wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 04:19 PM BST

PDB ID	:	1NK2
Title	:	VND/NK-2 HOMEODOMAIN/DNA COMPLEX, NMR, 20 STRUCTURES
Authors	:	Gruschus, J.M.; Tsao, D.H.H.; Wang, L.-H.; Nirenberg, M.; Ferretti, J.A.
Deposited on	:	1998-05-06

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

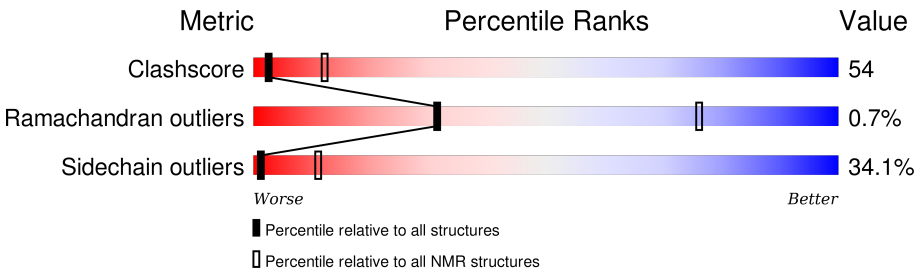
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 64%.

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	16	31% 38% 31%
2	B	16	6% 75% 19%
3	P	77	10% 34% 12% • 40%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 3 is the overall representative, medoid model (most similar to other models).

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	P:116-P:161 (46)	0.10	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 2 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2358 atoms, of which 1045 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called DNA (5'-D(*TP*GP*TP*GP*TP*CP*AP*AP*GP*TP*GP*GP*CP*TP*GP*T)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	16	513	158	184	58	98	15	0

- Molecule 2 is a DNA chain called DNA (5'-D(*AP*CP*AP*GP*CP*CP*AP*CP*TP*TP*GP*AP*CP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	16	501	154	180	62	90	15	0

- Molecule 3 is a protein called HOMEOBOX PROTEIN VND.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
3	P	77	1344	416	681	134	113	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	101	ALA	GLY	CONFLICT	UNP P22808

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: DNA (5'-D(*TP*GP*TP*GP*TP*CP*AP*AP*GP*TP*GP*GP*CP*TP*GP*T)-3')

Chain A: 31% 38% 31%

T1 G2 T3 G4 T5 C6 A7 A8 G9 T10 G11 G12 C13 T14 G15 T16

• Molecule 2: DNA (5'-D(*AP*CP*AP*GP*CP*CP*AP*CP*TP*TP*GP*AP*CP*AP*CP*A)-3')

Chain B: 6% 75% 19%

A17 C18 A19 G20 C21 C22 A23 C24 T25 T26 G27 A28 C29 A30 C31 A32

• Molecule 3: HOMEOBOX PROTEIN VND

Chain P: 10% 34% 12% • 40%

A101 S102 D103 G104 L105 P106 N107 K108 K109 R110 K111 R112 R113 V114 L115 F116 T117 K118 A119 Q120 T121 Y122 E123 L124 E125 R126 R127 F128 R129 Q130 Q131 R132 Y133 L134 S135 E138 R139 E140 H141 L142 A143 S144 L145 I146 R147 L148 T149 P150 T151 Q152 V153 K154 I155 W156 F157 Q158 N159 H160 R161

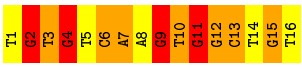
Y162 K163 T164 K165 R166 A167 Q168 N169 E170 K171 G172 Y173 E174 G175 H176 P177

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: DNA (5'-D(*TP*GP*TP*GP*TP*CP*AP*AP*GP*TP*GP*GP*CP*TP*GP*T)-3')

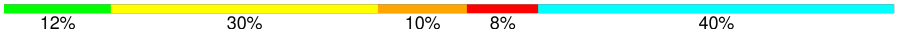
Chain A: 31% 44% 25%

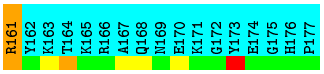


• Molecule 2: DNA (5'-D(*AP*CP*AP*GP*CP*CP*AP*CP*TP*TP*GP*AP*CP*AP*CP*A)-3')

Chain B:

• Molecule 3: HOMEOBOX PROTEIN VND

Chain P: 



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DISTANCE GEOMETRY/ SIMULATED ANNEALING*.

Of the 80 calculated structures, 20 were deposited, based on the following criterion: *LEAST RESTRAINT VIOLATION*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	**Classification**	**Version**
INSIGHT II 97 AUTHOR : MSI	refinement	MSI
NMR REFINE	structure solution	REFINE

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)	BMRB entry 4141
Number of chemical shift lists	3
Total number of shifts	1313
Number of shifts mapped to atoms	1313
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	64%

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	2.33±0.07	8±2/368 (2.1±0.4%)	2.00±0.05	13±3/568 (2.2±0.5%)
2	B	2.28±0.04	9±2/360 (2.4±0.5%)	2.03±0.05	21±5/552 (3.7±0.9%)
3	P	1.44±0.05	1±0/418 (0.2±0.1%)	1.61±0.03	6±2/564 (1.0±0.3%)
All	All	2.03	345/22920 (1.5%)	1.89	784/33680 (2.3%)

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	7.0±1.2
2	B	0.0±0.0	8.0±1.2
3	P	0.0±0.0	2.6±1.1
All	All	0	353

5 of 22 unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	11	DG	C6-N1	11.01	1.47	1.39	19	20
2	B	27	DG	C6-N1	8.02	1.45	1.39	12	20
1	A	9	DG	C6-N1	7.16	1.44	1.39	11	19
2	B	20	DG	C6-N1	7.12	1.44	1.39	19	19
1	A	13	DC	C4-C5	7.08	1.48	1.43	14	19

5 of 88 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	3	DT	P-O3'-C3'	12.28	134.43	119.70	15	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	22	DC	P-O3'-C3'	10.52	132.32	119.70	18	5
1	A	4	DG	P-O3'-C3'	10.20	131.94	119.70	4	2
1	A	2	DG	P-O3'-C3'	9.51	131.12	119.70	19	2
1	A	6	DC	P-O3'-C3'	9.05	130.56	119.70	2	2

There are no chirality outliers.

5 of 32 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
2	B	24	DC	Sidechain	20
2	B	26	DT	Sidechain	20
2	B	32	DA	Sidechain	20
1	A	11	DG	Sidechain	20
2	B	20	DG	Sidechain	19

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	329	184	184	31±5
2	B	321	180	180	23±4
3	P	407	411	409	48±7
All	All	21140	15500	15460	1966

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 54.

5 of 358 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
3:P:146:ILE:HD11	3:P:148:LEU:HD12	1.12	1.18	18	20
3:P:139:ARG:CZ	3:P:154:LYS:HG2	0.93	1.92	12	15
3:P:139:ARG:NH2	3:P:154:LYS:HG3	0.90	1.81	7	2
1:A:7:DA:H2"	1:A:8:DA:H5'	0.87	1.46	18	3
3:P:123:GLU:CD	3:P:146:ILE:HG22	0.86	1.91	16	8

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
3	P	46/77 (60%)	39±2 (84±3%)	7±2 (15±4%)	0±1 (1±1%)	31 76
All	All	920/1540 (60%)	774 (84%)	140 (15%)	6 (1%)	31 76

All 2 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	P	153	VAL	4
3	P	138	GLU	2

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
3	P	43/69 (62%)	28±2 (66±4%)	15±2 (34±4%)	1 11
All	All	860/1380 (62%)	567 (66%)	293 (34%)	1 11

5 of 28 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)
3	P	126	ARG	20
3	P	146	ILE	20
3	P	154	LYS	20
3	P	161	ARG	19
3	P	135	SER	19

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 64% for the well-defined parts and 69% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 4141

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	1000
Number of shifts mapped to atoms	1000
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	75	-0.27 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	74	0.25 ± 0.10	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	71	-0.74 ± 0.30	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 43%, i.e. 561 atoms were assigned a chemical shift out of a possible 1305. 0 out of 6 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	180/226 (80%)	90/90 (100%)	46/92 (50%)	44/44 (100%)
Sidechain	318/376 (85%)	200/224 (89%)	101/123 (82%)	17/29 (59%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	63/71 (89%)	33/37 (89%)	29/29 (100%)	1/5 (20%)
Overall	561/1305 (43%)	323/727 (44%)	176/460 (38%)	62/118 (53%)

7.1.4 Statistically unusual chemical shifts (i)

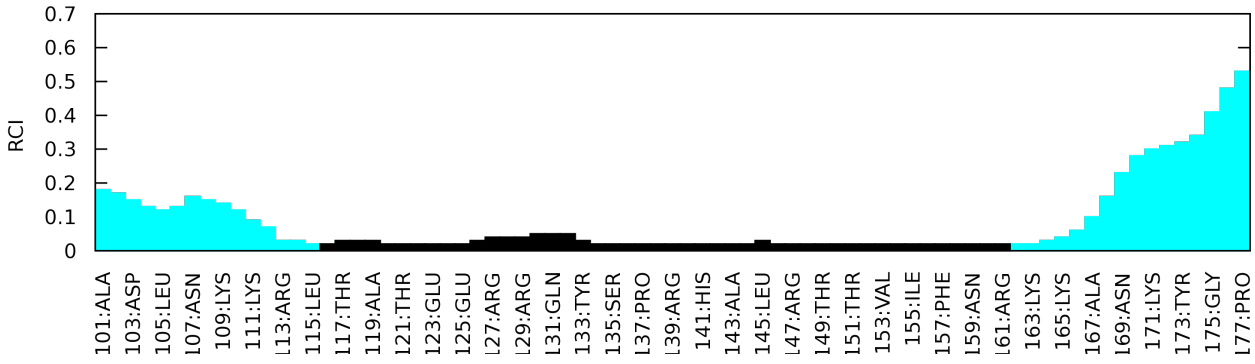
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
???	P	132	ARG	NE	110.80	92.63 – 76.73	16.4
???	P	124	LEU	HB2	-1.29	3.32 – -0.08	-8.6
???	P	152	GLN	HB3	3.39	3.37 – 0.67	5.1

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 P:



7.2 Chemical shift list 2

File name: BMRB entry 4141

Chemical shift list name: *assigned_chem_shift_list_2*

7.2.1 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	165
Number of shifts mapped to atoms	165
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

7.2.2 Chemical shift referencing (i)

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

7.2.3 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 10%, i.e. 136 atoms were assigned a chemical shift out of a possible 1305. 0 out of 6 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	0/226 (0%)	0/90 (0%)	0/92 (0%)	0/44 (0%)
Sidechain	0/376 (0%)	0/224 (0%)	0/123 (0%)	0/29 (0%)
Aromatic	0/71 (0%)	0/37 (0%)	0/29 (0%)	0/5 (0%)
Overall	136/1305 (10%)	136/727 (19%)	0/460 (0%)	0/118 (0%)

7.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.2.5 Random Coil Index (RCI) plots (i)

No *random coil index* (RCI) plot could be generated from the current chemical shift list (assigned_chem_shift_list_2). RCI is only applicable to proteins.

7.3 Chemical shift list 3

File name: BMRB entry 4141

Chemical shift list name: *assigned_chem_shift_list_3*

7.3.1 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	148
Number of shifts mapped to atoms	148
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.3.2 Chemical shift referencing (i)

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

7.3.3 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 11%, i.e. 138 atoms were assigned a chemical shift out of a possible 1305. 0 out of 6 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	0/226 (0%)	0/90 (0%)	0/92 (0%)	0/44 (0%)
Sidechain	0/376 (0%)	0/224 (0%)	0/123 (0%)	0/29 (0%)
Aromatic	0/71 (0%)	0/37 (0%)	0/29 (0%)	0/5 (0%)
Overall	138/1305 (11%)	138/727 (19%)	0/460 (0%)	0/118 (0%)

7.3.4 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.

Mol	**Chain**	**Res**	**Type**	**Atom**	**Shift, *ppm***	**Expected range, *ppm***	**Z-score**
2	B	24	DC	H5	3.48	7.21 – 4.01	-6.7

7.3.5 Random Coil Index (RCI) plots (i)

No *random coil index* (RCI) plot could be generated from the current chemical shift list (assigned_chem_shift_list_3). RCI is only applicable to proteins.