wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 08:55 PM BST

PDB ID	:	2JNB
Title	:	Solution Structure of RNA-binding protein 15.5K
Authors	:	Soss, S.E.; Flynn, P.F.
Deposited on	:	2007-01-04

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:

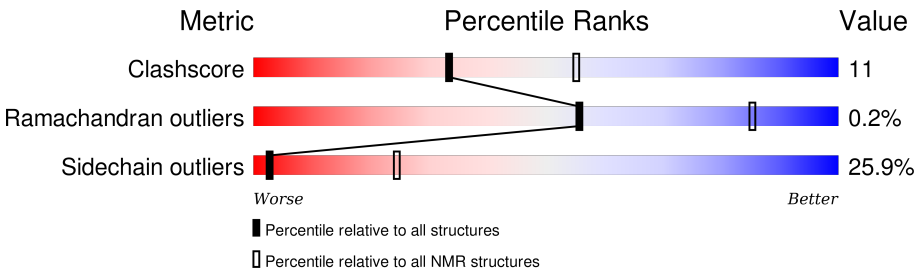
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 was not calculated.

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	144	51% 25% • 13% 11%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 7 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:10-A:70, A:79-A:127 (110)	0.66	7

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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2031 atoms, of which 1040 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called NHP2-like protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	128	2031	625	1040	175	185	6	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-15	MET	-	EXPRESSION TAG	UNP P55769
A	-14	GLY	-	EXPRESSION TAG	UNP P55769
A	-13	HIS	-	EXPRESSION TAG	UNP P55769
A	-12	HIS	-	EXPRESSION TAG	UNP P55769
A	-11	HIS	-	EXPRESSION TAG	UNP P55769
A	-10	HIS	-	EXPRESSION TAG	UNP P55769
A	-9	HIS	-	EXPRESSION TAG	UNP P55769
A	-8	HIS	-	EXPRESSION TAG	UNP P55769
A	-7	SER	-	EXPRESSION TAG	UNP P55769
A	-6	SER	-	EXPRESSION TAG	UNP P55769
A	-5	GLY	-	EXPRESSION TAG	UNP P55769
A	-4	ILE	-	EXPRESSION TAG	UNP P55769
A	-3	GLU	-	EXPRESSION TAG	UNP P55769
A	-2	GLU	-	EXPRESSION TAG	UNP P55769
A	-1	GLY	-	EXPRESSION TAG	UNP P55769
A	0	ARG	-	EXPRESSION TAG	UNP P55769

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: NHP2-like protein 1

Chain A: 51% 25% • 13% 11%

MET GLY HIS HIS HIS HIS HIS HIS SER SER GLY ILE GLU GLU GLY ARG M1 T2 E3 A4 D5 V6 N7 P8 K9 A10 Y11 P12 H17 L18 L23 D24 L25 Q28 Y32 L35 R36 K37 T45 L46 I50 S51 E52 F53 I54 V55 M56 A57 A60 E61 P62 L63

E64 I65 I66 L67 H68 L69 P70 L71 L72 C73 E74 D75 K76 N77 V78 P79 Y80 V81 S85 K86 R97 I100 S103 L112 K113 V128

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

The representative model is number 7. Colouring as in section 4.1 above.

- Molecule 1: NHP2-like protein 1

Chain A: 47% 25% 5% 13% 11%

MET GLY HIS HIS HIS HIS HIS HIS SER SER GLY ILE GLU GLU GLY ARG M1 T2 E3 A4 D5 V6 N7 P8 K9 A10 Y11 P12 L18 K21 L25 V26 Q27 Q28 L35 R36 K37 T43 L46 I50 S51 E52 F53 I54 V55 M56 A57 A60 E61 P62 L63 E64

I65 I66 L67 H68 L69 P70 L71 L72 C73 E74 D75 K76 N77 V78 V81 S85 K86 R97 I100 A101 C102 S103 V104 T105 S110 Q111 L112 I116 Q117 S118 I119 I123 L126 L127 V128

5 Refinement protocol and experimental data overview (i)

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

Of the 100 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	2.1
CYANA	refinement	2.1

No chemical shift data was provided. 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.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	853	900	900	19±4
All	All	17060	18000	18000	381

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 11.

5 of 163 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:42:ALA:HB1	1:A:54:ILE:HD13	0.98	1.31	10	1
1:A:57:ALA:HB3	1:A:60:ALA:HB2	0.95	1.34	16	14
1:A:46:LEU:HD12	1:A:69:LEU:HD23	0.94	1.39	7	1
1:A:23:LEU:HD21	1:A:88:ALA:HB1	0.89	1.43	8	5
1:A:42:ALA:HB3	1:A:69:LEU:HD22	0.85	1.46	4	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	110/144 (76%)	103±2 (93±2%)	7±2 (6±2%)	0±0 (0±0%)	56 85
All	All	2200/2880 (76%)	2053 (93%)	143 (6%)	4 (0%)	56 85

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)
1	A	85	SER	2
1	A	81	VAL	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
1	A	94/124 (76%)	70±3 (74±4%)	24±3 (26±4%)	3 24
All	All	1880/2480 (76%)	1394 (74%)	486 (26%)	3 24

5 of 71 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	35	LEU	18
1	A	25	LEU	18
1	A	17	HIS	15
1	A	103	SER	15
1	A	32	TYR	14

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)

No chemical shift data were provided