wwPDB NMR Structure Validation Summary Report (i)

Dec 6, 2016 – 10:05 PM EST

PDB ID	:	2NB1
Title	:	P63/p73 hetero-tetramerisation domain
Authors	:	Gebel, J.; Buchner, L.; Loehr, F.M.; Luh, L.M.; Coutandin, D.; Guentert, P.; Doetsch, V.
Deposited on	:	2016-01-19

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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

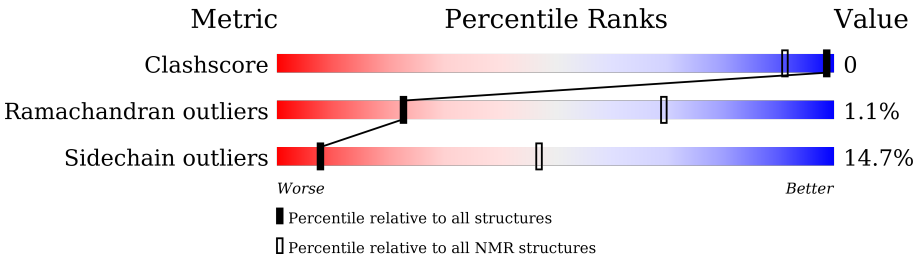
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20028442
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

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 20%.

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	60	65% · 32%
1	C	60	67% · 32%
2	B	50	82% · 14%
2	D	50	78% 8% 14%

2 Ensemble composition and analysis

This entry contains 20 models. Model 3 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:4-A:32, B:98-B:140, C:1004-C:1032, D:1098-D:1140 (144)	0.48	3
2	A:35-A:46 (12)	0.17	17
3	C:1035-C:1046 (12)	0.18	3

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3732 atoms, of which 1862 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Tumor protein 63.

Mol	Chain	Residues	Atoms						Trace
1	A	60	Total	C	H	N	O	S	0
			1016	320	503	89	102	2	
1	C	60	Total	C	H	N	O	S	0
			1016	320	503	89	102	2	

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	EXPRESSION TAG	UNP Q9H3D4
A	21	GLU	LYS	ENGINEERED MUTATION	UNP Q9H3D4
C	1001	SER	-	EXPRESSION TAG	UNP Q9H3D4
C	1021	GLU	LYS	ENGINEERED MUTATION	UNP Q9H3D4

- Molecule 2 is a protein called Tumor protein p73.

Mol	Chain	Residues	Atoms						Trace
2	B	50	Total	C	H	N	O	S	0
			850	266	428	73	81	2	
2	D	50	Total	C	H	N	O	S	0
			850	266	428	73	81	2	

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	93	GLY	-	EXPRESSION TAG	UNP O15350
B	94	SER	-	EXPRESSION TAG	UNP O15350
B	107	LYS	GLU	ENGINEERED MUTATION	UNP O15350
D	1093	GLY	-	EXPRESSION TAG	UNP O15350
D	1094	SER	-	EXPRESSION TAG	UNP O15350
D	1107	LYS	GLU	ENGINEERED MUTATION	UNP O15350

4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

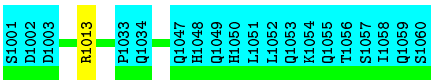
These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain 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: Tumor protein 63

Chain A:

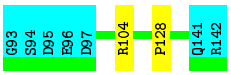
- Molecule 1: Tumor protein 63

Chain C:



- Molecule 2: Tumor protein p73

Chain B:



- Molecule 2: Tumor protein p73

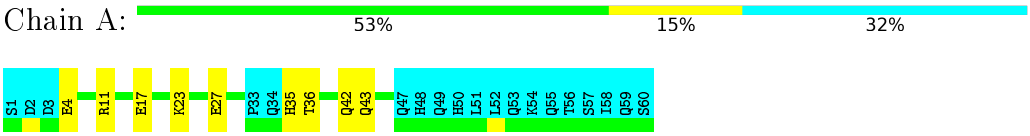
Chain D:

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

The representative model is number 3. Colouring as in section 4.1 above.

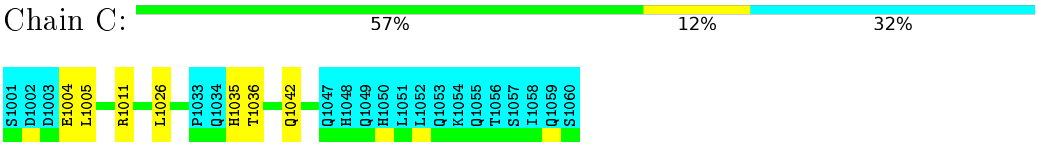
- Molecule 1: Tumor protein 63

Chain A:



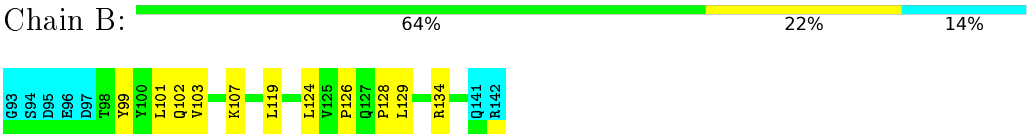
- Molecule 1: Tumor protein 63

Chain C:



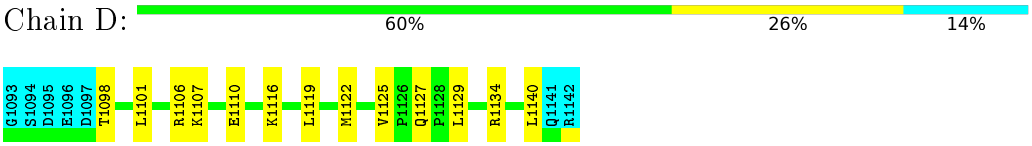
- Molecule 2: Tumor protein p73

Chain B:



- Molecule 2: Tumor protein p73

Chain D:



5 Refinement protocol and experimental data overview

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

Of the 200 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
CYANA	structure solution	3.97
OPALp	refinement	1.4

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)	2nb1_cs.cif
Number of chemical shift lists	1
Total number of shifts	2706
Number of shifts mapped to atoms	723
Number of unparsed shifts	1983
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	20%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality

6.1 Standard geometry

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	0.64±0.01	0±0/364 (0.0±0.0%)	0.95±0.04	0±0/492 (0.0±0.1%)
1	C	0.64±0.01	0±0/364 (0.0±0.0%)	0.96±0.04	0±1/492 (0.1±0.1%)
2	B	0.63±0.01	0±0/373 (0.0±0.0%)	0.99±0.04	0±0/502 (0.1±0.1%)
2	D	0.63±0.01	0±0/373 (0.0±0.0%)	0.98±0.05	1±1/502 (0.1±0.1%)
All	All	0.64	0/29480 (0.0%)	0.97	26/39760 (0.1%)

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	0.7±0.7
1	C	0.0±0.0	0.5±0.6
2	B	0.0±0.0	0.5±0.6
2	D	0.0±0.0	0.6±0.7
All	All	0	45

There are no bond-length outliers.

5 of 16 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(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	104	ARG	NE-CZ-NH2	-7.33	116.63	120.30	12	3
2	D	1104	ARG	NE-CZ-NH2	-7.27	116.67	120.30	4	4
2	B	106	ARG	NE-CZ-NH2	-6.52	117.04	120.30	6	1
2	D	1125	VAL	CA-CB-CG1	6.03	119.95	110.90	13	2
1	C	1011	ARG	NE-CZ-NH2	-5.66	117.47	120.30	12	1

There are no chirality outliers.

5 of 19 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	11	ARG	Sidechain	5
2	B	99	TYR	Sidechain	4
2	D	1099	TYR	Sidechain	4
1	A	13	ARG	Sidechain	3
1	C	1013	ARG	Sidechain	3

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	358	358	358	0±0
1	C	358	358	358	0±0
2	D	367	385	385	0±0
2	B	367	385	385	0±0
All	All	29000	29720	29720	11

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 0.

5 of 11 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:119:LEU:HD12	2:D:1108:ASN:HD21	0.49	1.67	14	1
1:C:1037:ILE:H	1:C:1037:ILE:HD12	0.47	1.70	5	1
2:B:101:LEU:HD22	2:D:1116:LYS:HE2	0.46	1.87	6	1
1:C:1028:LEU:HD22	2:D:1130:VAL:HG23	0.46	1.86	13	1
1:C:1036:THR:HG23	2:D:1125:VAL:HG21	0.46	1.87	10	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	41/60 (68%)	39±1 (95±2%)	2±1 (5±2%)	0±0 (0±1%)	59 88
1	C	41/60 (68%)	39±1 (96±2%)	2±1 (4±2%)	0±0 (0±1%)	59 88
2	B	43/50 (86%)	40±1 (93±2%)	2±1 (5±2%)	1±1 (2±2%)	13 53
2	D	43/50 (86%)	40±1 (93±2%)	2±1 (5±2%)	1±1 (2±2%)	14 55
All	All	3360/4400 (76%)	3166 (94%)	157 (5%)	37 (1%)	23 69

5 of 8 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	128	PRO	13
2	D	1128	PRO	12
2	B	126	PRO	4
2	D	1126	PRO	4
1	C	1004	GLU	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	40/59 (68%)	34±2 (86±6%)	6±2 (14±6%)	8 49
1	C	40/59 (68%)	34±3 (86±6%)	6±3 (14±6%)	8 47
2	B	42/48 (88%)	36±2 (86±5%)	6±2 (14±5%)	8 48
2	D	42/48 (88%)	35±2 (84±5%)	7±2 (16±5%)	7 44
All	All	3280/4280 (77%)	2799 (85%)	481 (15%)	8 47

5 of 128 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)
2	D	1104	ARG	14
2	D	1134	ARG	11
2	D	1137	GLN	10
1	C	1020	LEU	10
2	B	102	GLN	10

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates

There are no carbohydrates in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 20% for the well-defined parts and 22% for the entire structure.

7.1 Chemical shift list 1

File name: 2nb1_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping

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	2706
Number of shifts mapped to atoms	723
Number of unparsed shifts	1983
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

The following errors were found when reading this chemical shift list.

- Entity instance (chain) must be specified. First 5 (of 1983) occurences are reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
724	?	94	SER	CA	58.711	0.400	1
725	?	94	SER	CB	63.860	0.400	1
726	?	94	SER	HB2	3.928	0.020	2
727	?	94	SER	HB3	3.890	0.020	2
728	?	95	ASP	N	122.025	0.400	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	58	-0.56 ± 0.13	Should be applied
$^{13}C_{\beta}$	58	0.03 ± 0.08	None needed (< 0.5 ppm)
$^{13}C'$	57	-0.15 ± 0.11	None needed (< 0.5 ppm)
^{15}N	55	-0.80 ± 0.35	Should be applied

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 20%, i.e. 475 atoms were assigned a chemical shift out of a possible 2338. 4 out of 42 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	203/828 (25%)	81/330 (25%)	82/336 (24%)	40/162 (25%)
Sidechain	245/1366 (18%)	151/804 (19%)	91/492 (18%)	3/70 (4%)
Aromatic	27/144 (19%)	18/74 (24%)	9/68 (13%)	0/2 (0%)
Overall	475/2338 (20%)	250/1208 (21%)	182/896 (20%)	43/234 (18%)

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	41	ARG	NE	136.90	92.63 – 76.73	32.8
1	A	11	ARG	NE	134.86	92.63 – 76.73	31.6
1	A	13	ARG	NE	134.05	92.63 – 76.73	31.1

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: