecule 1: OR34

Chain A: 73% • 25%

M1 H2 H3 H4 H5 H6 H7 R8 S9 G10 C20 N21 G22 K23 T24 F25 D26 C53 R63 Y84 N85 D86 E99 G100 Q101 L102 E103 G104 V105 D106

### 4.2.11 Score per residue for model 11

- Molecule 1: OR34

Chain A: 67% 8% 25%

M1 H2 H3 H4 H5 H6 H7 R8 S9 G10 N16 C20 N21 G22 K23 T24 F25 D26 T33 E36 L44 C53 I76 K77 Y84 N85 D86 V89 E99 G100 Q101 L102 E103 G104 V105 D106

### 4.2.12 Score per residue for model 12

- Molecule 1: OR34

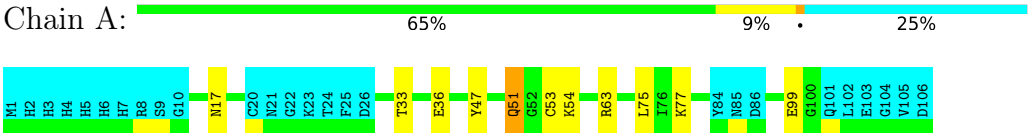
Chain A: 75% • 25%

M1 H2 H3 H4 H5 H6 H7 R8 S9 G10 C20 N21 G22 K23 T24 F25 D26 Y84 N85 D86 D92 Q101 L102 E103 G104 V105 D106

### 4.2.13 Score per residue for model 13

- Molecule 1: OR34

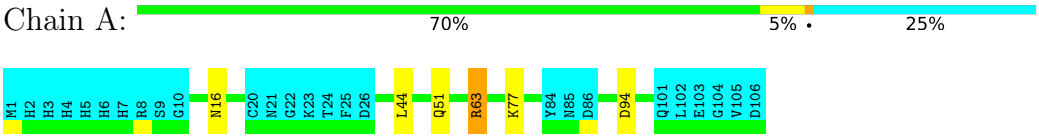
Chain A:




### 4.2.14 Score per residue for model 14

- Molecule 1: OR34

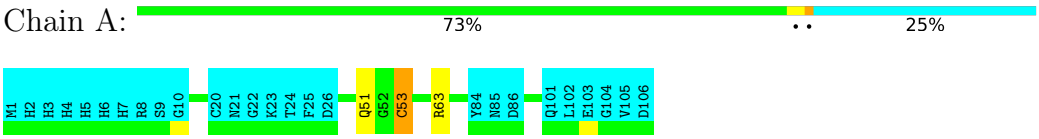
Chain A:




### 4.2.15 Score per residue for model 15

- Molecule 1: OR34

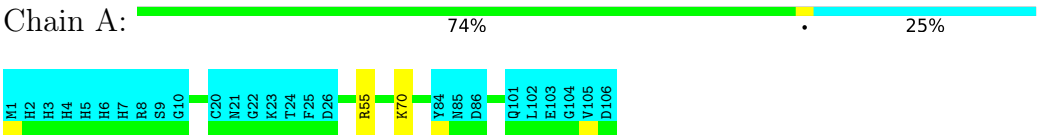
Chain A:




### 4.2.16 Score per residue for model 16

- Molecule 1: OR34

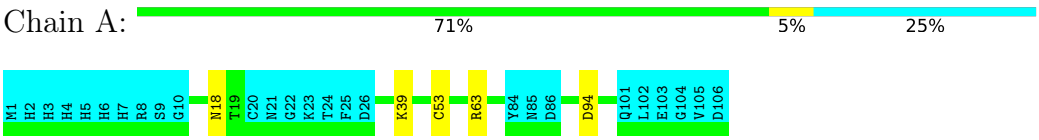
Chain A:




### 4.2.17 Score per residue for model 17

- Molecule 1: OR34

Chain A:



### 4.2.18 Score per residue for model 18

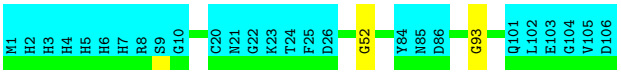
- Molecule 1: OR34

Chain A: 74% • 25%

### 4.2.19 Score per residue for model 19

- Molecule 1: OR34

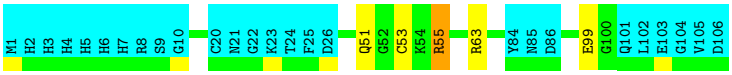
Chain A: 74% • 25%




### 4.2.20 Score per residue for model 20

- Molecule 1: OR34

Chain A: 71% • • 25%



# 5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing, distance geometry, molecular dynamics, null.*

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function.*

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

| Software name | Classification | Version |
|---|---|---|
| TALOS+ | geometry optimization | |
| CYANA | refinement | |
| CNS | 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 | 637 | 649 | 649 | 1±1 |
| All | All | 12740 | 12980 | 12980 | 22 |

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:51:GLN:HB2 | 1:A:53:CYS:SG | 0.54 | 2.43 | 20 | 1 |
| 1:A:33:THR:HB | 1:A:36:GLU:HG2 | 0.53 | 1.80 | 11 | 3 |
| 1:A:51:GLN:HG3 | 1:A:53:CYS:SG | 0.52 | 2.45 | 15 | 1 |
| 1:A:18:ASN:HB3 | 1:A:53:CYS:SG | 0.50 | 2.46 | 8 | 1 |
| 1:A:58:THR:HB | 1:A:75:LEU:HD21 | 0.48 | 1.85 | 9 | 1 |
| 1:A:16:ASN:HD21 | 1:A:44:LEU:HD21 | 0.48 | 1.69 | 14 | 1 |
| 1:A:16:ASN:HB3 | 1:A:27:TYR:HB3 | 0.47 | 1.86 | 5 | 1 |
| 1:A:16:ASN:HD21 | 1:A:56:SER:HB2 | 0.46 | 1.70 | 3 | 1 |
| 1:A:63:ARG:HA | 1:A:94:ASP:HA | 0.46 | 1.88 | 5 | 2 |
| 1:A:76:ILE:HD11 | 1:A:89:VAL:HG22 | 0.45 | 1.86 | 11 | 1 |
| 1:A:63:ARG:HB3 | 1:A:94:ASP:HA | 0.45 | 1.89 | 14 | 1 |
| 1:A:47:TYR:O | 1:A:51:GLN:HG3 | 0.44 | 2.12 | 13 | 1 |
| 1:A:51:GLN:NE2 | 1:A:53:CYS:SG | 0.42 | 2.92 | 13 | 1 |
| 1:A:16:ASN:HD21 | 1:A:44:LEU:HD11 | 0.42 | 1.73 | 11 | 1 |

*Continued on next page...*

*Continued from previous page...*

| Atom-1 | Atom-2 | Clash(Å) | Distance(Å) | Models | |
|---|---|---|---|---|---|
| | | | | Worst | Total |
| 1:A:72:ALA:HB2 | 1:A:96:VAL:HG11 | 0.41 | 1.92 | 8 | 1 |
| 1:A:18:ASN:ND2 | 1:A:53:CYS:SG | 0.41 | 2.93 | 17 | 1 |
| 1:A:55:ARG:HA | 1:A:55:ARG:NE | 0.41 | 2.30 | 20 | 1 |
| 1:A:62:ALA:HB1 | 1:A:68:ALA:HA | 0.41 | 1.91 | 5 | 1 |
| 1:A:78:VAL:O | 1:A:82:LEU:HG | 0.41 | 2.15 | 4 | 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 | 80/106 (75%) | 75±2 (94±2%) | 5±2 (6±2%) | 0±1 (1±1%) | 29 | 74 |
| All | All | 1600/2120 (75%) | 1499 (94%) | 92 (6%) | 9 (1%) | 29 | 74 |

All 5 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 | GLY | 3 |
| 1 | A | 52 | GLY | 2 |
| 1 | A | 53 | CYS | 2 |
| 1 | A | 100 | GLY | 1 |
| 1 | A | 93 | GLY | 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 | 71/94 (76%) | 69±1 (97±2%) | 2±1 (3±2%) | 46 | 90 |
| All | All | 1420/1880 (76%) | 1380 (97%) | 40 (3%) | 46 | 90 |

All 15 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 | 63 | ARG | 10 |
| 1 | A | 77 | LYS | 5 |
| 1 | A | 99 | GLU | 5 |
| 1 | A | 70 | LYS | 4 |
| 1 | A | 51 | GLN | 3 |
| 1 | A | 17 | ASN | 2 |
| 1 | A | 75 | LEU | 2 |
| 1 | A | 55 | ARG | 2 |
| 1 | A | 50 | LYS | 1 |
| 1 | A | 44 | LEU | 1 |
| 1 | A | 43 | GLU | 1 |
| 1 | A | 65 | LYS | 1 |
| 1 | A | 92 | ASP | 1 |
| 1 | A | 39 | LYS | 1 |
| 1 | A | 95 | THR | 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)

The completeness of assignment taking into account all chemical shift lists is 0% for the well-defined parts and 0% for the entire structure.