wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 03:26 PM BST

PDB ID	:	1JWE
Title	:	NMR Structure of the N-Terminal Domain of E. Coli Dnab Helicase
Authors	:	Weigelt, J.; Brown, S.E.; Miles, C.S.; Dixon, N.E.; Otting, G.
Deposited on	:	1999-01-22

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:

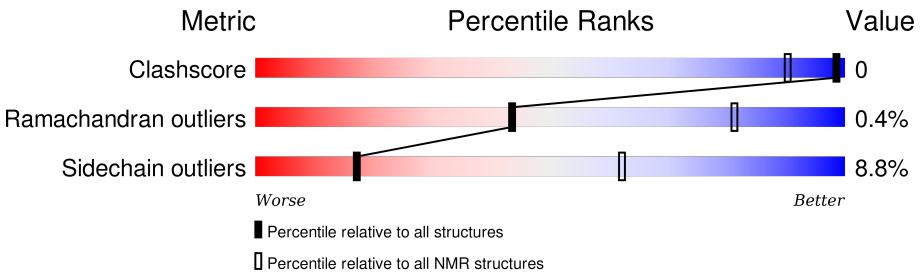
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 54%.

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	114	89% • 10%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 1 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:29-A:131 (103)	0.36	1

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 10 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1765 atoms, of which 873 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PROTEIN (DNAB HELICASE).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	114	1765	554	873	157	177	4	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	23	MET	LEU	SEE REMARK 999	UNP P0ACB0

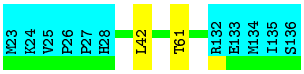
4 Residue-property plots (i)

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: PROTEIN (DNAB HELICASE)

Chain A:

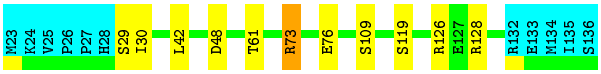


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: PROTEIN (DNAB HELICASE)

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *SIMULATED ANNEALING IN TORSION ANGLE SPACE*.

Of the 100 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
OPAL	refinement	
DYANA	structure solution	

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 5122
Number of chemical shift lists	1
Total number of shifts	957
Number of shifts mapped to atoms	897
Number of unparsed shifts	5
Number of shifts with mapping errors	55
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	54%

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	0.55±0.01	0±0/814 (0.0±0.0%)	0.99±0.03	1±1/1105 (0.1±0.1%)
All	All	0.55	0/16280 (0.0%)	0.99	17/22100 (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	1.1±0.9
All	All	0	21

There are no bond-length outliers.

5 of 11 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	46	ARG	NE-CZ-NH1	9.25	124.92	120.30	6	2
1	A	46	ARG	NE-CZ-NH2	-8.16	116.22	120.30	17	1
1	A	46	ARG	CD-NE-CZ	6.80	133.12	123.60	6	1
1	A	128	ARG	NE-CZ-NH2	-6.61	117.00	120.30	1	2
1	A	65	ARG	NE-CZ-NH2	-6.60	117.00	120.30	15	3

There are no chirality outliers.

5 of 9 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	60	TYR	Sidechain	4
1	A	126	ARG	Sidechain	3
1	A	73	ARG	Sidechain	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	92	ARG	Sidechain	2
1	A	53	ARG	Sidechain	2

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	801	775	775	1±1
All	All	16020	15500	15500	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 6 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:84:ILE:H	1:A:84:ILE:HD12	0.59	1.56	9	3
1:A:84:ILE:HD12	1:A:84:ILE:H	0.59	1.57	7	3
1:A:84:ILE:H	1:A:84:ILE:CD1	0.50	2.20	7	2
1:A:42:LEU:CD1	1:A:112:THR:HG21	0.41	2.46	16	1
1:A:66:HIS:C	1:A:66:HIS:CD2	0.41	2.93	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	103/114 (90%)	97±2 (94±2%)	5±2 (5±2%)	0±1 (0±1%)	43	81
All	All	2060/2280 (90%)	1943 (94%)	108 (5%)	9 (0%)	43	81

All 4 unique Ramachandran outliers are listed below. They are sorted by the frequency of occur-

rence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	114	SER	4
1	A	30	ILE	3
1	A	117	ASN	1
1	A	116	ALA	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	85/96 (89%)	78±2 (91±2%)	8±2 (9±2%)	17	62
All	All	1700/1920 (89%)	1550 (91%)	150 (9%)	17	62

5 of 43 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	61	THR	11
1	A	83	LEU	10
1	A	42	LEU	10
1	A	49	ASP	8
1	A	48	ASP	8

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 54% for the well-defined parts and 54% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 5122

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	957
Number of shifts mapped to atoms	897
Number of unparsed shifts	5
Number of shifts with mapping errors	55
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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

- Chemical shift has been reported more than once. All 5 occurences are reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
924	A	119	SER	N	115.4	0.2	1
925	A	119	SER	H	8.080	0.02	1
926	A	119	SER	HA	4.442	0.02	1
927	A	119	SER	HB2	3.823	0.02	2
928	A	119	SER	HB3	3.884	0.02	2

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

- Residue not found in structure. First 5 (of 55) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	114	THR	HG23	1.236	0.02	1
A	114	THR	HB	4.352	0.02	1
A	120	ILE	HA	4.114	0.02	1

Continued on next page...

Continued from previous page...

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	118	GLY	HA2	3.928	0.02	2
A	121	GLU	HB3	1.78	0.02	2

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	—
$^{13}C_{\beta}$	0	—	—
$^{13}C'$	0	—	—
^{15}N	113	0.66 ± 0.46	None needed (imprecise)

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 54%, i.e. 671 atoms were assigned a chemical shift out of a possible 1235. 0 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	300/509 (59%)	202/203 (100%)	0/206 (0%)	98/100 (98%)
Sidechain	334/649 (51%)	327/376 (87%)	0/241 (0%)	7/32 (22%)
Aromatic	37/77 (48%)	36/41 (88%)	0/33 (0%)	1/3 (33%)
Overall	671/1235 (54%)	565/620 (91%)	0/480 (0%)	106/135 (79%)

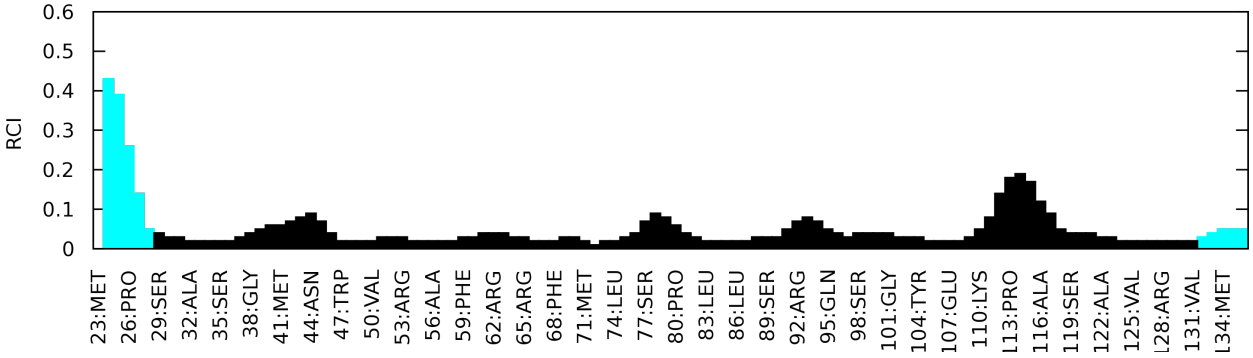
7.1.4 Statistically unusual chemical shifts (i)

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
0.6
0.5
0.4
0.3
0.2
0.1
0
23:MET
26:PRO
29:SER
32:ALA
35:SER
38:GLY
41:MET
44:ASN
47:TRP
50:VAL
53:ARG
56:ALA
59:PHE
62:ARG
65:ARG
68:PHE
71:MET
74:LEU
77:SER
80:PRO
83:LEU
86:LEU
89:SER
92:ARG
95:GLN
98:SER
101:GLY
104:TYR
107:GLU
110:LYS
113:PRO
116:ALA
119:SER
122:ALA
125:VAL
128:ARG
131:VAL
134:MET