Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 11:47 pm BST

PDB ID	:	2MKE
Title	:	Solution structure of CPEB1 ZZ domain in the free state
Authors	:	Afroz, T.; Skrisovska, L.; Belloc, E.; Boixet, J.G.; Mendez, R.; Allain, F.H.-T.
Deposited on	:	2014-02-06

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

The following versions of software and data (see references (1)) were used in the production of this report:

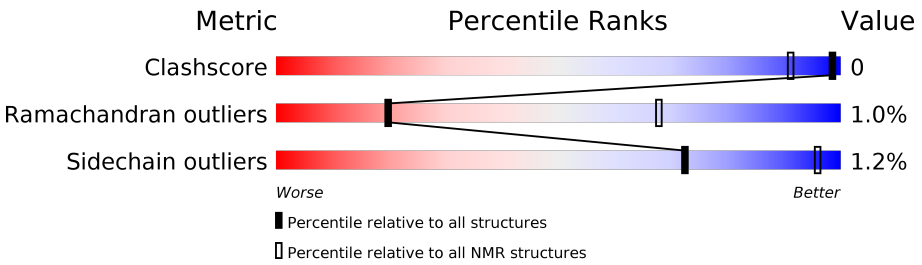
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 78%.

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	61	70% 30%

2 Ensemble composition and analysis (i)

This entry contains 26 models. Model 10 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:434-A:476 (43)	0.35	10

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 6 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 6, 7, 8, 10, 17, 20, 22, 25
2	2, 3, 4, 9, 11, 23, 26
3	13, 15, 16
4	5, 19, 21
5	12, 18
6	14, 24

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 987 atoms, of which 470 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	61	985	320	470	98	89	8	0

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	2	2	2

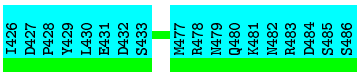
4 Residue-property plots (i)

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: Cytoplasmic polyadenylation element-binding protein 1

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: Cytoplasmic polyadenylation element-binding protein 1

Chain A:

4.2.2 Score per residue for model 2

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A:

4.2.3 Score per residue for model 3

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 69% • 30%

I426 D427 P428 Y429 L430 E431 D432 S433 K454 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.4 Score per residue for model 4

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 69% • 30%

I426 D427 P428 Y429 L430 E431 D432 S433 Q441 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.5 Score per residue for model 5

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 69% • 30%

I426 D427 P428 Y429 L430 E431 D432 S433 S439 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.6 Score per residue for model 6

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 69% • 30%

I426 D427 P428 Y429 L430 E431 D432 S433 R458 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.7 Score per residue for model 7

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

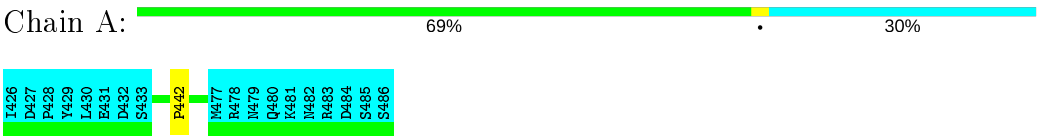
Chain A: 69% • 30%

I426 D427 P428 Y429 L430 E431 D432 S433 G469 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.8 Score per residue for model 8

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

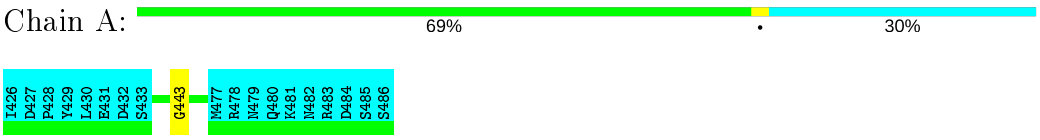
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

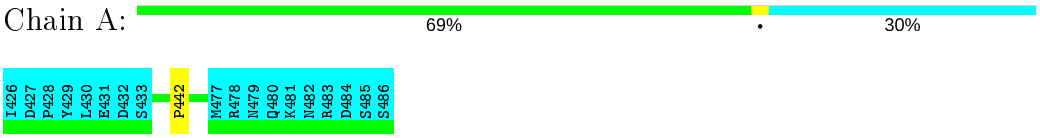
Chain A:



4.2.10 Score per residue for model 10 (medoid)

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

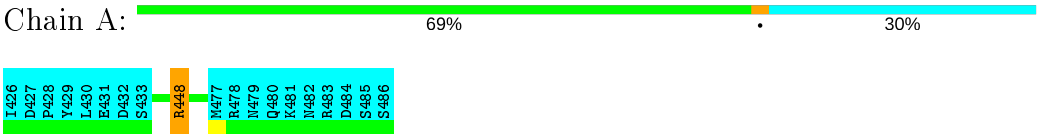
Chain A:



4.2.11 Score per residue for model 11

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

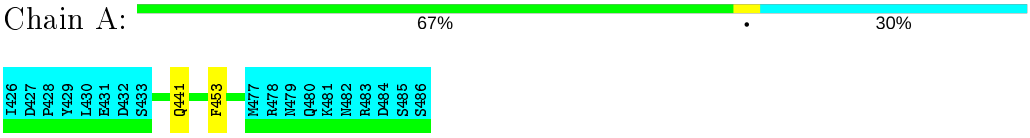
Chain A:



4.2.12 Score per residue for model 12

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A:



4.2.13 Score per residue for model 13

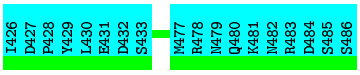
- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 70% 30%

4.2.14 Score per residue for model 14

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

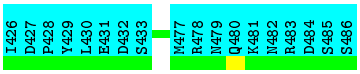
Chain A: 70% 30%



4.2.15 Score per residue for model 15

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 70% 30%



4.2.16 Score per residue for model 16

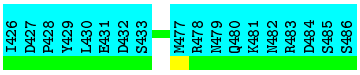
- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 66% 5% 30%

4.2.17 Score per residue for model 17

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 70% 30%



4.2.18 Score per residue for model 18

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 70% 30%

I426 D427 P428 Y429 L430 E431 D432 S433 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.19 Score per residue for model 19

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 69% • 30%

I426 D427 P428 Y429 L430 E431 D432 S433 G443 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.20 Score per residue for model 20

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 66% 5% 30%

I426 D427 P428 Y429 L430 E431 D432 S433 P442 D449 E468 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.21 Score per residue for model 21

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 67% • 30%

I426 D427 P428 Y429 L430 E431 D432 S433 G443 R458 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.22 Score per residue for model 22

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 70% 30%

I426 D427 P428 Y429 L430 E431 D432 S433 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.23 Score per residue for model 23

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 69% 30%

I426 D427 P428 Y429 L430 E431 D432 S433 P442 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.24 Score per residue for model 24

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 69% 30%

I426 D427 P428 Y429 L430 E431 D432 S433 M467 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.25 Score per residue for model 25

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 69% 30%

I426 D427 P428 Y429 L430 E431 D432 S433 F453 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

4.2.26 Score per residue for model 26

- Molecule 1: Cytoplasmic polyadenylation element-binding protein 1

Chain A: 69% 30%

I426 D427 P428 Y429 L430 E431 D432 S433 R471 M477 R478 N479 Q480 K481 N482 R483 D484 S485 S486

5 Refinement protocol and experimental data overview (i)

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

Of the 30 calculated structures, 26 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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	566
Number of shifts mapped to atoms	566
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	78%

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)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	364	329	330	0±0
All	All	9516	8554	8560	1

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.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:448:ARG:H	1:A:448:ARG:CD	0.40	2.29	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	43/61 (70%)	38±1 (88±2%)	5±1 (11±2%)	0±0 (1±1%)	20 68
All	All	1118/1586 (70%)	986 (88%)	121 (11%)	11 (1%)	20 68

All 5 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	442	PRO	4
1	A	443	GLY	3
1	A	453	PHE	2
1	A	469	GLY	1
1	A	439	SER	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	41/59 (69%)	41±1 (99±2%)	1±1 (1±2%)	72 96
All	All	1066/1534 (69%)	1053 (99%)	13 (1%)	72 96

All 9 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	441	GLN	3
1	A	458	ARG	2
1	A	468	GLU	2
1	A	450	GLN	1
1	A	454	LYS	1
1	A	471	ARG	1
1	A	467	MET	1
1	A	449	ASP	1
1	A	448	ARG	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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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 78% for the well-defined parts and 66% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

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	566
Number of shifts mapped to atoms	566
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	49	-0.46 ± 0.16	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	50	0.13 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	53	-0.54 ± 0.90	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 78%, i.e. 445 atoms were assigned a chemical shift out of a possible 574. 3 out of 4 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	163/209 (78%)	82/83 (99%)	41/86 (48%)	40/40 (100%)
Sidechain	211/262 (81%)	136/162 (84%)	73/85 (86%)	2/15 (13%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	71/103 (69%)	37/56 (66%)	30/40 (75%)	4/7 (57%)
Overall	445/574 (78%)	255/301 (85%)	144/211 (68%)	46/62 (74%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 66%, i.e. 533 atoms were assigned a chemical shift out of a possible 809. 4 out of 5 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	204/297 (69%)	102/118 (86%)	49/122 (40%)	53/57 (93%)
Sidechain	254/401 (63%)	163/245 (67%)	89/131 (68%)	2/25 (8%)
Aromatic	75/111 (68%)	39/60 (65%)	32/44 (73%)	4/7 (57%)
Overall	533/809 (66%)	304/423 (72%)	170/297 (57%)	59/89 (66%)

7.1.4 Statistically unusual chemical shifts (i)

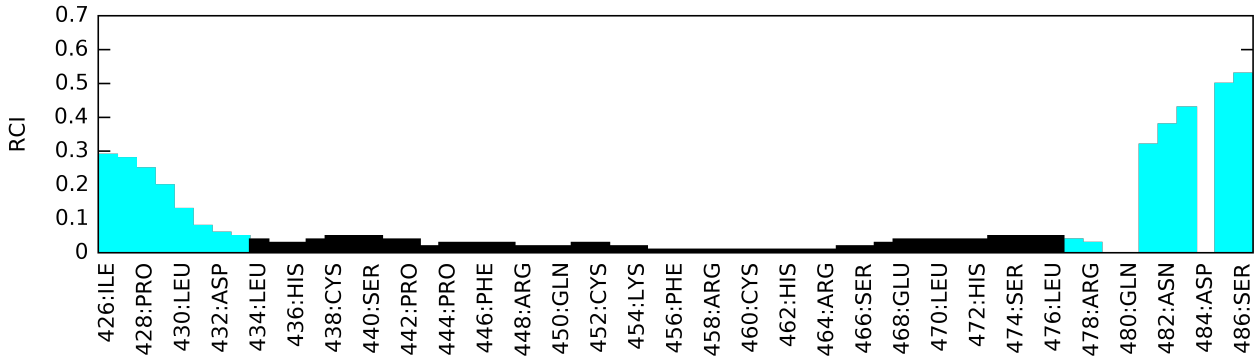
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	464	ARG	HG2	0.03	2.92 – 0.22	-5.7
1	A	464	ARG	HG3	0.03	3.00 – 0.10	-5.2
1	A	475	PRO	HB3	0.14	3.81 – 0.21	-5.2

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
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
426:ILE
428:PRO
430:LEU
432:ASP
434:LEU
436:HIS
438:CYS
440:SER
442:PRO
444:PRO
446:PHE
448:ARG
450:GLN
452:CYS
454:LYS
456:PHE
458:ARG
460:CYS
462:HIS
464:ARG
466:SER
468:GLU
470:LEU
472:HIS
474:SER
476:LEU
478:ARG
480:GLN
482:ASN
484:ASP
486:SER