wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 09:02 pm BST

PDB ID	:	1YH5
Title	:	Solution NMR Structure of Protein yggU from Escherichia coli. Northeast Structural Genomics Consortium Target ER14.
Authors	:	Aramini, J.M.; Xiao, R.; Huang, Y.J.; Acton, T.B.; Wu, M.J.; Mills, J.L.; Tejero, R.T.; Szyperski, T.; Montelione, G.T.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2005-01-06

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 (1)) were used in the production of this report:

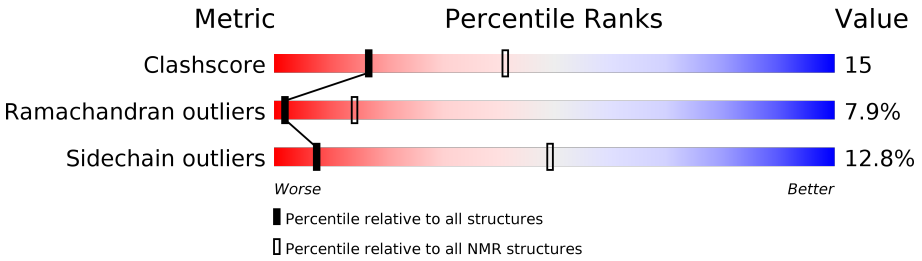
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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	108	39% 31% • • 26%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:7-A:24, A:29-A:90 (80)	0.57	4

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 1 single-model cluster was found.

Cluster number	Models
1	2, 4, 9, 10
2	3, 5, 8
3	1, 6
Single-model clusters	7

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1690 atoms, of which 852 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ORF, HYPOTHETICAL PROTEIN.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	108	1690	529	852	157	150	2	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	101	LEU	-	EXPRESSION TAG	UNP Q8XCU6
A	102	GLU	-	EXPRESSION TAG	UNP Q8XCU6
A	103	HIS	-	EXPRESSION TAG	UNP Q8XCU6
A	104	HIS	-	EXPRESSION TAG	UNP Q8XCU6
A	105	HIS	-	EXPRESSION TAG	UNP Q8XCU6
A	106	HIS	-	EXPRESSION TAG	UNP Q8XCU6
A	107	HIS	-	EXPRESSION TAG	UNP Q8XCU6
A	108	HIS	-	EXPRESSION TAG	UNP Q8XCU6

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: ORF, HYPOTHETICAL PROTEIN

Chain A: 39% 31% • • 26%

M1 D2 G3 V4 M5 S6 A7 N11 L15 V16 L17 R18 L19 Y20 I21 Q22 P23 K24 A25 S26 R27 D28 S29 I30 V31 G32 L33 H34 G35 D36 E37 V40 A41 I42 T43 A44 P45 P46 V47 D48 G49 Q50 A51 F58 L59 G60 R64 K74 G78 R79 H80 K81

N87 P88 Q89 Q90 I91 P92 P93 E94 V95 A96 A97 L98 I99 N100 L101 E102 H103 H104 H105 H106 H107 H108

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

The representative model is number 4. Colouring as in section 4.1 above.

- Molecule 1: ORF, HYPOTHETICAL PROTEIN

Chain A: 45% 21% 7% 26%

M1 D2 G3 V4 M5 S6 V10 N11 L15 L19 Y20 I21 Q22 P23 K24 A25 S26 R27 D28 S29 I30 D36 V40 A41 I42 T43 A44 P45 P46 V47 D48 G49 Q50 A51 N52 V56 R64 V65 A66 K67 G75 G78 R79 H80 N87 Q90 I91 P92 P93

E94 V95 A96 A97 L98 I99 N100 L101 E102 H103 H104 H105 H106 H107 H108

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *TORSION ANGLE DYNAMICS SIMULATED ANNEALING.*

Of the 56 calculated structures, 10 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
X-PLOR 2.0.4 (NIH), AUTOSTRUCTURE	refinement	2.0.0
VNMR 6.1B	structure solution	6.1B
NMRPipe	structure solution	2.1
SPARKY	structure solution	3.106
AUTOASSIGN	structure solution	1.9
AUTOSTRUCTURE	structure solution	2.0.0
HYPER	structure solution	2.70
TALOS	structure solution	2.1
PDBSTAT	structure solution	3.27
XPLOR	structure solution	2.0.4 (NIH), PSVS 1.0

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality (i)

6.1 Standard geometry (i)

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	2.9±0.3
All	All	0	29

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	64	ARG	Sidechain	10
1	A	79	ARG	Sidechain	10
1	A	18	ARG	Sidechain	9

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	617	647	655	20±6
All	All	6170	6470	6550	197

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 15.

5 of 120 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:18:ARG:O	1:A:19:LEU:HD12	0.71	1.86	3	3
1:A:21:ILE:N	1:A:21:ILE:HD12	0.69	2.02	4	5
1:A:19:LEU:N	1:A:19:LEU:HD22	0.67	2.04	7	1
1:A:45:PRO:O	1:A:51:ALA:HB2	0.66	1.91	1	9
1:A:19:LEU:HD22	1:A:19:LEU:N	0.64	2.08	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	80/108 (74%)	56±2 (70±3%)	18±2 (22±3%)	6±2 (8±3%)	2	14
All	All	800/1080 (74%)	562 (70%)	175 (22%)	63 (8%)	2	14

5 of 23 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	33	LEU	7
1	A	88	PRO	6
1	A	30	ILE	6
1	A	79	ARG	6
1	A	47	VAL	5

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	68/92 (74%)	59±1 (87±2%)	9±1 (13±2%)	7	49
All	All	680/920 (74%)	593 (87%)	87 (13%)	7	49

5 of 37 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	15	LEU	9
1	A	33	LEU	7
1	A	17	LEU	6
1	A	19	LEU	4
1	A	29	SER	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 carbohydrates 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