wwPDB NMR Structure Validation Summary Report

Apr 21, 2024 – 01:20 PM EDT

PDB ID	:	2MKH
BMRB ID	:	19775
Title	:	Solution structure of tandem RRM domains of cytoplasmic polyadenylation element binding protein 1 (CPEB1) in free state
Authors	:	Afroz, T.; Skrisovska, L.; Belloc, E.; Boixet, J.G.; Mendez, R.; Allain, F.H.-T.
Deposited on	:	2014-02-07

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

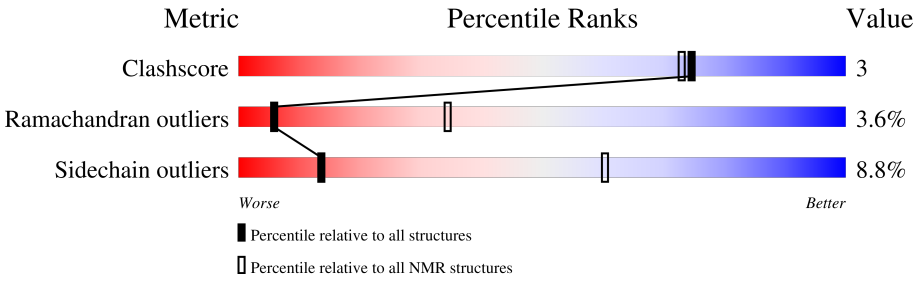
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance

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	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	217	87% 9%

2 Ensemble composition and analysis

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:218-A:304, A:309-A:429 (208)	1.80	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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3360 atoms, of which 1682 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Cytoplasmic polyadenylation element-binding protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	212	3360	1082	1682	289	299	8	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	218	MET	-	expression tag	UNP Q9BZB8

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Cytoplasmic polyadenylation element-binding protein 1

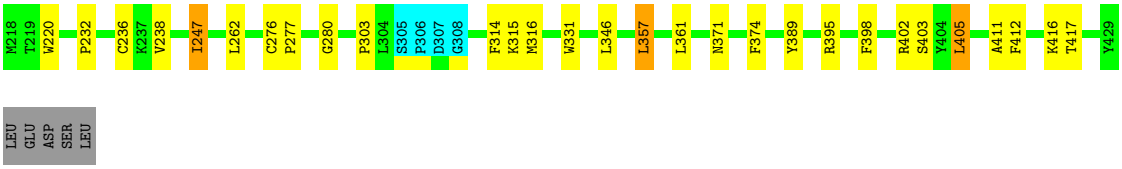
Chain A:

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: Cytoplasmic polyadenylation element-binding protein 1

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 30 calculated structures, 20 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
Amber	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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	2037
Number of shifts mapped to atoms	1991
Number of unparsed shifts	0
Number of shifts with mapping errors	46
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	68%

6 Model quality

6.1 Standard geometry

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	0.67±0.01	0±0/1699 (0.0± 0.0%)	1.03±0.03	2±1/2304 (0.1± 0.1%)
All	All	0.67	0/33980 (0.0%)	1.03	43/46080 (0.1%)

There are no bond-length outliers.

5 of 31 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	395	ARG	NE-CZ-NH1	8.31	124.46	120.30	1	1
1	A	312	TYR	CB-CG-CD2	-7.56	116.46	121.00	2	1
1	A	404	TYR	CB-CG-CD1	-7.24	116.66	121.00	13	1
1	A	229	TYR	CB-CG-CD2	-6.95	116.83	121.00	8	2
1	A	395	ARG	NE-CZ-NH2	-6.82	116.89	120.30	14	1

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts

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	1653	1663	1660	11±4
All	All	33060	33260	33200	213

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 3.

5 of 137 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:357:LEU:HD13	1:A:357:LEU:H	0.78	1.39	1	6
1:A:346:LEU:H	1:A:346:LEU:HD22	0.64	1.52	16	1
1:A:288:LEU:H	1:A:288:LEU:HD23	0.62	1.54	4	6
1:A:247:ILE:CD1	1:A:247:ILE:H	0.62	2.07	19	5
1:A:247:ILE:HD13	1:A:247:ILE:H	0.62	1.54	13	2

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	206/217 (95%)	168±4 (82±2%)	30±3 (15±1%)	7±3 (4±1%)	6	34
All	All	4120/4340 (95%)	3368 (82%)	605 (15%)	147 (4%)	6	34

5 of 55 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	411	ALA	19
1	A	224	LEU	7
1	A	287	GLU	6
1	A	246	ASP	6
1	A	386	LYS	6

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	181/189 (96%)	165±3 (91±1%)	16±3 (9±1%)	13	60

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	3620/3780 (96%)	3300 (91%)	320 (9%)	13 60

5 of 77 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	405	LEU	20
1	A	316	MET	19
1	A	247	ILE	18
1	A	220	TRP	17
1	A	262	LEU	17

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 monosaccharides 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

The completeness of assignment taking into account all chemical shift lists is 68% for the well-defined parts and 68% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *CPEB1RM12free*

7.1.1 Bookkeeping

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	2037
Number of shifts mapped to atoms	1991
Number of unparsed shifts	0
Number of shifts with mapping errors	46
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	12

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

- No matching atom found in the structure. First 5 (of 46) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	430	LEU	CA	51.0	0.000	.
1	A	430	LEU	CB	39.116	0.000	.
1	A	430	LEU	CD1	22.152	0.000	.
1	A	430	LEU	HA	4.213	0.000	.
1	A	430	LEU	H	8.13	0.002	.
1	A	430	LEU	N	123.888	0.000	.
1	A	430	LEU	HB2	2.859	0.001	.
1	A	430	LEU	HB3	2.859	0.001	.
1	A	430	LEU	HD11	0.679	0.002	.
1	A	430	LEU	HD12	0.679	0.002	.
1	A	430	LEU	HD13	0.679	0.002	.
1	A	431	GLU	CA	53.363	0.000	.
1	A	431	GLU	CB	27.454	0.035	.
1	A	431	GLU	CG	33.133	0.002	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	431	GLU	HA	4.024	0.003	.
1	A	431	GLU	HB2	1.89	0.002	.
1	A	431	GLU	HB3	1.659	0.000	.
1	A	431	GLU	HG2	2.026	0.022	.
1	A	431	GLU	HG3	1.983	0.018	.
1	A	431	GLU	H	7.569	0.000	.
1	A	431	GLU	N	121.913	0.000	.
1	A	432	ASP	CA	51.44	0.000	.
1	A	432	ASP	CB	38.203	0.002	.
1	A	432	ASP	HA	4.448	0.001	.
1	A	432	ASP	HB2	2.533	0.013	.
1	A	432	ASP	HB3	2.493	0.002	.
1	A	432	ASP	H	8.213	0.001	.
1	A	432	ASP	N	121.095	0.000	.
1	A	433	SER	CA	54.848	0.000	.
1	A	433	SER	CB	61.33	0.000	.
1	A	433	SER	HA	4.348	0.001	.
1	A	433	SER	H	7.908	0.001	.
1	A	433	SER	N	115.532	0.000	.
1	A	433	SER	HB2	3.708	0.000	.
1	A	433	SER	HB3	3.708	0.000	.
1	A	434	LEU	CA	53.854	0.000	.
1	A	434	LEU	CB	40.284	0.000	.
1	A	434	LEU	CD1	20.629	0.000	.
1	A	434	LEU	HA	4.054	0.000	.
1	A	434	LEU	H	7.837	0.000	.
1	A	434	LEU	N	130.618	0.000	.
1	A	434	LEU	HB2	1.413	0.001	.
1	A	434	LEU	HB3	1.413	0.001	.
1	A	434	LEU	HD11	0.633	0.000	.
1	A	434	LEU	HD12	0.633	0.000	.
1	A	434	LEU	HD13	0.633	0.000	.

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	152	2.98 ± 0.23	Should be applied
$^{13}C_{\beta}$	167	3.05 ± 0.13	Should be applied
$^{13}C'$	0	—	None (insufficient data)

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
^{15}N	195	-0.13 ± 0.23	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 68%, i.e. 1965 atoms were assigned a chemical shift out of a possible 2878. 0 out of 38 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	683/1027 (67%)	352/417 (84%)	144/416 (35%)	187/194 (96%)
Sidechain	1097/1575 (70%)	754/1028 (73%)	330/482 (68%)	13/65 (20%)
Aromatic	185/276 (67%)	112/135 (83%)	69/133 (52%)	4/8 (50%)
Overall	1965/2878 (68%)	1218/1580 (77%)	543/1031 (53%)	204/267 (76%)

7.1.4 Statistically unusual chemical shifts

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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	369	ILE	CG1	12.05	19.24 – 36.26	-9.2
1	A	237	LYS	HG2	3.61	0.13 – 2.61	9.0
1	A	237	LYS	HG3	3.61	0.04 – 2.67	8.6
1	A	390	PRO	HG3	-0.54	0.33 – 3.48	-7.8
1	A	390	PRO	HD2	1.01	1.93 – 5.38	-7.6
1	A	390	PRO	HD3	0.81	1.76 – 5.48	-7.6
1	A	311	GLU	CG	27.27	30.20 – 42.01	-7.5
1	A	331	TRP	HB2	0.91	1.51 – 4.87	-6.8
1	A	390	PRO	HG2	-0.05	0.41 – 3.45	-6.5
1	A	331	TRP	HB3	0.91	1.31 – 4.93	-6.1
1	A	373	LEU	HB2	-0.10	-0.07 – 3.30	-5.1
1	A	296	LEU	CD1	16.59	16.71 – 32.55	-5.1

7.1.5 Random Coil Index (RCI) plots

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: