



wwPDB NMR Structure Validation Summary Report ⓘ

Apr 27, 2016 – 03:57 AM BST

PDB ID : 2N4U
Title : NMR structure of Fbp28 WW domain E454Y mutant
Authors : Medina, J.; Macias, M.; Martin-Malpartida, P.; Scheraga, H.A.
Deposited on : 2015-07-01

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
<http://wwpdb.org/validation/2016/NMRValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

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

Cyrange : Kirchner and Güntert (2011)
NmrClust : Kelley et al. (1996)
MolProbity : 4.02b-467
Mogul : unknown
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : rb-20027457
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20027457

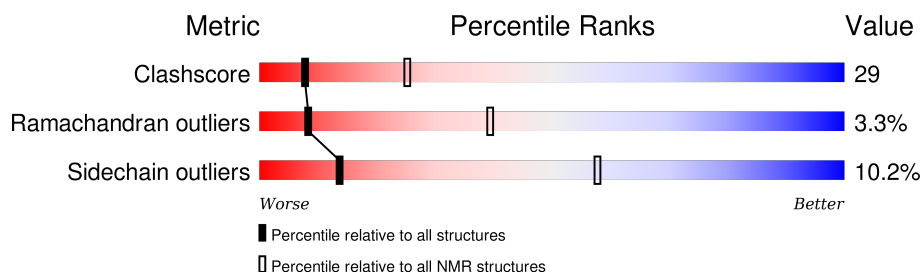
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 41%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	37	

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 20 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:434-A:460 (27)	0.17	20

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. No single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 603 atoms, of which 291 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Transcription elongation regulator 1.

Mol	Chain	Residues	Atoms					Trace
1	A	37	Total	C	H	N	O	0
			603	199	291	49	64	

There is a discrepancy between the modelled and reference sequences:

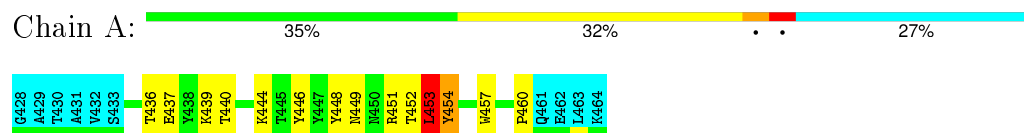
Chain	Residue	Modelled	Actual	Comment	Reference
A	454	TYR	GLU	engineered mutation	UNP O14776

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.

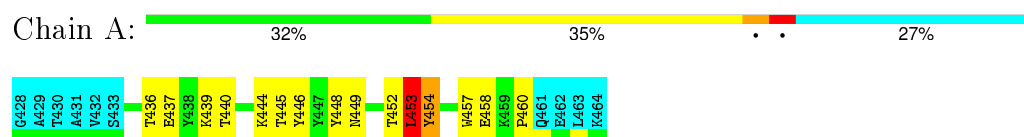
- Molecule 1: Transcription elongation regulator 1



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

The representative model is number 20. Colouring as in section 4.1 above.

- Molecule 1: Transcription elongation regulator 1



5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing*.

Of the 300 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
CNSSOLVE	structure solution	
CNSSOLVE	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	2n4u_cs.str
Number of chemical shift lists	1
Total number of shifts	228
Number of shifts mapped to atoms	228
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	41%

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](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.53±0.02	0±0/252 (0.0±0.0%)	0.70±0.03	1±1/345 (0.3±0.2%)
All	All	0.53	0/5040 (0.0%)	0.71	18/6900 (0.3%)

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	454	TYR	CB-CG-CD1	-6.54	117.08	121.00	13	16
1	A	454	TYR	CB-CG-CD2	5.26	124.16	121.00	13	2

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	242	217	217	13±2
All	All	4840	4340	4340	269

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 29.

5 of 36 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:449:ASN:HB3	1:A:454:TYR:HD1	0.98	1.14	13	20
1:A:449:ASN:HB3	1:A:454:TYR:CD1	0.94	1.97	2	20
1:A:449:ASN:O	1:A:453:LEU:HA	0.89	1.67	9	20
1:A:439:LYS:HE2	1:A:443:GLY:HA2	0.80	1.51	10	1
1:A:448:TYR:HA	1:A:454:TYR:O	0.80	1.77	1	20

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	27/37 (73%)	25±1 (94±2%)	1±1 (3±3%)	1±0 (3±1%)	8	39
All	All	540/740 (73%)	506 (94%)	16 (3%)	18 (3%)	8	39

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	453	LEU	18

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	25/32 (78%)	22±1 (90±4%)	3±1 (10±4%)	14	58
All	All	500/640 (78%)	449 (90%)	51 (10%)	14	58

5 of 8 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	437	GLU	17

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	453	LEU	13
1	A	439	LYS	8
1	A	458	GLU	5
1	A	452	THR	5

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](#)

The completeness of assignment taking into account all chemical shift lists is 41% for the well-defined parts and 40% for the entire structure.

7.1 Chemical shift list 1

File name: 2n4u_cs.str

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping [i](#)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	228
Number of shifts mapped to atoms	228
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	56

7.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	5	0.00 \pm 0.00	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	2	0.00 \pm 0.00	None needed (< 0.5 ppm)
$^{13}\text{C}'$	5	0.00 \pm 0.00	None needed (< 0.5 ppm)
^{15}N	4	0.00 \pm 0.00	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments [i](#)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 41%, i.e. 147 atoms were assigned a chemical shift out of a possible 357. 1 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	52/133 (39%)	43/53 (81%)	7/54 (13%)	2/26 (8%)
Sidechain	74/160 (46%)	70/95 (74%)	4/57 (7%)	0/8 (0%)

Continued on next page...

Continued from previous page...

	Total	¹H	¹³C	¹⁵N
Aromatic	21/64 (33%)	19/32 (59%)	2/30 (7%)	0/2 (0%)
Overall	147/357 (41%)	132/180 (73%)	13/141 (9%)	2/36 (6%)

7.1.4 Statistically unusual chemical shifts ⓘ

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	455	SER	C	3.42	183.48 – 165.88	-97.3
1	A	456	THR	C	9.92	183.43 – 165.73	-93.0
1	A	457	TRP	CD2	1.58	134.55 – 120.45	-89.3
1	A	461	GLN	C	4.19	186.20 – 166.50	-87.4
1	A	454	TYR	CD2	3.49	140.11 – 125.31	-87.3
1	A	453	LEU	C	1.55	187.06 – 166.96	-87.3
1	A	457	TRP	C	2.50	186.20 – 166.10	-86.4
1	A	457	TRP	CE3	1.50	129.06 – 111.96	-69.6
1	A	457	TRP	CG	4.17	118.98 – 101.98	-62.5
1	A	459	LYS	CE	-0.13	46.00 – 37.80	-51.3
1	A	460	PRO	CA	1.81	71.13 – 55.53	-39.4
1	A	461	GLN	N	8.88	138.01 – 101.71	-30.6
1	A	456	THR	HG23	7.85	2.29 – -0.01	29.2
1	A	456	THR	HG22	7.85	2.29 – -0.01	29.2
1	A	456	THR	HG21	7.85	2.29 – -0.01	29.2
1	A	461	GLN	CG	2.10	39.38 – 28.18	-28.3
1	A	457	TRP	N	6.71	142.48 – 100.78	-27.6
1	A	462	GLU	CA	1.06	67.86 – 46.86	-26.8
1	A	458	GLU	CA	1.16	67.86 – 46.86	-26.8
1	A	461	GLN	CA	3.93	67.31 – 45.91	-24.6
1	A	459	LYS	CD	0.83	34.86 – 23.06	-23.8
1	A	456	THR	N	2.97	139.35 – 91.55	-23.5
1	A	463	LEU	CG	1.58	32.55 – 21.05	-21.9
1	A	457	TRP	CA	7.11	70.54 – 44.84	-19.7
1	A	460	PRO	N	3.57	167.35 – 100.35	-19.4
1	A	463	LEU	CB	7.44	51.69 – 32.89	-18.5
1	A	460	PRO	HG2	7.39	3.48 – 0.38	17.6
1	A	455	SER	HB2	8.29	5.18 – 2.58	17.0
1	A	461	GLN	CB	1.84	38.36 – 19.96	-14.8
1	A	457	TRP	HD1	1.88	8.95 – 5.35	-14.6
1	A	454	TYR	HB2	8.32	4.76 – 1.06	14.6
1	A	463	LEU	CD1	1.33	32.77 – 16.57	-14.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	457	TRP	HB3	7.97	4.94 – 1.34	13.4
1	A	463	LEU	CD2	1.36	32.60 – 15.60	-13.4
1	A	453	LEU	CD2	2.06	32.60 – 15.60	-13.0
1	A	456	THR	CG2	7.18	27.15 – 15.95	-12.8
1	A	453	LEU	HD11	4.13	2.16 – -0.64	12.1
1	A	453	LEU	HD12	4.13	2.16 – -0.64	12.1
1	A	453	LEU	HD13	4.13	2.16 – -0.64	12.1
1	A	462	GLU	H	1.14	11.34 – 5.34	-12.0
1	A	458	GLU	H	1.24	11.34 – 5.34	-11.8
1	A	453	LEU	CD1	7.02	32.77 – 16.57	-10.9
1	A	454	TYR	HE1	9.18	7.86 – 5.56	10.7
1	A	459	LYS	HG2	3.71	2.67 – 0.07	9.0
1	A	460	PRO	HD2	6.72	5.45 – 1.85	8.5
1	A	449	ASN	HB3	0.00	4.41 – 1.11	-8.4
1	A	460	PRO	HA	1.73	6.05 – 2.75	-8.1
1	A	454	TYR	HB3	5.83	4.75 – 0.95	7.8
1	A	454	TYR	HD1	4.62	8.44 – 5.44	-7.7
1	A	458	GLU	HB3	0.35	3.10 – 0.90	-7.5
1	A	459	LYS	HD2	-0.06	2.76 – 0.46	-7.3
1	A	462	GLU	HB3	0.65	3.10 – 0.90	-6.1
1	A	463	LEU	HB3	3.75	3.34 – -0.26	6.1
1	A	458	GLU	HB2	0.82	3.08 – 0.98	-5.7
1	A	449	ASN	HD22	4.34	9.59 – 4.69	-5.7
1	A	459	LYS	HE2	1.89	3.87 – 1.97	-5.4

7.1.5 Random Coil Index (RCI) plots

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A:

