Full wwPDB NMR Structure Validation Report (i)

Apr 27, 2016 – 01:00 AM BST

PDB ID	:	2LHN
Title	:	RNA-binding zinc finger protein
Authors	:	Brockmann, C.; Neuhaus, D.; Stewart, M.
Deposited on	:	2011-08-12

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

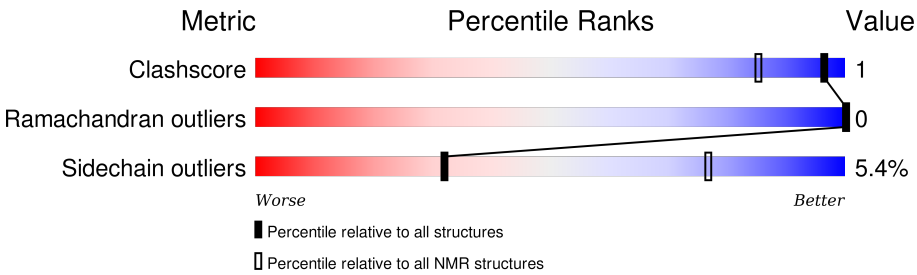
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 78%.

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	80	83% 8% 10%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 10 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:409-A:480 (72)	0.50	10

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 3 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1269 atoms, of which 626 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Nuclear polyadenylated RNA-binding protein NAB2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	80	1266	394	626	127	109	10	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	404	GLY	-	EXPRESSION TAG	UNP P32505
A	405	PRO	-	EXPRESSION TAG	UNP P32505
A	406	LEU	-	EXPRESSION TAG	UNP P32505
A	407	GLY	-	EXPRESSION TAG	UNP P32505
A	408	SER	-	EXPRESSION TAG	UNP P32505

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	3	3	3

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: Nuclear polyadenylated RNA-binding protein NAB2

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

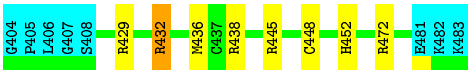
- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A:

4.2.2 Score per residue for model 2

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A:

4.2.4 Score per residue for model 4

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A:

4.2.5 Score per residue for model 5

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A:

4.2.6 Score per residue for model 6

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A:

4.2.7 Score per residue for model 7

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A:

4.2.8 Score per residue for model 8

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A: 81% 8% • 10%

G404 P405 L406 G407 S408 Q414 C415 R429 R432 R438 C448 E456 E481 K482 K483

4.2.9 Score per residue for model 9

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A: 75% 14% • 10%

G404 P405 L406 G407 S408 E409 Q414 C415 C426 R429 R432 R438 R445 C448 H452 R459 R477 E481 K482 K483

4.2.10 Score per residue for model 10 (medoid)

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A: 84% 5% • 10%

G404 P405 L406 G407 S408 R432 R438 R445 C448 H452 E481 K482 K483

4.2.11 Score per residue for model 11

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A: 71% 16% • 10%

G404 P405 L406 G407 S408 E413 Q414 C415 C421 R425 C426 K427 Y428 R429 H430 A431 R432 I435 R438 R445 C448 R459 R472 E481 K482 K483

4.2.12 Score per residue for model 12

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

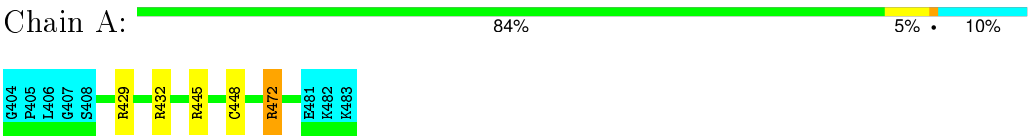
Chain A: 76% 11% • 10%

G404 P405 L406 G407 S408 E409 K410 C421 C426 M436 C437 R438 R445 C448 H452 E456 R459 R477 E481 K482 K483

4.2.13 Score per residue for model 13

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

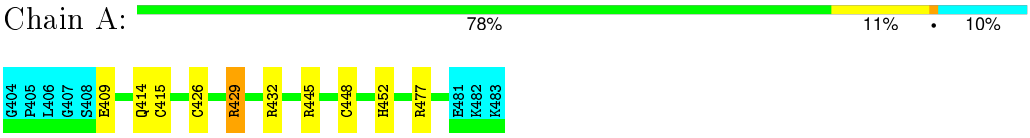
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

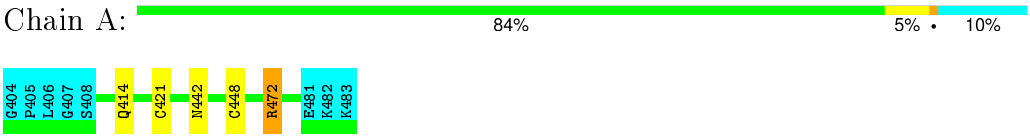
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

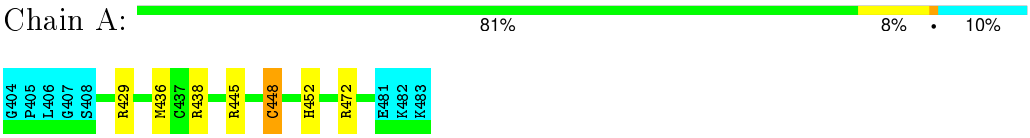
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

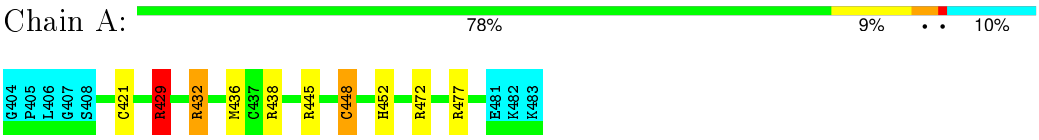
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

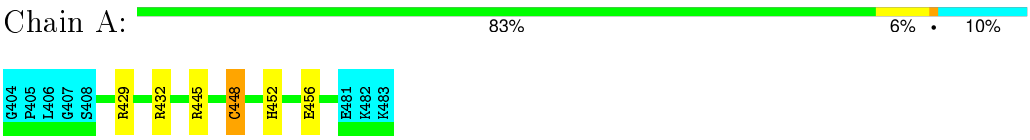
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

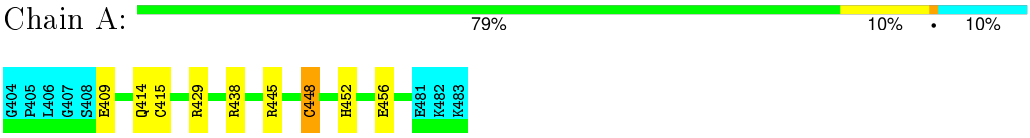
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

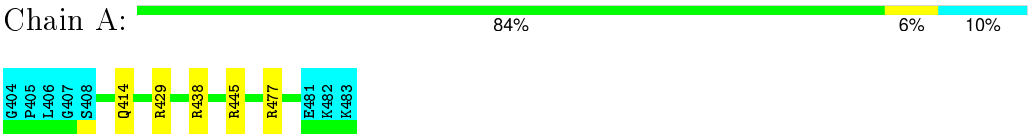
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: Nuclear polyadenylated RNA-binding protein NAB2

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 50 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
X-PLOR NIH	structure solution	
X-PLOR NIH	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)	2lhn_cs.str
Number of chemical shift lists	1
Total number of shifts	875
Number of shifts mapped to atoms	875
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	78%

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)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.75±0.00	0±0/598 (0.0±0.0%)	1.14±0.04	5±2/802 (0.6±0.2%)
All	All	0.75	0/11960 (0.0%)	1.14	91/16040 (0.6%)

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	0.2±0.5
All	All	0	5

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(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	459	ARG	NE-CZ-NH1	8.71	124.66	120.30	9	3
1	A	429	ARG	NE-CZ-NH1	8.41	124.50	120.30	8	15
1	A	445	ARG	NE-CZ-NH1	7.50	124.05	120.30	6	16
1	A	432	ARG	NE-CZ-NH1	6.43	123.52	120.30	4	11
1	A	438	ARG	NE-CZ-NH1	6.37	123.49	120.30	9	13
1	A	477	ARG	NE-CZ-NH1	6.26	123.43	120.30	20	5
1	A	472	ARG	NE-CZ-NH1	5.91	123.26	120.30	15	2
1	A	438	ARG	NE-CZ-NH2	-5.89	117.35	120.30	1	12
1	A	448	CYS	CB-CA-C	5.81	122.03	110.40	7	7
1	A	421	CYS	CA-CB-SG	-5.81	103.54	114.00	12	4
1	A	425	ARG	NE-CZ-NH1	5.50	123.05	120.30	6	3

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	426	CYS	Peptide	1
1	A	428	TYR	Sidechain	1
1	A	438	ARG	Sidechain	1
1	A	436	MET	Peptide	1
1	A	429	ARG	Sidechain	1

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	583	563	563	1±1
All	All	11720	11260	11260	25

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:448:CYS:HB3	1:A:452:HIS:CE1	0.58	2.32	18	12
1:A:448:CYS:CB	1:A:452:HIS:CE1	0.50	2.95	16	12
1:A:415:CYS:HB3	1:A:430:HIS:CD2	0.46	2.46	11	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	72/80 (90%)	70±1 (97±1%)	3±1 (3±1%)	0±0 (0±0%)	100	100
All	All	1440/1600 (90%)	1390 (97%)	50 (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	65/71 (92%)	62±2 (95±3%)	4±2 (5±3%)	32	77
All	All	1300/1420 (92%)	1230 (95%)	70 (5%)	32	77

All 17 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	448	CYS	9
1	A	414	GLN	9
1	A	472	ARG	7
1	A	415	CYS	6
1	A	456	GLU	6
1	A	429	ARG	5
1	A	409	GLU	5
1	A	477	ARG	5
1	A	432	ARG	4
1	A	426	CYS	3
1	A	435	ILE	3
1	A	436	MET	3
1	A	459	ARG	1
1	A	410	LYS	1
1	A	442	ASN	1
1	A	465	LYS	1
1	A	413	GLU	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 carbohydrates in this entry.

6.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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 78% for the well-defined parts and 79% for the entire structure.

7.1 Chemical shift list 1

File name: 2lhn_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	875
Number of shifts mapped to atoms	875
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	80	-0.21 ± 0.33	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	73	0.04 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	74	-1.93 ± 0.42	Should be applied

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 78%, i.e. 735 atoms were assigned a chemical shift out of a possible 939. 0 out of 6 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	280/352 (80%)	140/140 (100%)	72/144 (50%)	68/68 (100%)
Sidechain	399/500 (80%)	254/301 (84%)	139/164 (85%)	6/35 (17%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	56/87 (64%)	28/48 (58%)	28/34 (82%)	0/5 (0%)
Overall	735/939 (78%)	422/489 (86%)	239/342 (70%)	74/108 (69%)

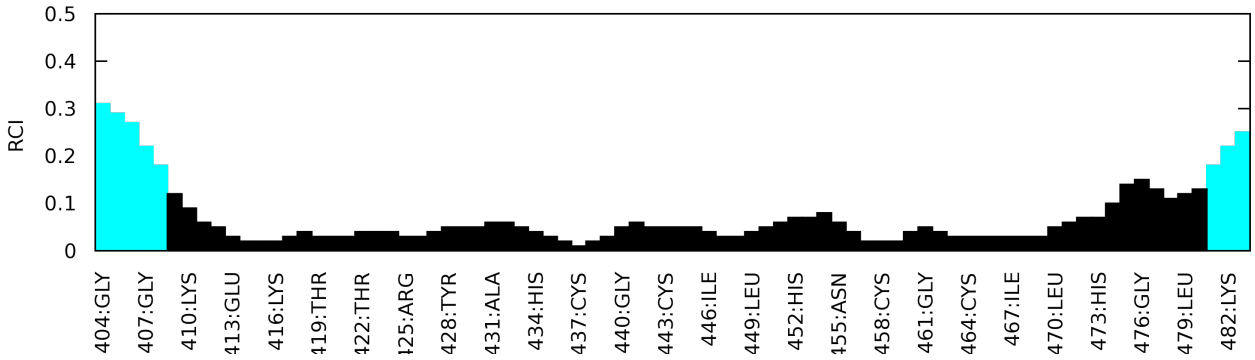
The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 79%, i.e. 814 atoms were assigned a chemical shift out of a possible 1031. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	308/390 (79%)	154/155 (99%)	80/160 (50%)	74/75 (99%)
Sidechain	450/554 (81%)	287/334 (86%)	157/183 (86%)	6/37 (16%)
Aromatic	56/87 (64%)	28/48 (58%)	28/34 (82%)	0/5 (0%)
Overall	814/1031 (79%)	469/537 (87%)	265/377 (70%)	80/117 (68%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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.

Random coil index (RCI) for chain A: