Identifying RNA-binding residues based on evolutionary conserved structural and energetic features

Yao Chi Chen[1], Karen Sargsyan[1], Jon Wright[1,2], Yi-Shuian Huang[1*], and Carmay Lim[1,2,*]

[1]Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan, [2]Genomics Research Center, Academia Sinica, Taipei 115, Taiwan, and [3]Department of Chemistry, National Tsing Hua University, Hsinchu 300, Taiwan

*To whom correspondence should be addressed. E-mail: carmay@gate.sinica.edu.tw or yishuian@ibms.sinica.edu.tw

MATERIALS AND METHODS

Plasmid construction

The various CPEB3 alanine mutants were generated by QuikChange Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's protocol. The pcDNA3.1-myc-hCPEB3 plasmid was used as the PCR template and the sets of sense and antisense primers used for mutagenesis were: R427A, 5'-GAGTAGAACGCTACTCTGCAAAGGTGTTTGTTGGAGG-3' and 5'-CCTCCAACAAACACCTTTGCAGAGTAGCGTTCTACTC-3'; R449A, 5'-CACTGCCAGCTTTCGCGCGTTTGGACCTCTCGTAG-3' and 5'-CTACGAGAGGTCCAAACGCGCGAAAGCTGGCAGTG-3'; D456A, 5'-TGGACCTCTCGTAGTAGCCTGGCCTCACAAAGCT-3' and 5'-AGCTTTGTGAGGCCAGGCTACTACGAGAGGTCCA-3'; K460A, 5'-AGTAGACTGGCCTCACGCAGCTGAAAGCAAGTCTTA-3' and 5'-TAAGACTTGCTTTCAGCTGCGTGAGGCCAGTCTACT-3'; S465A, 5'-CACAAAGCTGAAAGCAAGGCTTATTTTCCTCCTAAAGGC-3' and 5'-GCCTTTAGGAGGAAAATAAGCCTTGCTTTCAGCTTTGTG-3'; F474A, 5'-CTCCTAAAGGCTATGCCGCTCTGCTGTTCCAAGAGG-3' and 5'-CCTCTTGGAACAGCAGAGCGGCATAGCCTTTAGGAG-3', R514A, 5'-GACAAGCCAGTGCAAATTGCACCATGGAACCTAAGTGA-3' and 5'-TCACTTAGGTTCCATGGTGCAATTTGCACTGGCTTGTC-3'. To construct F430A and G432A mutants, the sense primer, F430A 5'-TACTCTAGAAAGGTGGCTGTTGGAGGACTTCCT-3' or G432A 5'-TACTCTAGAAAGGTGTTTGTTGCAGGACTTCCTCCTGATA-3', along with the antisense primer 5'-CCGCTCGAGTCAGCTCCAGCGGAAC-3' were used to PCR amplify the RNA-binding domain of CPEB3. The amplified DNA fragments were digested with XbaI and XhoI and cloned to the XbaI and XhoI-linearized pcDNA3.1-myc-hCPEB3 vector. All constructs were sequenced to confirm the mutations.

Cell culture and transfection

HEK-293T cells were cultured in DMEM with 10% fetal bovine serum (Invitrogen). Transfection of plasmid DNAs was carried out using lipofectamine 2000 (Invitrogen) following the manufacture's protocol. In general, a 35-mm dish of 293T cells was transfected with the mixture of 2 µg plasmid DNA and 5 µl lipofectamine 2000 and incubated overnight for CPEB3 production.

The 35-mm dish of 293T cells expressing myc-tagged CPEB3 variants was lysed in 140 µl of 2X gel retardation buffer (20 mM Hepes, pH 7.4, 100 mM KCl, 2 mM

$MgCl_2$, 0.1% TritonX-100, 10% glycerol, 0.5 mM DTT and 1X protease inhibitor (Roche)) on ice for 15 min, followed by centrifugation at 10,000 rpm for 5 minutes. The resulting supernatants were used for RNA-binding assay. The 1904 RNA probe identified previously as the CPEB3-binding sequence (1) was labelled by *in vitro* transcription with α^{32}P-UTP.

REFERENCES

1. Huang, Y.S., Kan, M.C., Lin, C.L. and Richter, J.D. (2006) CPEB3 and CPEB4 in neurons: analysis of RNA-binding specificity and translational control of AMPA receptor GluR2 mRNA. *EMBO J*, **25**, 4865-4876.

Supplementary Table S1

Dataset I		Dataset II		RMSD (Å)	Sequence id (%)
PDB code	chain identifier	PDB code	chain identifier		
1a9n	C				
1av6	A	2gaf	A	0.68	98
1c0a	A	1eqr	A	1.64	100
1ddl	A				
1dfu*	P	1b75*	A	4.26	100
1di2	A				
1f7u	A	1bs2	A	2.22	100
1feu*	A				
1fjg*	M				
1fxl	A				
1gax	A				
1gtf	L	2zcz	A	1.78	97
1h4s	A	1h4t	A	1.28	100
1hq1	A				
1j1u	A	1u7x	A	1.5	98
1jbs	A	1aqz	A	0.71	99
1jid	A				
1k8w	A	1r3f	A	1.95	98
1m8w	A	3gvt	A	1.24	91
1mms*	A	2k3f*	A	4.17	100
1n78	A	1j09	A	1.86	100
1ooa	A	2v2t	B	8.87	99
1q2r	A	1enu	A	0.87	100
1qf6	A	1evk	A	1.32	100
1sds	C	1ra4	A	0.35	100
1ser	A				
1vq8*	1				
1vq8*	3				
1vq8*	A				
1vq8*	B				
1vq8*	C				
1vq8*	D				
1vq8*	E				
1vq8*	H				
1vq8*	J				

1vq8*	K				
1vq8*	L				
1vq8*	M				
1vq8*	N				
1vq8*	O				
1vq8*	P				
1vq8*	Q				
1vq8*	R				
1vq8*	U				
1vq8*	V				
1vq8*	W				
1vq8*	X				
1wpu	A	1wpt	A	2.23	100
1yz9	A	1rc5	A	1.37	98
1zh5	A				
1zho	A	487d	H	5.78	98
2a8v	A	1a63	A	1.9	100
2anr	A				
2asb	A	1k0r	A	1.08	98
2bgg	A	1w9h	A	0.97	98
2bu1	A	1msc	A	3.34	99
2e9t	A	1u09	A	0.78	100
2fk6	A	1y44	B	1.99	99
2fmt	A	1fmt	A	1.17	100
2gic	A	3hhw	K	1.74	99
2j01*	I				
2j01*	R	1gd8*	A	1.25	100
2q66	A	1fa0	A	3.55	96
2r8s	L	1dee	A	3.65	95
2vqe*	B				
2vqe*	C				
2vqe*	D				
2vqe*	F				
2vqe*	G	1rss*	A	2.93	100
2vqe*	H	1an7*	A	1.05	100
2vqe*	I				
2vqe*	J				
2vqe*	K				
2vqe*	N				
2vqe*	P	1emw*	A	2.42	100
2vqe*	R				

2vqe*	S	1qkf*	A	4.09	100
2vqe*	T				
2zue	A				
3gib	A	1u1t	A	0.93	93
3iev	A				

* denotes ribosomal proteins

Supplementary Table S2. Performance of DR_bind1 and BindN+ for the same number of predictions made by DR_bind1 based on the RNA-bound structures of 17 proteins whose sequences are nonhomologous to those in PRINR25 compared.

Methods	DR_bind1	BindN+
precision	**0.74**	0.47
sensitivity	0.08	0.05
specificity	1	0.99
accuracy	0.90	0.89
mcc	0.22	0.12

Supplementary Table S3. Performance of DR_bind1 based on 83 DNA-bound protein structures.

	DR_bind1[a]
TP	177
FP	84
TN	18,866
FN	1,922
Sensitivity	0.08
Specificity	1
Precision	**0.68**
Accuracy	0.90
MCC	0.22