Full wwPDB NMR Structure Validation Report (i)

Mar 5, 2022 – 09:32 AM EST

PDB ID	:	2JYN
Title	:	A novel solution NMR structure of protein yst0336 from Saccharomyces cerevisiae. Northeast Structural Genomics Consortium target YT51/Ontario Centre for Structural Proteomics target yst0336
Authors	:	Wu, B.; Yee, A.; Fares, C.; Lemak, A.; Gutmanas, A.; Semest, A.; Arrowsmith, C.H.; Northeast Structural Genomics Consortium (NESG); Ontario Centre for Structural Proteomics (OCSP)
Deposited on	:	2007-12-14

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

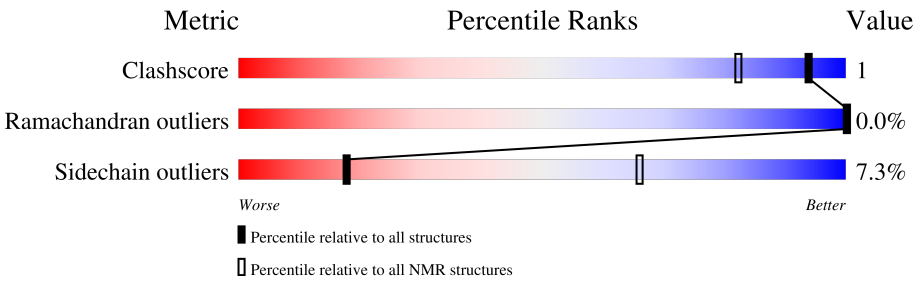
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	146	86% • 10%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 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:15-A:146 (132)	0.45	8

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 5 clusters and 6 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2439 atoms, of which 1204 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called UPF0368 protein YPL225W.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	146	2439	797	1204	201	234	3	0

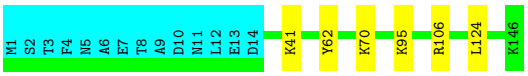
4 Residue-property plots (i)

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: UPF0368 protein YPL225W

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: UPF0368 protein YPL225W

Chain A:

4.2.2 Score per residue for model 2

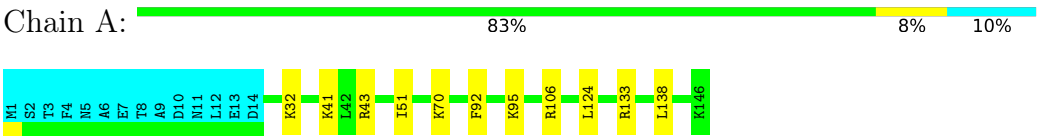
• Molecule 1: UPF0368 protein YPL225W

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: UPF0368 protein YPL225W

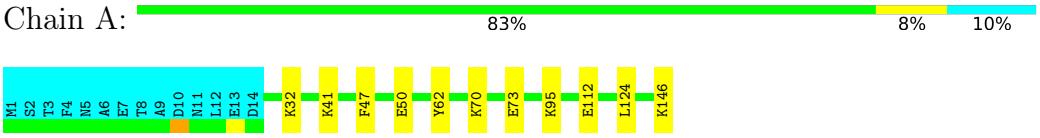
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: UPF0368 protein YPL225W

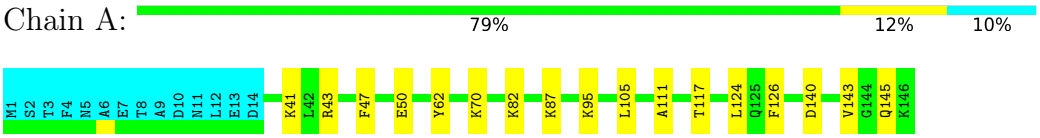
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: UPF0368 protein YPL225W

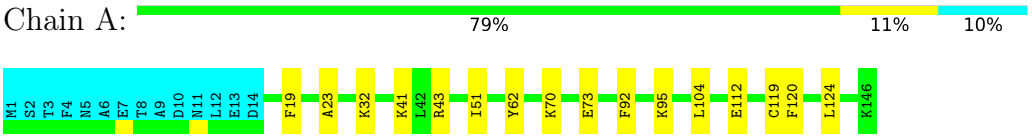
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: UPF0368 protein YPL225W

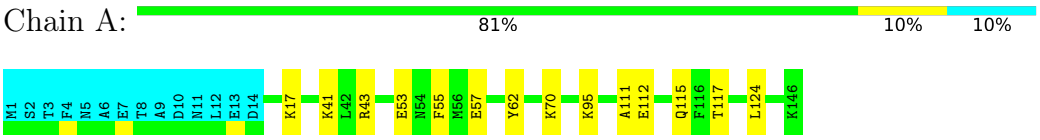
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: UPF0368 protein YPL225W

Chain A:



4.2.8 Score per residue for model 8 (medoid)

- Molecule 1: UPF0368 protein YPL225W

Chain A: 82% 9% 10%

M1 S2 T3 F4 N5 A6 E7 T8 A9 D10 N11 L12 E13 D14 K41 I51 Y52 E53 E57 E61 Y62 K70 F92 K95 Q115 L124 L138 K146

4.2.9 Score per residue for model 9

- Molecule 1: UPF0368 protein YPL225W

Chain A: 79% 10% • 10%

M1 S2 T3 F4 N5 A6 E7 T8 A9 D10 N11 L12 E13 D14 G39 S40 K41 F55 Y62 V68 K69 K70 E73 K95 R106 A111 E112 T117 R123 L124 A132 N139 K146

4.2.10 Score per residue for model 10

- Molecule 1: UPF0368 protein YPL225W

Chain A: 83% 8% 10%

M1 S2 T3 F4 N5 A6 E7 T8 A9 D10 N11 L12 E13 D14 K17 V21 K41 L44 F55 Y62 K70 F89 K95 L124 I131 K146

4.2.11 Score per residue for model 11

- Molecule 1: UPF0368 protein YPL225W

Chain A: 79% 12% 10%

M1 S2 T3 F4 N5 A6 E7 T8 A9 D10 N11 L12 E13 D14 V22 A23 K41 L42 R43 I51 F55 Y62 K70 E73 K82 K87 F92 K95 L104 L124 R133 L138 K146

4.2.12 Score per residue for model 12

- Molecule 1: UPF0368 protein YPL225W

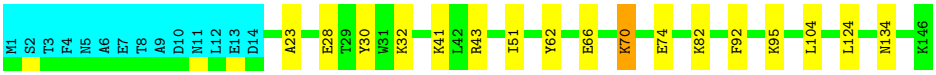
Chain A: 83% 7% • 10%

M1 S2 T3 F4 N5 A6 E7 T8 A9 D10 N11 L12 E13 D14 K32 R43 Y62 K69 K70 E74 E75 A81 K95 Q115 L124 K146

4.2.13 Score per residue for model 13

- Molecule 1: UPF0368 protein YPL225W

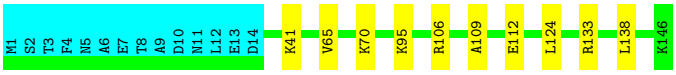
Chain A: 79% 11% • 10%



4.2.14 Score per residue for model 14

- Molecule 1: UPF0368 protein YPL225W

Chain A: 84% 7% 10%



4.2.15 Score per residue for model 15

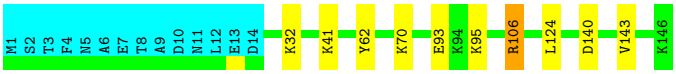
- Molecule 1: UPF0368 protein YPL225W

Chain A: 83% 7% • 10%

4.2.16 Score per residue for model 16

- Molecule 1: UPF0368 protein YPL225W

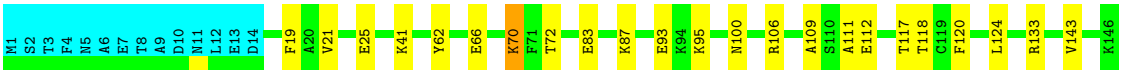
Chain A: 84% 6% • 10%



4.2.17 Score per residue for model 17

- Molecule 1: UPF0368 protein YPL225W

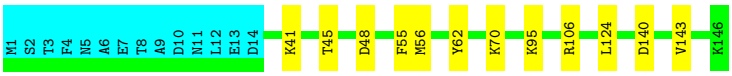
Chain A: 75% 15% • 10%



4.2.18 Score per residue for model 18

- Molecule 1: UPF0368 protein YPL225W

Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: UPF0368 protein YPL225W

Chain A:

4.2.20 Score per residue for model 20

- Molecule 1: UPF0368 protein YPL225W

Chain A:

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 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	2.1
CNS	refinement	1.0

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	1128	1112	1111	3±2
All	All	22560	22240	22220	55

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:133:ARG:HB2	1:A:138:LEU:HB2	0.55	1.78	3	3
1:A:107:THR:HG21	1:A:119:CYS:SG	0.52	2.44	20	1
1:A:66:GLU:HB3	1:A:70:LYS:HE2	0.52	1.81	17	1
1:A:73:GLU:HB3	1:A:119:CYS:SG	0.51	2.46	6	1
1:A:111:ALA:HB1	1:A:117:THR:HB	0.49	1.85	7	4
1:A:139:ASN:HD22	1:A:139:ASN:H	0.48	1.51	9	1
1:A:109:ALA:HB2	1:A:133:ARG:HD2	0.48	1.85	14	2
1:A:45:THR:H	1:A:48:ASP:HB3	0.48	1.69	18	1
1:A:23:ALA:HB1	1:A:104:LEU:HB2	0.48	1.86	6	3
1:A:110:SER:HA	1:A:141:TRP:CZ2	0.45	2.47	15	1
1:A:44:LEU:HD12	1:A:131:ILE:HG13	0.45	1.89	10	1
1:A:53:GLU:O	1:A:57:GLU:HG2	0.44	2.13	1	3
1:A:19:PHE:HZ	1:A:120:PHE:HB3	0.44	1.73	2	2
1:A:19:PHE:CZ	1:A:120:PHE:HB3	0.44	2.48	17	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:75:GLU:O	1:A:81:ALA:HB2	0.43	2.13	12	1
1:A:66:GLU:O	1:A:70:LYS:HD3	0.43	2.14	13	2
1:A:69:LYS:HG3	1:A:70:LYS:HD3	0.43	1.91	12	1
1:A:83:GLU:O	1:A:87:LYS:HD3	0.43	2.14	17	1
1:A:140:ASP:O	1:A:143:VAL:HG22	0.42	2.14	18	4
1:A:73:GLU:O	1:A:77:LYS:HG2	0.42	2.15	2	1
1:A:106:ARG:HG3	1:A:133:ARG:NH2	0.42	2.29	14	1
1:A:68:VAL:HG13	1:A:132:ALA:HB1	0.42	1.90	9	1
1:A:51:ILE:HA	1:A:92:PHE:CZ	0.41	2.51	6	5
1:A:28:GLU:O	1:A:32:LYS:HG2	0.41	2.16	20	2
1:A:47:PHE:O	1:A:50:GLU:HG2	0.41	2.16	4	2
1:A:69:LYS:HB3	1:A:138:LEU:HD23	0.41	1.93	20	1
1:A:105:LEU:HD13	1:A:126:PHE:HA	0.41	1.93	5	1
1:A:21:VAL:O	1:A:25:GLU:HG2	0.41	2.16	17	1
1:A:73:GLU:O	1:A:77:LYS:HG3	0.40	2.16	1	1
1:A:55:PHE:CD2	1:A:128:ALA:HB1	0.40	2.51	15	1
1:A:106:ARG:HA	1:A:118:THR:HG22	0.40	1.91	17	1
1:A:69:LYS:HB2	1:A:109:ALA:HB2	0.40	1.93	19	1
1:A:106:ARG:HD3	1:A:106:ARG:H	0.40	1.76	16	1
1:A:65:VAL:HG13	1:A:138:LEU:HD21	0.40	1.93	14	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	131/146 (90%)	125±2 (95±2%)	6±2 (5±2%)	0±0 (0±0%)	100	100
All	All	2620/2920 (90%)	2491 (95%)	128 (5%)	1 (0%)	100	100

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	39	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	119/131 (91%)	110±1 (93±1%)	9±1 (7±1%)	18	66
All	All	2380/2620 (91%)	2206 (93%)	174 (7%)	18	66

All 32 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	70	LYS	20
1	A	124	LEU	20
1	A	41	LYS	19
1	A	95	LYS	19
1	A	62	TYR	18
1	A	43	ARG	9
1	A	32	LYS	8
1	A	106	ARG	8
1	A	112	GLU	7
1	A	55	PHE	6
1	A	82	LYS	4
1	A	146	LYS	3
1	A	73	GLU	3
1	A	115	GLN	3
1	A	74	GLU	3
1	A	93	GLU	3
1	A	28	GLU	2
1	A	104	LEU	2
1	A	134	ASN	2
1	A	87	LYS	2
1	A	145	GLN	2
1	A	61	GLU	1
1	A	138	LEU	1
1	A	123	ARG	1
1	A	139	ASN	1
1	A	89	PHE	1
1	A	30	TYR	1
1	A	72	THR	1
1	A	100	ASN	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	143	VAL	1
1	A	56	MET	1
1	A	18	GLN	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)

No chemical shift data were provided