Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2024 – 09:04 AM EDT

PDB ID	:	2MWF
BMRB ID	:	25315
Title	:	NMR structure of FBP28 WW2 mutant Y438R DN
Authors	:	Macias, M.J.; Scheraga, H.; Sunol, D.; Todorovski, T.
Deposited on	:	2014-11-04

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:

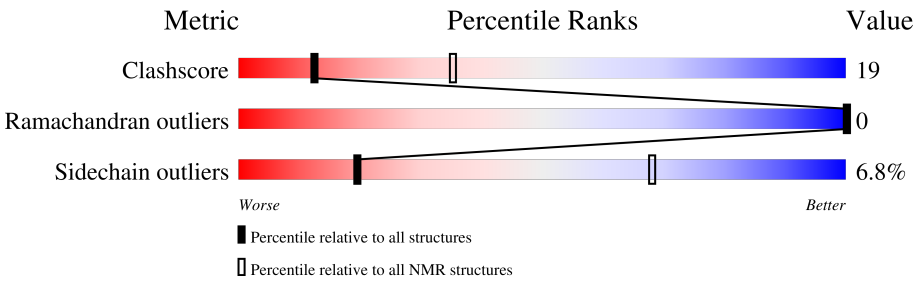
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 is 45%.

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	32	53% 31% 16%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 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:435-A:461 (27)	0.50	11

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 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 541 atoms, of which 261 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	32	541	175	261	47	58	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	438	ARG	TYR	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, 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: Transcription elongation regulator 1

Chain A: 53% 31% 16%

S433 E434 W435 T436 E437 R438 K439 Y446 Y447 Y448 N449 N450 R451 T452 W457 E462 L463 K464

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: Transcription elongation regulator 1

Chain A: 59% 25% 16%

S433 E434 W435 T436 E437 R438 T445 Y446 Y447 W457 E462 L463 K464

4.2.2 Score per residue for model 2

• Molecule 1: Transcription elongation regulator 1

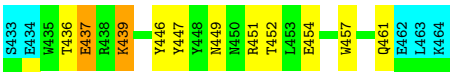
Chain A: 50% 28% 6% 16%

S433 E434 W435 T436 E437 R438 Y446 Y447 Y448 N449 N450 R451 T452 L453 E454 W457 E462 L463 K464

4.2.3 Score per residue for model 3

- Molecule 1: Transcription elongation regulator 1

Chain A: 50% 28% 6% 16%



4.2.4 Score per residue for model 4

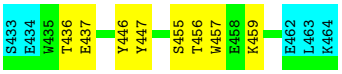
- Molecule 1: Transcription elongation regulator 1

Chain A: 50% 34% 16%

4.2.5 Score per residue for model 5

- Molecule 1: Transcription elongation regulator 1

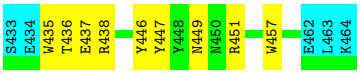
Chain A: 59% 25% 16%



4.2.6 Score per residue for model 6

- Molecule 1: Transcription elongation regulator 1

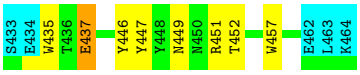
Chain A: 56% 28% 16%



4.2.7 Score per residue for model 7

- Molecule 1: Transcription elongation regulator 1

Chain A: 59% 22% • 16%



4.2.8 Score per residue for model 8

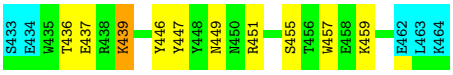
- Molecule 1: Transcription elongation regulator 1

Chain A: 47% 34% • 16%

4.2.9 Score per residue for model 9

- Molecule 1: Transcription elongation regulator 1

Chain A: 53% 28% • 16%



4.2.10 Score per residue for model 10

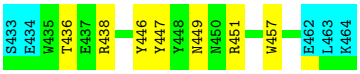
- Molecule 1: Transcription elongation regulator 1

Chain A: 38% 38% 9% 16%

4.2.11 Score per residue for model 11 (medoid)

- Molecule 1: Transcription elongation regulator 1

Chain A: 62% 22% 16%



4.2.12 Score per residue for model 12

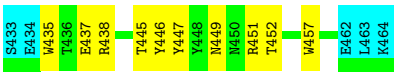
- Molecule 1: Transcription elongation regulator 1

Chain A: 47% 31% 6% 16%

4.2.13 Score per residue for model 13

• Molecule 1: Transcription elongation regulator 1

Chain A: 53% 31% 16%



4.2.14 Score per residue for model 14

• Molecule 1: Transcription elongation regulator 1

Chain A: 47% 38% 16%

4.2.15 Score per residue for model 15

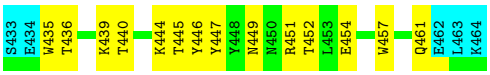
• Molecule 1: Transcription elongation regulator 1

Chain A: 47% 38% 16%

4.2.16 Score per residue for model 16

• Molecule 1: Transcription elongation regulator 1

Chain A: 41% 44% 16%



4.2.17 Score per residue for model 17

• Molecule 1: Transcription elongation regulator 1

Chain A: 53% 31% 16%

4.2.18 Score per residue for model 18

- Molecule 1: Transcription elongation regulator 1

Chain A: 53% 31% 16%

4.2.19 Score per residue for model 19

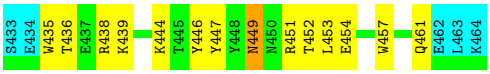
- Molecule 1: Transcription elongation regulator 1

Chain A: 56% 22% 6% 16%

4.2.20 Score per residue for model 20

- Molecule 1: Transcription elongation regulator 1

Chain A: 41% 41% • 16%



5 Refinement protocol and experimental data overview (i)

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
CNS	structure solution	
CNS	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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	204
Number of shifts mapped to atoms	204
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	45%

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	238	220	220	9±2
All	All	4760	4400	4400	173

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 19.

All 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:HD21	1:A:451:ARG:HB2	0.97	1.16	20	13
1:A:452:THR:HB	1:A:454:GLU:HG3	0.88	1.43	2	4
1:A:436:THR:O	1:A:447:TYR:HA	0.71	1.85	5	18
1:A:439:LYS:HD3	1:A:439:LYS:H	0.70	1.45	10	2
1:A:439:LYS:H	1:A:439:LYS:HD2	0.69	1.46	3	1
1:A:446:TYR:HB3	1:A:457:TRP:CH2	0.68	2.23	15	15
1:A:452:THR:HG23	1:A:454:GLU:HG3	0.64	1.68	15	1
1:A:452:THR:HG23	1:A:454:GLU:HB2	0.64	1.70	3	1
1:A:452:THR:HB	1:A:454:GLU:HG2	0.60	1.73	8	1
1:A:449:ASN:HD21	1:A:451:ARG:CB	0.60	2.10	2	7
1:A:444:LYS:HD3	1:A:457:TRP:HZ3	0.59	1.56	15	4
1:A:455:SER:HB2	1:A:457:TRP:HE1	0.58	1.58	17	3
1:A:452:THR:HB	1:A:454:GLU:CG	0.55	2.30	20	3
1:A:447:TYR:OH	1:A:459:LYS:HG2	0.54	2.02	5	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:437:GLU:HB3	1:A:447:TYR:CZ	0.54	2.37	18	4
1:A:437:GLU:HB3	1:A:447:TYR:CE2	0.54	2.38	18	8
1:A:435:TRP:CZ3	1:A:449:ASN:ND2	0.53	2.77	4	11
1:A:444:LYS:HD3	1:A:457:TRP:CZ3	0.53	2.38	15	1
1:A:452:THR:O	1:A:453:LEU:HB2	0.51	2.05	20	5
1:A:446:TYR:HB3	1:A:457:TRP:CZ3	0.50	2.41	18	15
1:A:447:TYR:HB2	1:A:456:THR:OG1	0.50	2.06	14	2
1:A:438:ARG:HD3	1:A:446:TYR:OH	0.49	2.07	10	7
1:A:449:ASN:ND2	1:A:451:ARG:HB2	0.49	2.02	20	5
1:A:440:THR:OG1	1:A:444:LYS:HG3	0.48	2.07	16	1
1:A:435:TRP:CZ2	1:A:451:ARG:HD3	0.48	2.44	13	2
1:A:439:LYS:HB3	1:A:445:THR:HG22	0.46	1.87	18	1
1:A:437:GLU:HB2	1:A:447:TYR:CE2	0.46	2.45	5	1
1:A:435:TRP:CE3	1:A:449:ASN:HB2	0.46	2.45	14	2
1:A:457:TRP:N	1:A:457:TRP:CD1	0.45	2.85	4	2
1:A:451:ARG:HA	1:A:451:ARG:NE	0.45	2.27	3	1
1:A:449:ASN:ND2	1:A:451:ARG:H	0.45	2.09	11	7
1:A:449:ASN:HB3	1:A:454:GLU:O	0.45	2.11	20	3
1:A:449:ASN:OD1	1:A:451:ARG:HB2	0.45	2.11	9	1
1:A:442:ASP:OD2	1:A:444:LYS:HE2	0.45	2.12	10	1
1:A:459:LYS:HB3	1:A:459:LYS:HZ3	0.44	1.72	9	1
1:A:438:ARG:HB3	1:A:446:TYR:CE1	0.44	2.47	11	2
1:A:436:THR:HB	1:A:448:TYR:O	0.44	2.13	17	1
1:A:451:ARG:HA	1:A:451:ARG:HE	0.44	1.71	3	1
1:A:438:ARG:HD2	1:A:446:TYR:OH	0.44	2.12	13	1
1:A:447:TYR:O	1:A:455:SER:HA	0.44	2.12	9	2
1:A:438:ARG:HB3	1:A:446:TYR:CE2	0.44	2.48	14	1
1:A:449:ASN:HD21	1:A:451:ARG:HG2	0.43	1.73	18	1
1:A:461:GLN:H	1:A:461:GLN:CD	0.43	2.17	3	1
1:A:437:GLU:HG2	1:A:437:GLU:O	0.43	2.13	19	1
1:A:457:TRP:CD1	1:A:457:TRP:N	0.42	2.87	8	4
1:A:449:ASN:HD21	1:A:451:ARG:CG	0.42	2.28	18	1
1:A:449:ASN:HD21	1:A:451:ARG:HG3	0.42	1.74	6	1
1:A:439:LYS:HD2	1:A:439:LYS:C	0.41	2.35	19	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	27/32 (84%)	26±1 (97±3%)	1±1 (3±3%)	0±0 (0±0%)	100	100
All	All	540/640 (84%)	525 (97%)	15 (3%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	25/30 (83%)	23±1 (93±6%)	2±1 (7±6%)	19	68
All	All	500/600 (83%)	466 (93%)	34 (7%)	19	68

All 9 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	452	THR	7
1	A	439	LYS	6
1	A	445	THR	5
1	A	437	GLU	5
1	A	449	ASN	4
1	A	458	GLU	3
1	A	461	GLN	2
1	A	450	ASN	1
1	A	444	LYS	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 45% for the well-defined parts and 45% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

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	204
Number of shifts mapped to atoms	204
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	7

7.1.2 Chemical shift referencing (i)

No chemical shift referencing corrections were calculated (not enough data).

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 45%, i.e. 172 atoms were assigned a chemical shift out of a possible 383. 0 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}**H**	^{13}**C**	^{15}**N**
Backbone	54/134 (40%)	54/54 (100%)	0/54 (0%)	0/26 (0%)
Sidechain	100/198 (51%)	100/124 (81%)	0/62 (0%)	0/12 (0%)
Aromatic	18/51 (35%)	18/24 (75%)	0/25 (0%)	0/2 (0%)
Overall	172/383 (45%)	172/202 (85%)	0/141 (0%)	0/40 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 45%, i.e. 203 atoms were assigned a chemical shift out of a possible 451. 0 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

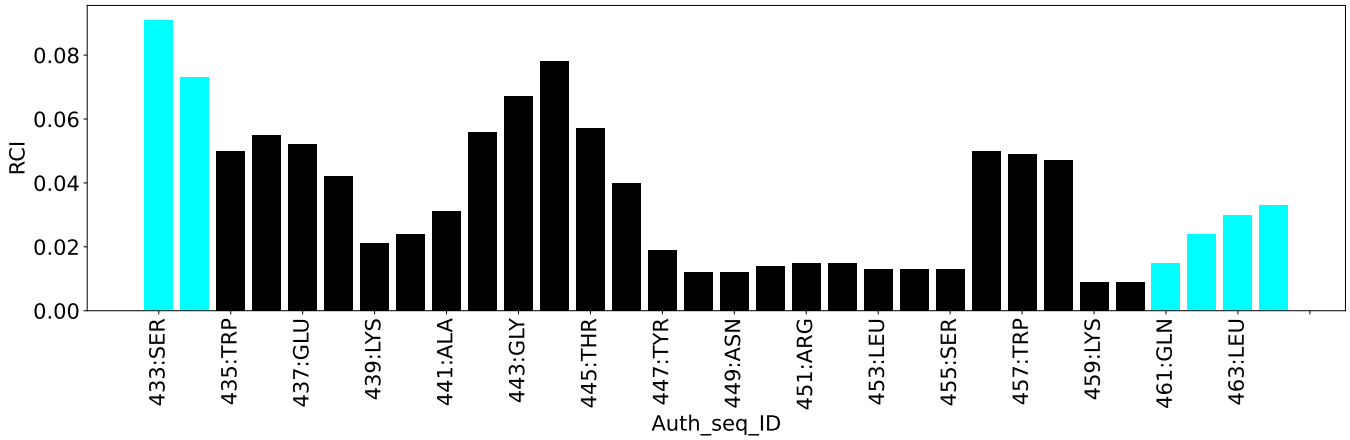
	Total	^{1}H	^{13}C	^{15}N
Backbone	63/159 (40%)	63/64 (98%)	0/64 (0%)	0/31 (0%)
Sidechain	122/241 (51%)	122/151 (81%)	0/77 (0%)	0/13 (0%)
Aromatic	18/51 (35%)	18/24 (75%)	0/25 (0%)	0/2 (0%)
Overall	203/451 (45%)	203/239 (85%)	0/166 (0%)	0/46 (0%)

7.1.4 Statistically unusual chemical shifts (i)

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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	440	THR	HG1	6.20	0.08 – 2.19	24.0
1	A	449	ASN	HB3	-0.21	1.12 – 4.38	-9.1
1	A	438	ARG	HB3	-0.27	0.43 – 3.11	-7.6
1	A	459	LYS	HD3	0.28	0.54 – 2.65	-6.3
1	A	460	PRO	HG3	-0.06	0.33 – 3.48	-6.2
1	A	460	PRO	HG2	0.11	0.41 – 3.45	-6.0
1	A	449	ASN	HD22	4.24	4.69 – 9.61	-5.9

7.1.5 Random Coil Index (RCI) plots (i)

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: