FEBS Letters 577 (2004) 134–140 FEBS 28914

Cooperative binding of the hnRNP K three KH domains to mRNA targets

Agnieszka Paziewska[a], Lucjan S. Wyrwicz[b], Janusz M. Bujnicki[c], Karol Bomsztyk[d], Jerzy Ostrowski[a,*]

[a]*Department of Gastroenterology, Medical Center for Postgraduate Education, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, 02-781 Warsaw, Poland*
[b]*BioInfoBank Institute, 60-744 Poznan, Poland*
[c]*Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Warsaw, Poland*
[d]*Department of Medicine, University of Washington, Seattle, WA 98195, USA*

Received 24 July 2004; revised 11 August 2004; accepted 29 August 2004

Available online 13 October 2004

Edited by Gianni Cesareni

Abstract The heterogeneous nuclear ribonucleoprotein (hnRNP) K homology (KH) domain is an evolutionarily conserved module that binds short ribonucleotide sequences. KH domains most often are present in multiple copies per protein. In vitro studies of hnRNP K and other KH domain bearing proteins have yielded conflicting results regarding the relative contribution of each KH domain to the binding of target RNAs. To assess this RNA-binding we used full-length hnRNP K, its fragments and the yeast ortholog as baits in the yeast three-hybrid system. The results demonstrate that in this heterologous in vivo system, the three KH domains bind RNA synergistically and that a single KH domain, in comparison, binds RNA weakly.

Keywords: hnRNP K protein; Yeast Hek2p protein; Protein-RNA interaction; Three-hybrid system; KH domain

1. Introduction

RNA–protein interactions play a critical role in gene expression and other cellular processes [1,2]. The diversity of RNA-directed processes could not have emerged without the evolution of RNA–protein interaction selectivity. There are only a few well described RNA-binding motifs, including the ribonucleoprotein (RNP) domain, arginine-glycine-glycine (RGG) boxes, zinc fingers and the heterogeneous nuclear ribonucleoprotein (hnRNP) K homology (KH) domain. Yet, the RNA repertoire is vast [3].

The KH domains are evolutionarily ancient and are well represented in the three superkingdoms [4]. The KH domains have a β1–α1–α2–β2–β3–α3 configuration, in which three stranded antiparalleled β-sheets are supported by three α-helices. KH domains are present predominately as several copies per protein, but there are a few examples of proteins containing single domain [5].

Members of the poly(rC)-binding class of proteins, which includes hnRNP K and hnRNP E1/2, contain three asymmetrically spread KH domains where the middle domain is closer to the N-terminal KH domain [6]. Multiple studies have been done to explore RNA-binding of proteins containing KH domains. The majority of these studies were done in vitro [7–11]. Dejgaard and Leffers [7] found that the poly(rC)-binding of hnRNP K is mediated by the third KH domain (KH3), while the binding of hnRNP E1 and E2 was shown to require the first and second KH domains (KH1 and KH3, respectively). Similarly, Ito et al. [8] found that hnRNP K binds dC-rich single-stranded DNA via the carboxyl terminus containing the KH3 domain. In contrast, Siomi et al. [12] suggested that all three hnRNP K KH domains, KH1, KH2 and KH3, play a role in binding to poly(rC) under stringent conditions (1 M NaCl concentration). Furthermore, the authors stated that at physiological ionic strengths no conclusion could be made regarding the relative contribution of each of the KH domains to RNA-binding. Using systematic evolution of ligands by exponential amplification (SELEX), Thisted et al. [11] identified the target for αCP-2KL (alternatively spliced hnRNP E2) RNA sequence that consists of three C-rich patches, suggesting a three prong interaction between this protein and its cognate RNA. In the same SELEX study, the selected RNA target that bound hnRNP K consisted of a single 6–7 nt long C-rich box, suggesting that only one of the three KH domains participated in the RNA interaction. It is also plausible that each of the three KH domains binds the same consensus RNA sequence.

The frequency of a specific sequence composed of 7 nt is 1 in 16 384 ($n = 4^7$). It has been estimated that there are ~32 000 human genes [3, Table 23] with an average transcript size of ~2500 nt ([3, Table 21]). If we assume that a KH domain binds 7 nt, then this region could have as many as ~5000 mRNA targets in the human transcriptome. If a single KH domain binds 6nt rather than 7 nt sequence, there could be as many as ~20 000 mRNA targets. hnRNP K has three KH domains and if each binds a different sequence there could be three times as many transcript targets; numbers of mRNAs that could be as high as those that compose the entire transcriptome. Thus, binding of a single KH domain to a short nucleotide stretch is conceptually problematic because it raises the question of how RNA–protein interaction selectivity is achieved. Studies that searched for KH domain RNA targets using full-length hnRNP K as bait in the yeast three hybrid screen found an RNA consensus sequence composed of three C-rich boxes [13,14]. These studies suggested that all three KH domains are

* Corresponding author. Fax: +48-22-6447601.
E-mail address: jostrow@warman.com.pl (J. Ostrowski).

0014-5793/$22.00
doi:10.1016/j.febslet.2004.08.086

involved in the binding. Longer RNA sequence would not only generate higher affinity of binding, but would also restrict the potential number of targets (greater selectivity).

In an effort to resolve the controversy regarding the role of distinct KH domains in RNA-binding, we carried out multiple three hybrid screens using as a bait either full-length or deletion fragments of hnRNP K. We suggest that the KH domains bind RNA cooperatively, providing an important basis for achieving hnRNP K–RNA interaction selectivity.

2. Materials and methods

The experimental procedures, including cloning of KH domains, used standard molecular biology techniques described previously [13]. To construct the bait plasmids, the cDNAs of full-length murine hnRNP K [15] and *Saccharomyces cerevisiae* hnRNP K-like protein, Hek2p [16], were used as templates for PCR. The PCR primers were designed to contain a *Bam*HI site at the 5′ end and a *Sma*I site at the 3′. The PCR products were subcloned into pCR2.1-TOPO (Invitrogen). Then, the plasmids were digested with *Bam*HI and *Sma*I, and the gel-purified inserts were subcloned in frame with LexA into pOAA plasmid.

All point mutations were generated using the QuickChange Site-Directed Mutagenesis Kit (Stratagene).

All constructs were confirmed by sequence analysis.

Construction of hybrid RNA libraries for the yeast three-hybrid screens. Hybrid RNA libraries were generated from total RNA isolated from the human AGS gastric carcinoma cell line [13]. cDNA fragments smaller than 200 bp were purified from the gel and ligated to *Sma*I-digested and dephosphorylated pIIIMS2-2 vector. The plasmids were transformed into *Escherichia coli* DH5α strain and cells were plated on Luria–Bertani (LB)-agar. The initial titer was about 6×10^5 bacterial colonies. Then, the library was amplified once on LB-agar plates and plasmid DNA was isolated.

Yeast three-hybrid screens. Saccharomyces cerevisiae strain L40 was transformed with pOAA plasmids and the screening procedure for isolating plasmids that interacted with bait constructs was performed essentially as described previously [13]. Briefly, the yeast strains containing expressed pOAA bait constructs were transformed with human or yeast RNA library. Positive clones were initially selected for histidine prototrophy and then assayed for *lacZ* activity using a filter β-galactosidase assay with X-Gal. Next, we used (−)HIS media supplemented with 15 mM 3-aminotriazol to exclude clones with weak interactions. The quantitative measurement of β-galactosidase activity was carried out in at least three independent colonies for each hybrid RNA-expressing plasmid by using Mammalian β-Galactosidase Assay Kit (Pierce).

Statistical analysis. Differences between groups were analyzed by the Mann–Whitney test or paired two-tailed *t* test using Statistica PL software. A P value of <0.05 was considered to be statistically significant.

3. Results and discussion

hnRNP K is composed of multiple domains including three KH domains, as well as the KI region which contains five RGG boxes [17]. Thus, there are several modules within hnRNP K that could mediate RNA binding. We used full-length hnRNP K, its fragments and point mutants as baits in the yeast three hybrid system to define the contribution of the different domains to RNA binding. The library of baits is shown in Fig. 1.

The results of the screens of human hybrid RNA library are summarized in Tables 1 and 2. Altogether, we carried out 11 screens of human RNA clones generated from total cellular RNA. Of the 321 true positive clones isolated across all of the screens, 33 sets of clones were found, each representing unique transcripts; and several clones were derived from mitochondrial RNA. Eighteen unique clones were selected by full-length hnRNP K. Thirteen of those clones contained asymmetrically spaced three C-rich clusters ribonucleotide sequences with the middle box situated closer to the 5′ box. These C-rich clusters and their uneven arrangement are nearly identical to the previously defined hnRNP K-binding consensus sequence, $CCAUCN_{2-7}(A/U)CCC(A/U)N_{7-18}UCA(C/U)C$ [13,14]. Four of the targets are identical or nearly identical to the one C-rich box consensus sequence identified by SELEX, $UC_{3-4}(A/U)$

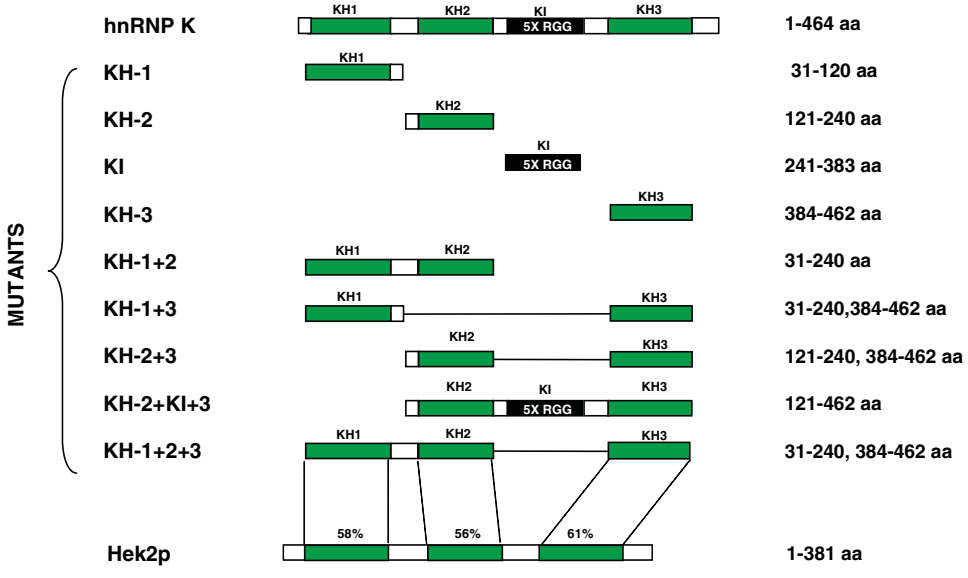


Fig. 1. Maps of full-length hnRNP K, its deletion fragments and the yeast ortholog, Hek2p, used as baits in three hybrid screens. Numbers of N-terminal and C-terminal amino acids of the fragments are indicated. KH and KI, K homology and K interactive domains, respectively. Percent conservation of KH domains across species is shown. Fragments of hnRNP K cDNA were PCR amplified using primers encoding *Bam*HI site at the 5′ end and *Sma*I site at the 3′. The PCR products were subcloned into pCR2.1-TOPO. Then, digested and gel-purified inserts corresponding to single KH domains or their different arrangements were subcloned in frame with LexA into pOAA.

Table 1
Summary of the three-hybrid screens of human RNA library using full-length hnRNP K, its mutants and the yeast ortholog Hek2 as baits

	Baits	Number of colonies screened	Total number of positive clones	Unique clones	β-Galactosidase activity	
					Range	Median
1	Full-length K	16 000	60	18	42–971	372
2	KH-1+2+3	14 000	47	13	1–246	106
3	KH-1+2	10 000	25	11	1–987	45
4	KH-1+3	12 000	17	5	28–94	86
5	KH-2+KI+KH3	12,000	31	11	35–346	92
6	KH-2+3	10 000	52	10	11–349	63
7	KH-1	16 000	3	3	10–36	28
8	KH-2	20 000	28	2	84–95	89
9	KH-3	12 000	22	5	12–157	22
10	KI (5 X RGG)	12 000	0	0	0	0
11	Hek2p	18 000	36	16	1–1172	202

(U/A) [11]. Among the 18 clones, only one did not contain a recognizable consensus sequence other than one short stretch of Us and Cs. Homopolymers of these ribonucleotides bind hnRNP K in vitro [16,18]. Thirteen clones were selected by the hnRNP K mutant containing all three KH domains (KH-1+2+3). Eleven clones each were selected by KH-1+2 and KH-2+KI+KH3, and 10 clones were selected by KH-2+3. Single domain mutants (KH-1, KH-2 or KH-3) and the KH-1+3 mutant each selected five or fewer unique clones. Out of 18 unique clones isolated with the full-length hnRNP K two were pulled out with KH-2, one with KH-3 and none with KH-1 domain (Table 2). This indicates that as a rule, in the context of full-length protein a single KH domain does not act independently. Rather the KH domains act collectively. The fact that KH-1+3 pulled out only five clones while the other two-KH domain baits, KH-1+2 and KH-2+3, isolated 10–11 clones suggests that two contiguous domains in hnRNP K may fold better for RNA-binding. With one exception (clone 8, Table 2), all the deletion mutants cloned RNA sequences nearly identical to those previously identified consensus sequences that bind hnRNP K [11,13,14]. For example, four out of five clones isolated with KH3 domain contain one or two C-rich boxes nearly identical to the previously defined SELEX consensus sequence [11]. These results suggest that these deletion mutants are overall appropriately folded.

No clones were identified using the KI region as a bait, even though this domain contains five RGG boxes [17]. Yeast hnRNP K ortholog does not have RGG boxes (Fig. 1), yet it binds some of the clones as strong as K (see below). This is in agreement with a study where mutation of the RGG boxes did not alter the binding of hnRNP K to poly(rC) in vitro [12]. Our results and those of others [12] are surprising, since RGG boxes are thought to mediate RNA-binding [19–21]. There are at least two possibilities to explain this discrepancy. (i) Ramos et al. [22] noted that the affinity of RGG peptide from Fragile X Mental Retardation Protein (FMRP) for binding to a SELEX-derived RNA is approximately two orders higher than the binding of the protein to native transcripts. Thus, it is plausible that the interaction of RGG boxes alone with RNA is of low affinity and could not be detected in this heterologous in vivo system. (ii) Darnell et al. [20] showed that all the FMRP mRNA targets, binding mediated by RGG boxes, are almost exclusively neuronally expressed. It is conceivable that the AGS gastric carcinoma library used in our screen may not have hnRNP K RGG boxes mRNA targets.

One of the advantages, and there are several, offered by the yeast three-hybrid system is that using β-galactosidase activity [23] provides a quantitative measure to assess the relative affinity of protein–RNA interactions. Although several identical RNA sequences were bound both by full-length hnRNP K and its fragments, the binding activities were significantly different. The median activity of RNA interaction with full-length hnRNP K was 372 units of β-galactosidase activity, whereas for all deletion mutants the median activity ranged between 22 and 106 units (Table 1). These results demonstrate that the affinity of interaction is the highest when all three KH domains participate in RNA binding.

The flexibility of the highly conserved GxxG loop connecting helix α1 and α2 is important for RNA binding [4,24–26]. Moreover, glycine residues have a steric advantage because their small size permits closer contact between protein and RNA, allowing main chain hydrogen bonding. Substitution with a large amino acid such as Tyr would sterically prevent such an interaction (Fig. 2). To determine if mutations in the GKGG motif of hnRNP K's KH3 domain have an effect on the RNA binding affinity, the double point mutant of Gly^{404} and Gly^{407} was generated by substituting these residues with Tyr, G404Y/G407Y. Using the G404Y/407Y mutant as a bait 42 positive clones, 26 representing unique transcripts, were isolated from 20 000 independent clones. Among them there were 13 that we also isolated with the wild type hnRNP K. Quantitative assay of β-galactosidase activity for these 13 clones revealed higher binding affinity of the wild-type hnRNP K compared to the hnRNP K lacking the GxxG motif within the KH3 domain ($n = 13$, $P < 0.05$). The decreased binding affinity of the G404Y/407Y mutant to two RNA targets is illustrated in Fig. 4. Although the mutation in the KH3 domain GxxG loop decreased the affinity, it did not block the binding. This is in agreement with our other results showing that all three KH domains contribute to the RNA binding. This minimizes the effect of mutating residues within any single KH domain.

The effect of these mutations on the overall protein structure cannot be ruled out. It remains possible that the decrease of the binding affinity is caused not only by introducing a steric hindrance in Gly → Tyr mutations or removing important contacts in Lys → Ala mutations, but also by the mild destabilization of the backbone of the RNA-binding site. For RNA-binding, the overall protein structural framework has to be correct. Because the hnRNP K mutants retain the ability to

Table 2
RNA sequences isolated in the three hybrid screen with hnRNP K, its deletion fragments and its yeast ortholog Hek2p

Clone	K	Hek2p	KH-1+2+3	KH-2+KI+3	KH-2+3	KH-1+2	KH-1+3	KH-1	KH-2	KH-3	Sequence 5'=>3'
23			x		x	x	x				CUUGCCAGGCUCUGAAUGGUCUGACUCUGCU**UCCCA**$_{ACCCAUUGUA}$**UCCCC**$_{CA}$$_{GACCAUUUAUACACUAGG}$
20	x									x	CUUAAA$_{C}$**CCCCU**$_{UAUUUCUAGGACUAUGAGAAUCGAA}$**CCCA**$_{UCCCUGAGAAUCCAAAAUUCUCCGUGCCACCUAUCACACCCCAUCCUAAAGUAAGGUCAG}
17		x								x	CUGUAUUUGCCAUCAUAGGAGGCUUCAUUCACUGAUU**UCCCCU**$_{AUUCUCAGAGCUACA}$**CCCU**$_{AGACCAAACCUACGCCAAAAUCCAUUUCACUAUCAUAUUC}$ $_{AUCGGCGUAAAUCUAACUUUCUUCCCACAACACUUUCUCGG}$
18	x	x		x	x	x			x		CUUCCUACACUAUUAAAGUUUACCACAA**CCA**$_{C}$**C**$_{AC}$**CCCA**$_{UCAUACGUCUU}$**UCACC**$_{CACAGCACCAAUCCUACG}$
Hk18								x		x	CUGACCAGGGAACUCCUUCUGCAGGCGGCUUCUCUGUG**U**$_{A}$**CCC**$_{AGCCAUGCCUUGCGCACUCCUGCCUCGGCAGUUACGCUUUGAGG}$
1						x		x			CC**UCCC**$_{A}$**A**$_{AGUGCUGGGAUUACAGAGGAAUUCCAGACUAGCCUGG}$
6a			x		x		x			x	CCCACCACAACCUCUGCCUCCUGGGUUCAAGUCAAGCGA**UUC**$_{U}$**CCU**$_{GCCUCAGCCUUCCGAGUAG}$
6b	x		x								CUCUGCCUCCUGGGUUCAUGCCAUUCUCCUGCCUCAGCC**UCCC**$_{G}$**A**$_{GUAGGUGGGACUACAGGCGCCCGCCACCACGCCCAG}$
8		x			x			x		x	CCGUCUACUGCACAUGCGCGGGUCCGCGGGCAGCCGG
9	x		x	x	x		x		x		CCCCUGC**CCA**$_{C}$**C**$_{UCAAC}$**UCC**$_{U}$**CU**$_{CUCCCCACCCCCU}$**UCAC**$_{U}$**C**$_{ACUUCCUCCAAG}$
30	x		x		x		x				CUGUGUUCCAAUGAAUCCUACCUCUUGCCCAG**UCCCA**$_{GGCAGAGUAAGCAGGG}$
33					x		x				CCCUGCUCACCACCCC**U**$_{U}$**CCCCU**$_{UACCUGCACCACAGGCUCUGGCUGACUCUGCCCAGG}$
47	x					x					CAGGUUAUUGAUAGGAAUUUGGGCAGGGGAAUGGAUAGAGCAGCCCGUGAUAAGGCAACCUGGCUUGG
10		x									CUGUCUCANAAGCAU**CC**$_{C}$**U**$_{UGCAACC}$**ACC**$_{A}$**A**$_{ACAGUGC}$**CACC**$_{AUAGCACCUUUCCAUGUCACAUCUGUGUCCACUCACUUCUCAGAACCAAAGGG}
14		x									CCCAACCCGU**CAUC**$_{UACUC}$**U**$_{A}$**CCA**$_{UCUUUGCAGGCACAC}$**UCAUC**$_{ACAGCGCUAA}$
19		x									CCUGACCUCAAUCUCUUGUGCUGCAGAGUGAGGCUGUGAAUCUGCCUCCUGCUGGCUCUGGGCACAGUCUCUACACUGCCUGGGCACUGCCCCUGGCUUCCAACUUU $_{UCUCAGAUGCCCUUCUUUGGGCUCCAGGGUUCUCGG}$
21	x	x			x	x					CUGCCCAUUCUUCCAUGCAUCCCUCCAUCCAUUCCUCCAUGCACCCACCCAUCCAUUUCUCCAUGC $_{AUGCAU}$**CCAUC**$_{CUUCCA}$**UCC**$_{A}$**U**$_{UCAUCCAUCCAUC}$**CAUC**$_{CAUCCAUGG}$
22	x	x		x		x					CCAUCAAUUUCAUCACAACAAUUAUCAAUAUAAAACCCCCUG**CCAU**$_{A}$**ACCCA**$_{AUACCAAACGCCCC}$**UC**$_{U}$**UC**$_{GUCUGAUCCGUCCUAAUCACAGCAGUCCU}$ $_{ACUUCUCCUAUCUCUCCCAGUCCUAG}$
24	x		x	x		x					CCAGCACCACCGGUCAGAGACUGU**CC**$_{C}$**U**$_{UUCUCCC}$**CCCA**$_{CCGCUCAGAGCCCUC}$**UC**$_{U}$**CC**$_{UACCUGUUCCCCCAUGACUGCUCCACUUUCUGUGUGGCAGGU}$ $_{GCUCCUCCCUGGG}$
25	x		x			x					CCCGUAUUUACCCUAUAGCACC**CC**$_{C}$**UC**$_{UACC}$**CCCU**$_{CUAGAGCC}$**CAC**$_{UGUAAAG}$
26a	x		x	x		x					CUCC**CCAUC**$_{CAU}$**UCC**$_{A}$**CU**$_{AAGAGC}$**CACC**$_{UCCAACACUCGAUAAAAUCCCUACAUUCACCAUACUUCAAGUAUGUGUGUGACCUGAUUCCUCCUGGACCU}$ $_{UGGUACCAAGAG}$
26b	x			x							CUCC**CCAUC**$_{CAU}$**UCC**$_{A}$**CU**$_{AAGAGC}$**CACC**$_{UCCAACACUCGAUAAAAUCCCUACAUUCACCAUACUUCAAGUAUGUGUGUGACCUGAUUCCUCCUGGACCU}$ $_{UGGUACCAAGGGGGCAGGGUGCAAAAGGCUAUUACCCUGACCCUCCACUGAG}$
27a	x	x		x		x					CCCA**CCAU**$_{AAUUACC}$**CCCA**$_{UACUCCUUACACUAU}$**UC**$_{C}$**UC**$_{AUCACCCAACUAAAAAU}$
27b			x								CCCA**CCAU**$_{AAUUACC}$**CCCA**$_{UACUCCUUACACUAU}$**UC**$_{C}$**UC**$_{AUCACCCAACUAAAAAUACCAUACU}$
28	x		x			x					CCCCCAACAGGCAUCACCCCGCUAAAUC**CC**$_{C}$**U**$_{AGAAG}$**UCCCA**$_{CUCCUAAACA}$**CAUC**$_{CGUAUUACUCGCAUCAGGAGUAUCAAUCACCUGAGA}$
41			x		x						CCGAGCCGGA**CCCCA**$_{CCCCAC}$**UCCCC**$_{CCCAG}$
54				x	x						CUAUCACCACUCUGUUCGCAGCAGUCUGCG**CCCU**$_{UACACAAAAUGACAUCAAAAAAAUCGUAG}$
56kh			x								CUGUGCACUUGCCACACUUACCCCG**UCCCA**$_{CCUUCUACACAG}$
49	x										AAGCCAGGUUGCCUUAUCACGGGCUGCUCUAU**CCAU**$_{UCCCCUG}$**CCCA**$_{AAUUCCUAUCAA}$**U**$_{A}$**ACC**$_{U}$
52	x										CCCGUCACUUGUUCGUGCGCCAGCACGCCAGGCACGUCGAGCACCAGCCCGGCAGGAACGCCGGCUUCGAUCAGCAGCGGC**UCCCA**$_{UUCGGCGG}$
K15				x							CUCUGGGUCCCUCUUUAAUCUCCUUC**CC**$_{C}$**U**$_{UAUCUUC}$**UCCCU**$_{UCUCAUCACC}$**UC**$_{C}$**UC**$_{CUUGGGGGAAAGG}$
Hk12	x	x									CUAACAUGACUAACA**CC**$_{C}$**U**$_{UAAU}$**UCC**$_{A}$**U**$_{CCACCCUCCUC}$**UC**$_{C}$**CU**$_{AGAGAGAGAGA}$
32		x									CCGGACAUGUCUCACUUUUAAUCUCU**CCA**$_{C}$**C**$_{AGC}$**UCC**$_{U}$**CA**$_{CGGUCUGACU}$**UCA**$_{A}$**C**$_{GAAAUUAAAACGCUGCGUCCUAACGCACAUAAUGUGCCCGUAG}$
35		x		x							CCACAU**UCCCA**$_{CACGGCAACAAUUAAACCGCAG}$**UCCCCA**$_{CUACUCCCUAUUACAUUUCCAACACUAAGACCAAGAGA}$
38		x									CCCAUACUGCAAGAUGUACGCCAUCACACUUUAACGUGCUCUGACUACUUGUAAGUAAAUGGUUUCAGAAUCUAUUUCAC**UCCCCU**$_{CUCGGGGUUCUUUUCACC}$ $_{UU}$**UCCC**$_{U}$**CA**$_{CGGUACUAGUUCACUAUCGGUGUCUGAUUAGU}$
39		x		x							CUCUGGGUCCCUCUUUAAUCUCCUUC**CC**$_{C}$**U**$_{UAUCUUC}$**UCCCU**$_{UCUCAUCACC}$**UC**$_{C}$**UC**$_{CUUGGGGGAAAGG}$
40		x									CUGUCUCAGAAGCAUCCCUUGCAACCACCAAACAGUG**CCA**$_{C}$**C**$_{AUAGC}$**ACC**$_{U}$**U**$_{UCCAUG}$**UCAC**$_{AUCUGUGUCCACUCACUUCUCAGAACCAAAGAGAGA}$
Hk3	x	x									CUACGCCCUGAUCGGCGCACUGCGAGCAGUAGCCCAAACAAUCUCAUAUGAAGUCACCCUAGCCAUCAUUCUACUAUCAACAUUACUAAUAAGUGGCUCCU $_{UUAA}$**C**$_{C}$**U**$_{CUCC}$**ACCCU**$_{UAUCACAACACAAGAA}$**CACC**$_{UCUGAUUACUCCUGCCAUCAUGACCCUUGG}$
	18	16	13	11	10	11	5	3	2	5	

Sequences similar to the previously derived 3′ box C-rich hnRNP K-binding motif [13,14] are shown in large red font, and those similar to the 1 box SELEX results [11] in large blue font. Mismatched ribonucleotides are shown in small green font. (For interpretation of the references to colour in this table legend, the reader is referred to the web version of this article.)

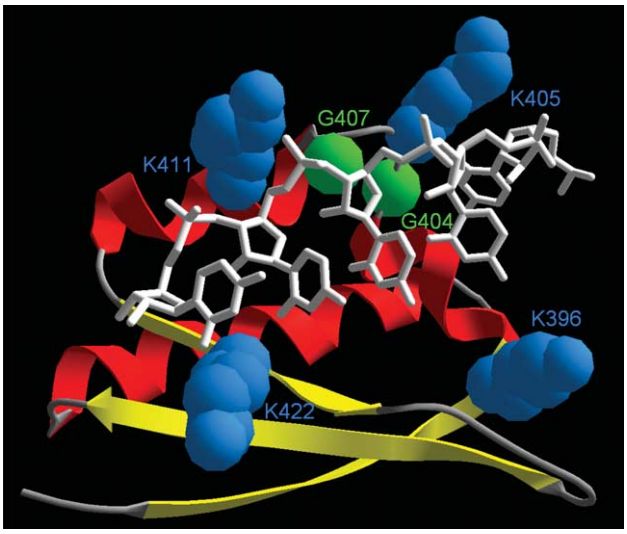
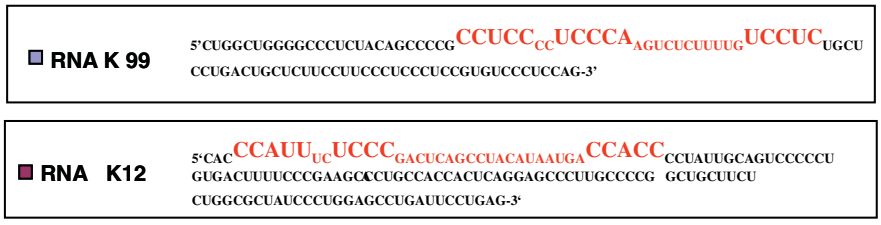


Fig. 2. Localization of predicted RNA-binding residues, based on the solution structure of the KH3-nucleic acid complex [30]. α-helices and β-strands are shown as red spirals and yellow arrows, respectively. Gly residues substituted with Tyr are shown as green balls, the side chains of Lys residues replaced by Ala are shown in blue and labeled. In the G404Y/G407Y mutant, the Gly ⇒ Tyr substitutions inflict steric clashes of side-chains with the "wild type" backbone and therefore most likely induce conformational change of the RNA-binding loop. Substitution of the Lys residues removes the positive charge and potential sites of protein-RNA interactions.

bind RNA, we strongly believe that the effect of mutations is local without significant misfolding of the RNA-binding site or that of the KH domain as a whole. Nonetheless, the ultimate evidence for this assertion can be provided only by a combination of biophysical and high-resolution structural analyses.

To further examine the importance of KH domain cooperative binding to RNA, we used a β-galactosidase assay to compare the affinity of interaction between different hnRNP K deletion mutants and two cognate RNA sequences (RNA K12 and RNA K99) that had been selected in the yeast three-hybrid screen with hnRNP K as bait [14]. Both RNA sequences contained the consensus hnRNP K-binding motif consisting of three C-rich boxes [13,14]. The full-length hnRNP K showed the highest RNA binding activity. The binding activity of the protein bait composed of the three KH domains but lacking the KI domain was half of that seen with the full-length hnRNP K bait (Fig. 3). A contribution of the KI region may reflect either the importance of the spacer between the KH-2 and KH-3 domains or the RRG may be contributing to the RNA binding. The latter is less likely, since the KI region bait did not isolate any clones (Table 1). Another possibility is incomplete folding of the mutated/truncated proteins.

In agreement with the data shown in the tables and Fig. 3, single KH domains showed very low levels of RNA binding activity. These measurements demonstrate that the RNA-binding activity of the full-length hnRNP K is greater than the sum of the activities generated by individual RNA-binding



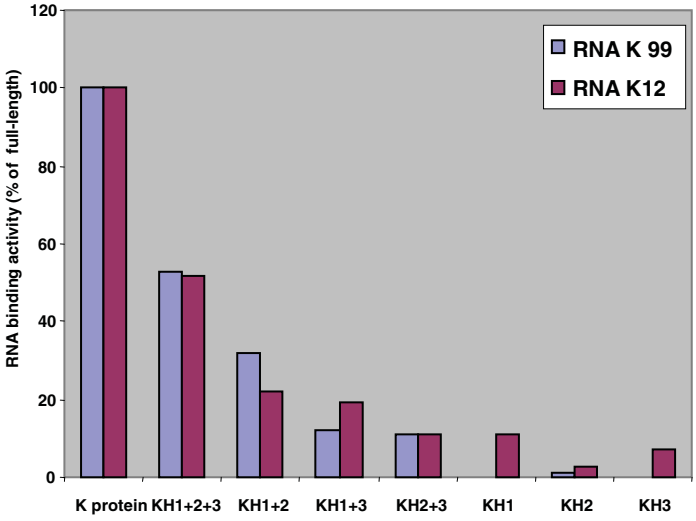


Fig. 3. Quantitative measurement of RNA-binding activity of full-length hnRNP K and its K deletion mutants to two RNAs. β-Galactosidase activity was determined in three independent colonies for each hybrid RNA-expressing plasmid and all assays were repeated in quadruplicate. Results are expressed as the percentage of RNA binding activity measured with the full-length hnRNP K bait and represent means. The two RNAs used are shown in the boxes above. The three asymmetrically spaced C-rich ribonuclotide boxes compose a stretch that matches the empirically derived K binding sequence [13,14] (shown in red).

domains, indicating a synergism between the three KH domains when binding these target RNAs. This analysis also suggests that relative arrangement of the KH domains (compare full-length K with KH-1+2+3) is critical to achieve maximal binding activity.

Mutation of K^{396}, K^{405}, $K^{411}$and K^{422} to Ala removes the side chains of Lys residues and eliminates the positively charged ε-amino groups, which are potentially interacting with nucleotides (Fig. 2). Next, we tested if these mutations will affect RNA-binding. As shown in Fig. 4, all K ⇒ A point mutations impair the binding affinity to two selected RNA sequences to the same extent as mutations in GKGG motif within KH3 domain (Fig. 4). However, these mutations did not completely block hnRNP K binding to RNA.

Considering that more than a billion years separate yeast and mammals [27], the primary structure of the three corresponding KH domains in these species is remarkably conserved [17]. In fact, the similarity of the respective KH domains across species is greater than that between KH domains in the same protein [4]. Moreover, hnRNP K and its yeast orthologs exhibit similar binding patterns to ribonucleotide homopolymers in vitro, suggesting that the RNA-binding specificity was also conserved [16].

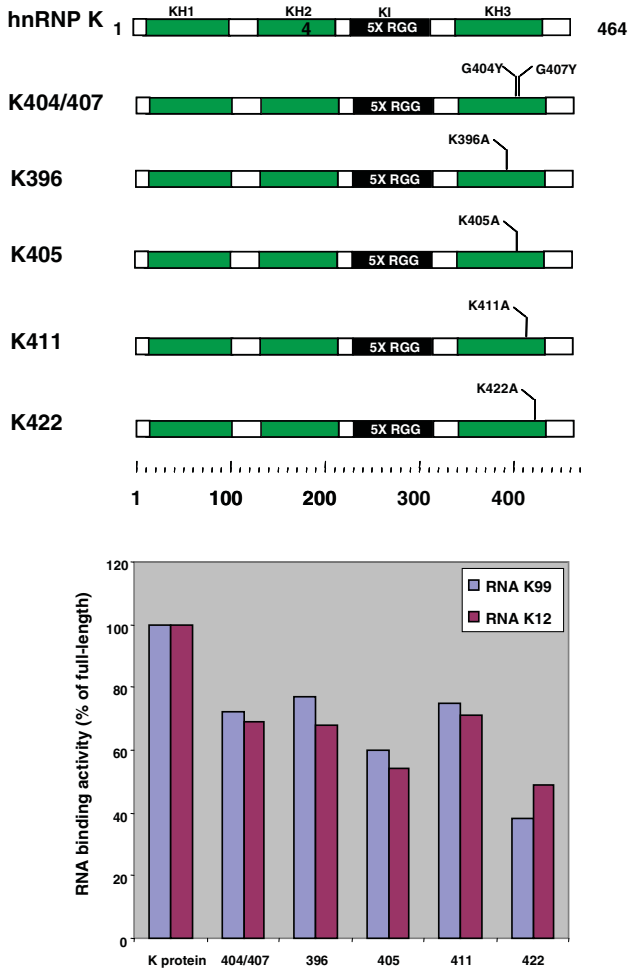


Fig. 4. Quantitative measurement of RNA binding activity of full-length hnRN K and two hnRNP K point mutants to RNA targets. RNA sequences as in Fig. 3. β-Galactosidase activity was determined in three independent colonies for each hybrid RNA-expressing plasmid and all assays were repeated in quadruplicate. Results are expressed as the percentage of RNA binding activity measured with full-length hnRNP K bait and represent means.

The homology modeling tools accessed through the Protein Structure Prediction MetaServer at BioInfo.PL [28] revealed that three regions similar to KH domain can create conserved structure of this fold (data not shown). The sequence similarity between Hek2p KH domains and KH3 domain from hnRNP K is high, with *E* values of 3 e − 17, 3 e − 14, and 2 e − 18 for KH1, KH2 and KH3 domains, respectively (PDB-Blast through the MetaServer).

The 3D-Jury system [29] was used to compare the similarity of predicted models with empirically defined hnRNP K KH3 domain structure. The models for all three KH domains achieved a score >50 in the 3D-Jury system. This analysis revealed that Hek2p KH domains create a highly conserved KH fold (http://lucjan.bioinfo.pl/supplemental/paziewska 2004). This analysis suggests that hnRNP K and its yeast othologs can bind similar RNA sequences via KH domains.

To test this, we used Hek2p protein as a bait and screened 18 000 clones from the human RNA library (Tables 1 and 2). Of 36 positive clones, 16 represented unique transcripts; nine of them were nuclear and seven were mitochondrion encoded. Six unique RNA sequences were also found in the screens using hnRNP K as bait. The binding activity of Hek2p with majority of the human RNAs was significantly lower compared to the activity found for hnRNP K ($P < 0.001$), but the binding activity of the same six clones found both by hnRNP K and Hek2p did not differ significantly (Table 2, shaded boxes). All of the six clones contained the asymmetrically spaced three C-rich boxes sequence that represents the RNA consensus sequence that binds hnRNP K [13,14]. Hek2p does not contain RGG boxes [17] (Fig. 1). In agreement with the failure to isolate any clones with the KI domain, these results confirm that the RGG boxes play little or no role in the binding of hnRNP K to these RNA targets. Three of the Hekp2 target RNA clones had two C-rich boxes, identical or nearly identical to the sequence isolated by SELEX [11], and two clones had no recognizable sequences for binding the yeast protein.

To further assess the specificity of the interactions between HnRNP K and RNAs, we have tested whether two selected RNA sequences (RNA K12 and RNA K99) can be bound by the yeast ortholog of the mammalian hnRNP K, Hek2p. As shown in Fig. 5, RNA-binding activity of Hek2p was lower than the RNA-binding activity of hnRNP K, but for the RNA K99 it was still strong. These results suggest that the synergistic RNA binding by the KH domains is evolutionarily conserved and that the tertiary structure of hnRNP K appears to be conserved. Cooperation between the three KH domains that generate such high affinity binding is likely to have played an important evolutionary role in conservation of not only the structure of each one of the KH domains but also the inter-domain fit.

We are suggesting that the synergistic binding of hnRNP K to RNA reflects a three prong interaction where each KH domain binds one of the C-rich boxes [14]. In comparison, a single KH domain binding to a short nucleotide sequence [30] does not provide sufficiently high RNA-binding affinity (Tables 1 and 2, Fig. 3) and limits the potential for in vivo RNA–protein interactions. It is plausible that multiple KH

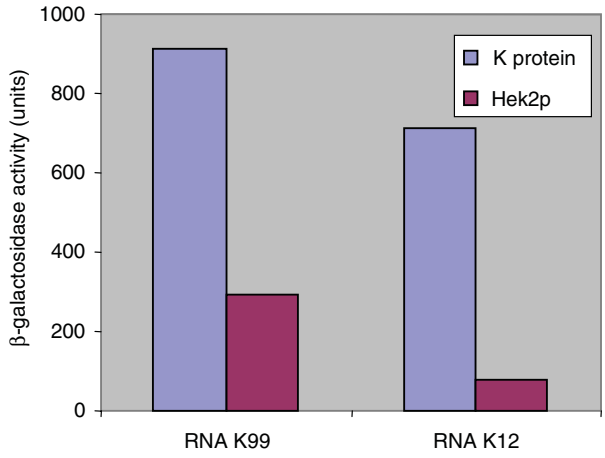


Fig. 5. Comparison of RNA-binding activity of full-length hnRNP K and yeast Hek2p to two selected RNAs. RNA sequences as in Fig. 3. The values of β-galactosidase activity are averages for four clones, each assayed in quadruplicate.

domains not only increase affinity but also select for longer nucleotide stretches, key determinants of selectivity. Although they represent a small fraction of KH domain containing proteins, there are some that contain only a single KH domain, for example Sam68, a member of the STAR/GSG family of RNA-binding proteins [31]. How is this solo KH domain-to-RNA-binding affinity and selectivity maintained? It turns out that in these proteins the KH domains are flanked by another domain that participates in RNA-binding [26]. Structural study of splicing factor 1 (SF1), a member of STAR/GSG family, has shown that the KH-QUA2 region defines an enlarged KH domain fold that binds the target RNA sequence [32]. This explains why the single Sam68 KH domain is necessary for RNA-binding, but it is not sufficient [33,34]. Moreover, Sam68 forms dimers [35]. It has been shown that dimerization of another member of this family of factors, GLD-1, is critical for high affinity binding to RNA [36]. Thus, the enlarged fold of the KH domain and multimerization of Sam68 would yield higher binding affinity to a longer RNA sequence compared to that provided by just one KH domain.

Structural studies of single KH domains complexed with DNA and RNA have provided great insight into the molecular basis of their sequence-specific binding [25,30,32,37–39]. The present study is evidence that the three KH domains of hnRNP K bind RNA cooperatively. Structural studies of the full-length protein bound to a natural target RNA are thus needed to define the molecular basis for the RNA-binding synergism exhibited by three KH domains. The molecular basis for this cooperation is likely to be conserved from yeast to man.

Acknowledgements: This work was supported by a grant from the Polish Committee for Scientific Research (KBN 4PO5A 084 19) and CMKP Grant (501-1-2-08-04/01) to J.O. J.M.B. was supported by the EHMBO/HHMI Young Investigator Programme and by a Fellowship for Young Scientists from the Foundation for Polish Science. K.B. work is supported by the NIH and the Juvenile Diabetes Research Foundation. We thank Oleg Denisenko, Daniel S. Schullery and Sanjeev Jain for critical reading of the manuscript.

References

[1] Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002) Nat. Rev. Mol. Cell Biol. 3, 195–205.
[2] Mattick, J.S. (2004) Nat. Rev. Genet. 5, 316–323.
[3] Consortium, I.H.G.S. (2001) Nature 409, 860–921.
[4] Grishin, N.V. (2001) Nucleic. Acids Res. 29, 638–643.
[5] Gibson, T.J., Thompson, J.D. and Heringa, J. (1993) FEBS Lett. 324, 361–366.
[6] Makeyev, A.V. and Liebhaber, S.A. (2002) RNA 8, 265–278.
[7] Dejgaard, K. and Leffers, H. (1996) Eur. J. Biochem. 241, 425–431.
[8] Ito, K., Sato, K. and Endo, H. (1994) Nucleic Acids Res. 22, 53–58.
[9] Siomi, H., Siomi, M.C., Nussbaum, R. and Dreyfuss, G. (1993) Cell 74, 291–298.
[10] Siomi, H., Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1993) Nucleic. Acids Res. 21, 1193–1198.
[11] Thisted, T., Lyakhov, D.L. and Liebhaber, S.A. (2001) J. Biol. Chem. 276, 17484–17496.
[12] Siomi, H., Choi, M., Siomi, M.C., Nussbaum, R. and Dreyfuss, G. (1994) Cell 77, 33–39.
[13] Ostrowski, J., Wyrwicz, L., Rychlewski, L. and Bomsztyk, K. (2002) J. Biol. Chem. 277, 6303–6310.
[14] Klimek-Tomczak, K., Wyrwicz, L.S., Jain, S., Bomsztyk, K. and Ostrowski, J. (2004) J. Mol. Biol. (in press).
[15] Ostrowski, J., Van Seuningen, I., Seger, R., Rouch, C.T., Sleath, P.R., McMullen, B.A. and Bomsztyk, K. (1994) J. Biol. Chem. 269, 17626–17634.
[16] Denisenko, O. and Bomsztyk, K. (2002) Mol. Cell Biol. 22, 286–297.
[17] Bomsztyk, K., Denisenko, O. and Ostrowski, J. (2004) BioEssays 26, 629–638.
[18] Van Seuningen, I., Ostrowski, J. and Bomsztyk, K. (1995) Biochemistry 34, 5644–5650.
[19] Kiledjian, M. and Dreyfuss, G. (1992) EMBO J. 11, 2655–2664.
[20] Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T. and Darnell, R.B. (2001) Cell 107, 489–499.
[21] Adinolfi, S., Bagni, C., Musco, G., Gibson, T., Mazzarella, L. and Pastore, A. (1999) RNA 5, 1248–1258.
[22] Ramos, A., Hollingworth, D. and Pastore, A. (2003) RNA 9, 1198–1207.
[23] Jaeger, S., Eriani, G. and Martin, F. (2004) FEBS Lett. 556, 7–12.
[24] Baber, J.L., Levens, D., Libutti, D. and Tjandra, N. (2000) Biochemistry 39, 6022–6032.
[25] Baber, J.L., Libutti, D., Levens, D. and Tjandra, N. (1999) J. Mol. Biol. 289, 949–962.
[26] Adinolfi, S., Bagni, C., Castiglione Morelli, M.A., Fraternali, F., Musco, G. and Pastore, A. (1999) Biopolymers 51, 153–164.
[27] Hardison, R.C. (2003) PLoS Biol. 1, E58.
[28] Bujnicki, J.M., Elofsson, A., Fischer, D. and Rychlewski, L. (2001) Bioinformatics 17, 750–751.
[29] Ginalski, K., Elofsson, A., Fischer, D. and Rychlewski, L. (2003) Bioinformatics 19, 1015–1018.
[30] Braddock, D.T., Baber, J.L., Levens, D. and Clore, G.M. (2002) EMBO J. 21, 3476–3485.
[31] Lukong, K.E. and Richard, S. (2003) Biochim. Biophys. Acta 1653, 73–86.
[32] Liu, Z. et al. (2001) Science 294, 1098–1102.
[33] Lin, Q., Taylor, S.J. and Shalloway, D. (1997) J. Biol. Chem. 272, 27274–27280.
[34] Barlat, I., Maurier, F., Duchesne, M., Guitard, E., Tocque, B. and Schweighoffer, F. (1997) J. Biol. Chem. 272, 3129–3132.
[35] Chen, T., Damaj, B.B., Herrera, C., Lasko, P. and Richard, S. (1997) Mol. Cell. Biol. 17, 5707–5718.
[36] Ryder, S.P., Frater, L.A., Abramovitz, D.L., Goodwin, E.B. and Williamson, J.R. (2004) Nat. Struct. Mol. Biol. 11, 20–28.
[37] Backe, P.H., Ravelli, R.B., Garman, E. and Cusack, S. (2004) Acta Crystallogr. D 60, 784–787.
[38] Lewis, H.A., Musunuru, K., Jensen, K.B., Edo, C., Chen, H., Darnell, R.B. and Burley, S.K. (2000) Cell 100, 323–332.
[39] Braddock, D.T., Louis, J.M., Baber, J.L., Levens, D. and Clore, G.M. (2002) Nature 415, 1051–1056.