Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 04:00 PM BST

PDB ID	:	1NH5
Title	:	AUTOMATIC ASSIGNMENT OF NMR DATA AND DETERMINATION OF THE PROTEIN STRUCTURE OF A NEW WORLD SCORPION NEUROTOXIN USING NOAH/DIAMOD
Authors	:	Xu, Y.; Jablonsky, M.J.; Jackson, P.L.; Krishna, N.R.; Braun, W.
Deposited on	:	2002-12-18

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:

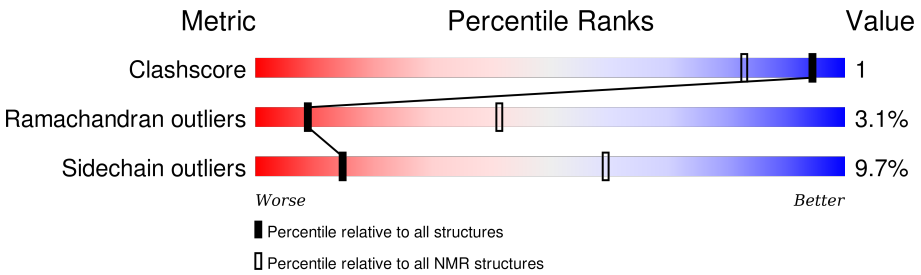
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 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	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	85% 10% • •

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 17 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	A:1-A:58 (58)	0.49	17

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 5 clusters and 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 838 atoms, of which 403 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Neurotoxin 5.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	60	838	261	403	75	89	10	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	60	GLY	-	INSERTION	UNP P58779

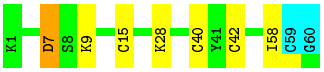
4 Residue-property plots (i)

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: Neurotoxin 5

Chain A:



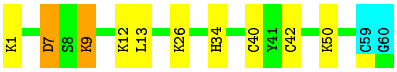
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Neurotoxin 5

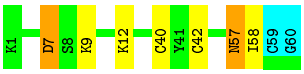
Chain A:



4.2.2 Score per residue for model 2

- Molecule 1: Neurotoxin 5

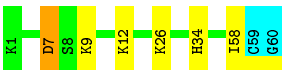
Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: Neurotoxin 5

Chain A:



4.2.4 Score per residue for model 4

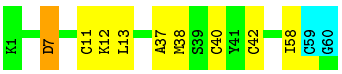
- Molecule 1: Neurotoxin 5

Chain A:

4.2.5 Score per residue for model 5

- Molecule 1: Neurotoxin 5

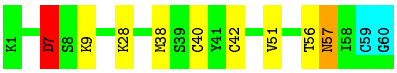
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Neurotoxin 5

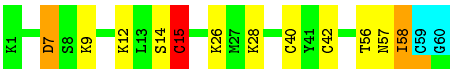
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Neurotoxin 5

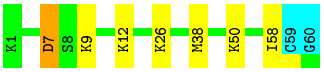
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Neurotoxin 5

Chain A: 85% 10%



4.2.9 Score per residue for model 9

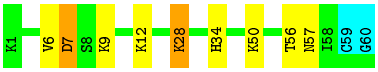
- Molecule 1: Neurotoxin 5

Chain A: 87% 10%

4.2.10 Score per residue for model 10

- Molecule 1: Neurotoxin 5

Chain A: 82% 12%



4.2.11 Score per residue for model 11

- Molecule 1: Neurotoxin 5

Chain A: 85% 10%

4.2.12 Score per residue for model 12

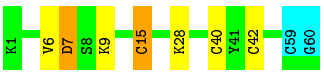
- Molecule 1: Neurotoxin 5

Chain A: 82% 12%

4.2.13 Score per residue for model 13

- Molecule 1: Neurotoxin 5

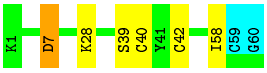
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Neurotoxin 5

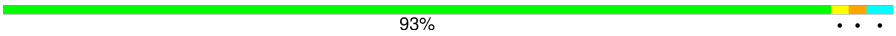
Chain A:

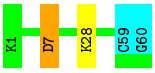


4.2.15 Score per residue for model 15

- Molecule 1: Neurotoxin 5

Chain A:

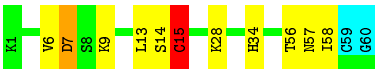




4.2.16 Score per residue for model 16

- Molecule 1: Neurotoxin 5

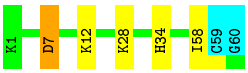
Chain A:



4.2.17 Score per residue for model 17 (medoid)

- Molecule 1: Neurotoxin 5

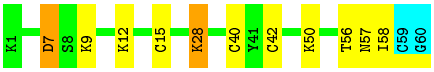
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Neurotoxin 5

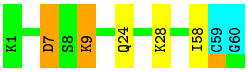
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: Neurotoxin 5

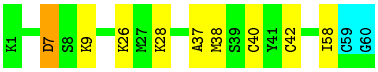
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: Neurotoxin 5

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *SELF-CORRECTING DG/ VTF*.

Of the 50 calculated structures, 20 were deposited, based on the following criterion: *VTF VALUES*.

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

Software name	Classification	Version
NOAH/DIAMOD/FANTOM	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)	BMRB entry 5672
Number of chemical shift lists	1
Total number of shifts	363
Number of shifts mapped to atoms	354
Number of unparsed shifts	0
Number of shifts with mapping errors	9
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	45%

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)

There are no covalent bond-length or bond-angle outliers.

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	0.3±0.6
All	All	0	6

There are no bond-length outliers.

There are no bond-angle outliers.

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	58	ILE	Peptide	3
1	A	51	VAL	Peptide	1
1	A	28	LYS	Peptide	1
1	A	7	ASP	Peptide	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	425	396	399	1±1
All	All	8500	7920	7980	24

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:40:CYS:SG	1:A:42:CYS:SG	0.67	2.93	18	11
1:A:9:LYS:HB3	1:A:9:LYS:HZ3	0.50	1.66	19	1
1:A:9:LYS:HZ3	1:A:9:LYS:HB3	0.49	1.67	1	1
1:A:57:ASN:CG	1:A:57:ASN:O	0.48	2.52	2	1
1:A:37:ALA:O	1:A:38:MET:C	0.47	2.53	20	2
1:A:14:SER:O	1:A:15:CYS:SG	0.45	2.74	16	3
1:A:56:THR:O	1:A:57:ASN:C	0.42	2.58	16	5

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	57/60 (95%)	49±2 (86±4%)	6±2 (11±3%)	2±1 (3±1%)	9	41
All	All	1140/1200 (95%)	979 (86%)	126 (11%)	35 (3%)	9	41

All 4 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	7	ASP	18
1	A	28	LYS	11
1	A	15	CYS	5
1	A	57	ASN	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	48/49 (98%)	43±2 (90±4%)	5±2 (10±4%)	15	59
All	All	960/980 (98%)	867 (90%)	93 (10%)	15	59

All 18 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	7	ASP	20
1	A	9	LYS	15
1	A	12	LYS	9
1	A	26	LYS	7
1	A	50	LYS	6
1	A	34	HIS	5
1	A	15	CYS	5
1	A	13	LEU	5
1	A	58	ILE	4
1	A	6	VAL	4
1	A	28	LYS	3
1	A	38	MET	2
1	A	39	SER	2
1	A	1	LYS	2
1	A	11	CYS	1
1	A	57	ASN	1
1	A	23	ASN	1
1	A	24	GLN	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)

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: BMRB entry 5672

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	363
Number of shifts mapped to atoms	354
Number of unparsed shifts	0
Number of shifts with mapping errors	9
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Residue not found in structure. All 9 occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	34	MET	HG3	2.96	-1.0	1
A	34	MET	HB3	2.47	-1.0	1
A	34	MET	H	8.95	-1.0	1
A	34	MET	HA	4.685	-1.0	1
A	34	MET	HE1	2.12	-1.0	1
A	34	MET	HE2	2.12	-1.0	1
A	34	MET	HB2	2.44	-1.0	1
A	34	MET	HG2	2.93	-1.0	1
A	34	MET	HE3	2.12	-1.0	1

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. 290 atoms were assigned a chemical shift out of a possible 638. 0 out of 5 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	111/286 (39%)	111/114 (97%)	0/116 (0%)	0/56 (0%)
Sidechain	171/313 (55%)	171/187 (91%)	0/113 (0%)	0/13 (0%)
Aromatic	8/39 (21%)	8/20 (40%)	0/18 (0%)	0/1 (0%)
Overall	290/638 (45%)	290/321 (90%)	0/247 (0%)	0/70 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 45%, i.e. 295 atoms were assigned a chemical shift out of a possible 651. 0 out of 5 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	115/296 (39%)	115/118 (97%)	0/120 (0%)	0/58 (0%)
Sidechain	172/316 (54%)	172/189 (91%)	0/114 (0%)	0/13 (0%)
Aromatic	8/39 (21%)	8/20 (40%)	0/18 (0%)	0/1 (0%)
Overall	295/651 (45%)	295/327 (90%)	0/252 (0%)	0/72 (0%)

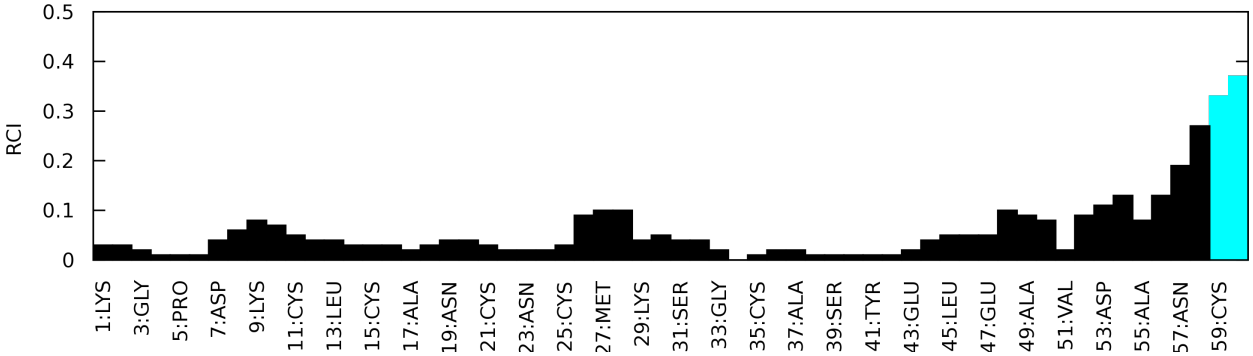
7.1.4 Statistically unusual chemical shifts (i)

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:


RCI
0.5
0.4
0.3
0.2
0.1
0
1:LYS
3:GLY
5:PRO
7:ASP
9:LYS
11:CYS
13:LEU
15:CYS
17:ALA
19:ASN
21:CYS
23:ASN
25:CYS
27:MET
29:LYS
31:SER
33:GLY
35:CYS
37:ALA
39:SER
41:TYR
43:GLU
45:LEU
47:GLU
49:ALA
51:VAL
53:ASP
55:ALA
57:ASN
59:CYS