wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 03:32 PM BST

PDB ID	:	1JGN
Title	:	Solution structure of the C-terminal PABC domain of human poly(A)-binding protein in complex with the peptide from Paip2
Authors	:	Kozlov, G.; Siddiqui, N.; Coillet-Matillon, S.; Ekiel, I.; Gehring, K.
Deposited on	:	2001-06-26

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:

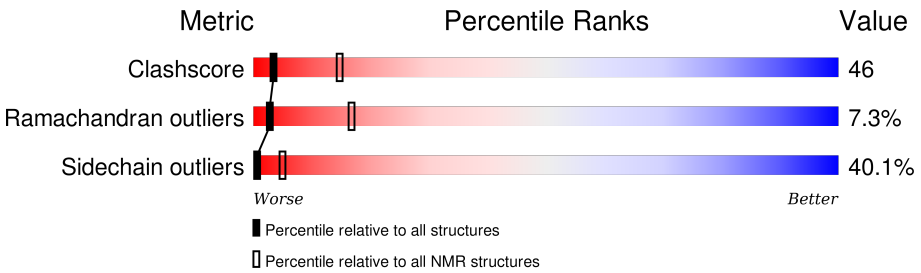
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 68%.

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	98	13% 43% 13% • 30%
2	B	22	14% 32% 5% 50%

2 Ensemble composition and analysis (i)

This entry contains 30 models. Model 16 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:8-A:15, A:19-A:79, B:5-B:13, B:18-B:19 (80)	0.21	16

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	2, 3, 5, 7, 9, 11, 12, 14, 15, 16, 17, 19, 20, 24, 25, 27, 30
2	1, 10, 18
3	6, 8
4	28, 29
5	13, 21
Single-model clusters	4; 22; 23; 26

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1815 atoms, of which 924 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called polyadenylate-binding protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	98	1469	452	747	124	141	5	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP P11940
A	2	PRO	-	CLONING ARTIFACT	UNP P11940
A	3	LEU	-	CLONING ARTIFACT	UNP P11940
A	4	GLY	-	CLONING ARTIFACT	UNP P11940
A	5	SER	-	CLONING ARTIFACT	UNP P11940

- Molecule 2 is a protein called polyadenylate-binding protein-interacting protein 2.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	22	346	109	177	29	31	0

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: polyadenylate-binding protein 1

Chain A: 13% 43% 13% • 30%

G1 P2 L3 G4 S5 P6 L7 T8 M11 L12 A15 P16 P17 Q18 E19 Q20 K21 Q22 M23 L24 G25 E26 R27 L28 F29 P30 L31 I32 Q33 A34 M35 H36 P37 T38 L39 A40 G41 K42 I43 T44 G45 M46 L47 L48 E49 I50 D51 N52 S53 E54 L55 L56 H57 M58 L59 E60 S61 P62

E63 S64 L65 R66 S67 K68 V69 D70 E71 A72 V73 A74 V75 L76 H79 Q80 A81 K82 E83 A84 A85 Q86 K87 A88 V89 N90 S91 A92 T93 G94 V95 P96 T97 V98

- Molecule 2: polyadenylate-binding protein-interacting protein 2

Chain B: 14% 32% 5% 50%

V1 V2 K3 S4 N5 L6 N7 A10 K11 E12 F13 V14 P15 G16 V17 K18 Y19 G20 N21 I22

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

The representative model is number 16. Colouring as in section 4.1 above.

- Molecule 1: polyadenylate-binding protein 1

Chain A: 19% 32% 18% • 30%

G1 P2 L3 G4 S5 P6 L7 T8 M11 L12 A15 P16 P17 Q18 E19 Q20 K21 Q22 M23 L24 G25 E26 R27 L28 F29 P30 L31 I32 Q33 A34 M35 H36 P37 T38 L39 A40 G41 K42 I43 T44 G45 M46 L47 L48 E49 I50 D51 N52 S53 E54 L55 M58 L59 E60 S61 S64

L65 R66 S67 K68 V69 D70 E71 A72 V73 A74 V75 L76 H79 Q80 A81 K82 E83 A84 A85 Q86 K87 A88 V89 N90 S91 A92 T93 G94 V95 P96 T97 V98

- Molecule 2: polyadenylate-binding protein-interacting protein 2

Chain B: 23% 9% 18% 50%

V1
V2
K3
S4
N5
L6
K11
E12
F13
V14
P15
G16
V17
K18
Y19
G20
N21
I22

5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 30 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
ARIA	structure solution	0.9
CNS	refinement	0.9

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 5085
Number of chemical shift lists	2
Total number of shifts	996
Number of shifts mapped to atoms	996
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	68%

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	524	541	541	57±7
2	B	94	91	91	5±3
All	All	18540	18960	18960	1738

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 46.

5 of 312 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:39:LEU:HD23	1:A:76:LEU:HD12	0.89	1.43	28	1
1:A:38:THR:HG21	1:A:73:VAL:HG13	0.85	1.48	13	11
1:A:31:LEU:HD13	1:A:65:LEU:HD21	0.84	1.49	15	19
1:A:28:LEU:HD22	1:A:44:THR:HA	0.82	1.52	1	16
1:A:15:ALA:HB1	1:A:19:GLU:HB3	0.78	1.55	7	19

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	69/98 (70%)	55±1 (80±2%)	10±2 (15±2%)	3±1 (5±1%)	5 26
2	B	11/22 (50%)	6±1 (58±9%)	2±2 (21±14%)	2±1 (22±13%)	0 2
All	All	2400/3600 (67%)	1846 (77%)	379 (16%)	175 (7%)	3 17

5 of 11 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	23	MET	30
1	A	39	LEU	29
2	B	10	ALA	22
2	B	18	LYS	18
1	A	28	LEU	17

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	58/79 (73%)	34±3 (59±4%)	24±3 (41±4%)	0 4
2	B	10/19 (53%)	6±1 (64±12%)	4±1 (36±12%)	1 9
All	All	2040/2940 (69%)	1221 (60%)	819 (40%)	0 5

5 of 54 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	32	ILE	30
1	A	31	LEU	30
1	A	24	LEU	30
1	A	23	MET	30
1	A	48	LEU	30

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 68% for the well-defined parts and 60% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 5085

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	876
Number of shifts mapped to atoms	876
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
$^{13}C_{\alpha}$	70	-1.07 ± 0.28	Should be applied
$^{13}C_{\beta}$	54	-0.73 ± 0.31	Should be applied
$^{13}C'$	0	—	—
^{15}N	88	1.37 ± 0.32	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 61%, i.e. 591 atoms were assigned a chemical shift out of a possible 976. 3 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	255/392 (65%)	132/156 (85%)	58/160 (36%)	65/76 (86%)
Sidechain	329/537 (61%)	221/315 (70%)	106/203 (52%)	2/19 (11%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	7/47 (15%)	7/26 (27%)	0/18 (0%)	0/3 (0%)
Overall	591/976 (61%)	360/497 (72%)	164/381 (43%)	67/98 (68%)

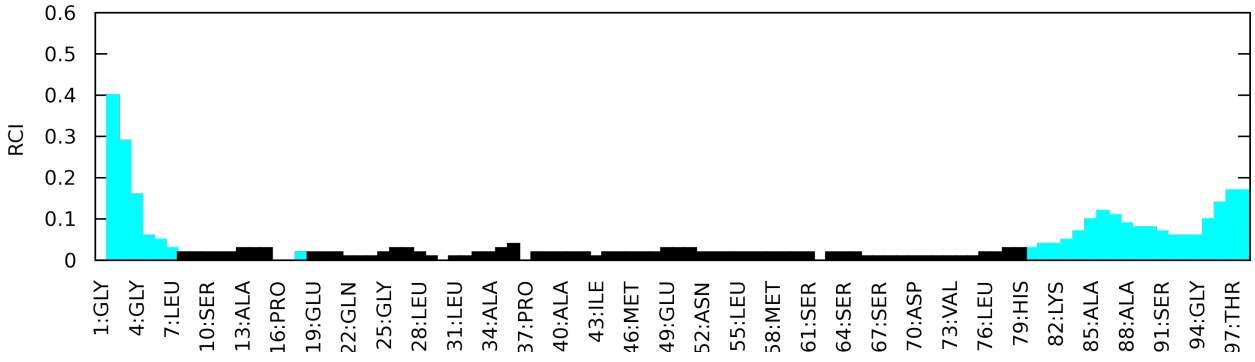
7.1.4 Statistically unusual chemical shifts (i)

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:



7.2 Chemical shift list 2

File name: BMRB entry 5085

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	120
Number of shifts mapped to atoms	120
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)

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	17	-2.65 ± 0.41	Should be applied

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 7%, i.e. 69 atoms were assigned a chemical shift out of a possible 976. 0 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	31/392 (8%)	21/156 (13%)	0/160 (0%)	10/76 (13%)
Sidechain	29/537 (5%)	29/315 (9%)	0/203 (0%)	0/19 (0%)
Aromatic	9/47 (19%)	9/26 (35%)	0/18 (0%)	0/3 (0%)
Overall	69/976 (7%)	59/497 (12%)	0/381 (0%)	10/98 (10%)

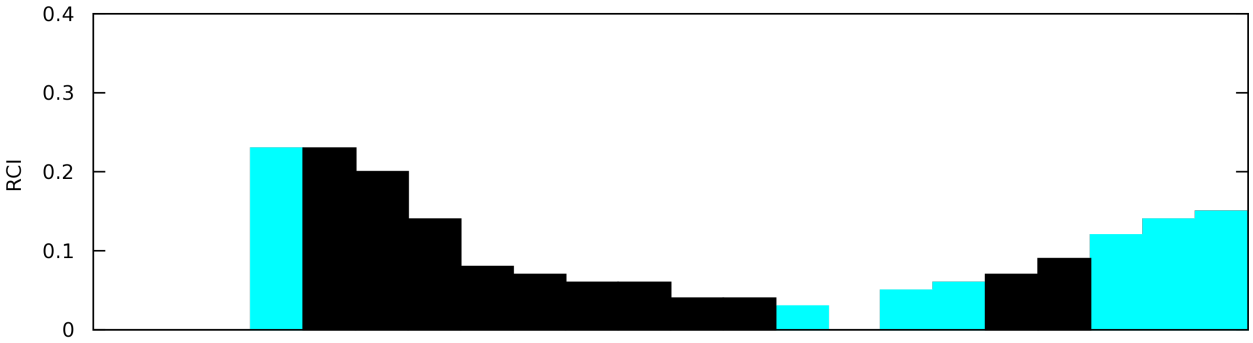
7.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.2.5 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 B:


0.4
0.3
0.2
0.1
0
RCI