Arch Virol (1996) 141: 685–699

Archives of Virology

RNA binding properties of core protein of the flavivirus Kunjin

A. A. Khromykh and **E. G. Westaway**

Sir Albert Sakzewski Virus Research Centre, Royal Children's Hospital, Brisbane, Australia

Accepted October 13, 1995

Summary. Kunjin virus (KUN) C is a typical flavivirus core protein which is truncated in vivo to a mature form of 105 residues enriched in lysine and arginine. In order to study the possible association of KUN C with RNA in vitro, we prepared several recombinant C proteins with specific deletions, each fused at the amino-terminus to glutathione-S-transferase (GST) and expressed in *E. coli*. They were reacted with KUN RNA probes transcribed in vitro from cDNA representing the 5′ untranslated region (5′ UTR, 93 of 96 nucleotides), the 3′ UTR (624 nucleotides), and the 5′ UTR plus most of the C coding region (5′ core, 440 nucleotides). Fusion protein C107 (incorporating mature C) bound strongly to all KUN RNA probes with apparent specificity, being completely resistant to inhibition by 800 mM NaCl, and to competition by a large excess of tRNA. In reactions with labelled KUN RNA probes putative binding sites were identified in the isolated amino-terminal (32 residues) and carboxy-terminal (26 residues) basic amino acid domains; this binding was strongly competed by unlabelled KUN UTR probes but weakly or not at all by tRNA. These small domains probably acted co-operatively in binding of mature C to KUN RNA probes. The KUN RNA-core protein binding reactions are similar to those reported with other viral coat or capsid proteins and viral RNAs.

Introduction

There are relatively few reports of in vitro binding of coat or core proteins to genomic sequences of positive-stranded RNA viruses; a common but not universal feature is the high proportion of lysine/arginine residues associated with these binding reactions. These residues impart a positive charge to the core protein facilitating its binding to (negatively charged) RNA during the process of virus assembly. Both specific and nonspecific binding reactions have been reported. Examples include alphaviruses [7, 8], mouse hepatitis virus [13, 15] and several plant viruses [6, 18, 24]. Stem loop structures in the binding region of RNA also have been reported. A predicted binding site of the nucleocapsid protein of

mouse hepatitis coronavirus corresponds to sequences within a potential hairpin loop involving nucleotides 52 through 76 (in the leader region) of genomic RNA; it was proposed that such binding may play a role in regulation of transcription [22]. Turnip crinkle virus coat protein binds to two sites in genomic RNA able to associate in a stem loop structure surrounding the termination codon; it was proposed that this structure is involved in RNA replication [24]. A stem loop structure between nucleotides 11 and 127 from the 3′ end of alfalfa mosaic virus RNA3 binds viral coat protein, thus possibly contributing to initiation of infection and to assembly of virus particles [18].

The association of the flavivirus core protein with virion RNA in intracellular and extracellular virus particles is well documented (for review see [3]). However, how and where this association occurs in cells is not understood. Nucleocapsids have not been isolated from flavivirus-infected cells [28], and core-like particles have been observed by electron microscopy only in mosquito cells infected with the PR-159 strain of dengue-2 virus (but not with the NGC strain or with other flavivirus species; reviewed by Hase et al. [11]). The flavivirus genomic RNA of about 11 kilobases is flanked by about 100 to 130 nucleotides in the 5′ untranslated region (UTR) and 400 to 600 nucleotides in the 3′ UTR (see [3, 12]). As noted above, binding of coat or capsid proteins to stem loops in the UTRs of some positive-stranded RNA viruses is possibly involved in encapsidation and/or RNA synthesis; for flaviviruses similar interactions with core protein may involve the folded stem loop structures proposed in approximately the first 80 nucleotides of the 5′ UTR and the last 80 nucleotides of the 3′ UTR [1, 2, 10, 14, 19, 26]. Before any sequence data were available, a role was proposed for flavivirus C in control of the switch of RNA plus strand as a template for transcription to translation by attachment of C to the 3′ end, hence blocking polymerase attachment for initiation of copying from plus strand templates [27]. Flavivirus core protein C is relatively small (about 13 kilodaltons) and about 21% of the amino acids are lysine or arginine (see [5, 19]). The carboxy terminal sequence of 18 amino acid residues of KUN C protein provides the hydrophobic signal sequence preceding prM that is cleaved posttranslationally, reducing C to its mature form of 105 residues [21].

The purpose of the present work was to investigate associations of C with RNA which could be observed in vitro, as part of a continuing study of events during flavivirus replication. Initially, we have analysed the binding of KUN C representing the mature form (i.e. with the carboxy-terminal 16 amino acids deleted), and with additional deletions in regions of basic amino acids, to KUN RNA probes representing the 5′ and 3′ UTRs and to nonspecific RNA. The UTR regions were chosen as RNA probes in this initial study because of their possible involvement in replication or assembly, and because of reports noted above of binding of coat or capsid proteins to 3′ UTR regions of several plant viruses [6, 18, 24] or to the 5′ leader sequence of a coronavirus [15, 22]. For this study we exploited the use of glutathione-S-transferase (GST)-core fusion proteins attached to an insoluble matrix (Glutathione Sepharose 4B) for rapid RNA-protein binding assays, and extended some of these results by mobility shift

assays [17] which were possible because of the relatively small size of the RNA probes. KUN UTR probes bound strongly to fusion proteins incorporating full length or mature C, and to the isolated amino-terminal or carboxy-terminal domains of mature C.

Materials and methods

Construction of plasmids, expression and purification of recombinant proteins

The 460 base pair fragment representing the sequence coding for the first 107 amino acids of KUN C protein, plus the complete 5′ UTR and an upstream SP6 promoter, were PCR amplified (using appropriate primers) from cDNA synthesized from KUN virion RNA ligated (5′ end to 3′ end) as described previously [12]. The PCR fragment was then cloned into pBluescript II SK vector (Stratagene); from the resulting plasmid named pBSSKcore, a cDNA fragment containing the sequence coding for the first 107 amino acids of C protein and the preceding 7 nucleotides of the 5′ UTR was excised by digestion with *Bgl*II and *Bam*HI, and subcloned into *Bam*HI digested pGEX-3X expression vector for expression from a single open reading frame of C protein fused at its amino terminus to GST (Fig. 1a, C107). Constructs expressing GST fused to the first 32 or 83 amino acids of C protein (C32 or C83 respectively), were obtained by similar methodology. Constructs expressing GST fused to the carboxy-terminal region of C protein, such as C49–107 (amino acids 49–107) and C82–107 (amino acids 82–107), or to the middle region of C protein, such as C49–83 (amino acids 49–83) and C49–86 (amino acids 49–86), were obtained by direct cloning of digested PCR fragments into pGEX-3X vector. The construct C63 expressing the amino-terminal 63 amino acids of C protein was obtained by digestion of plasmid construct C107 with restriction endonuclease *Sma*I at the *Sma*I site adjacent to the codon for amino acid 63 and the *Sma*I site in the multiple cloning site of the vector, and subsequent religation of the digested plasmid.

Each of the described constructs (Fig. 1a) was expressed using 100 ml cultures of *E. coli* DH5α, induced by adding 0.3 mM isopropyl-B-D-thiogalactoside (IPTG; Progene). 2 h later cells were pelleted and ruptured by sonication in 5 ml of RIPA buffer (50 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease inhibitors. Cell debris was removed by centrifugation and fusion proteins were purified from the resulting supernatant by affinity chromatography with 300 µl of Glutathione Sepharose 4B slurry using a batch method essentially as described in the Manual (GST Gene Fusion System, Pharmacia P-L Biochemicals Inc., 1993). Fusion proteins were eluted off the Glutathione Sepharose beads in 300 µl of elution buffer containing 50 mM Tris-HCl pH 8.0, 10 mM glutathione, and 100 mM NaCl.

Preparation of RNA probes

5′ UTR and 5′ core RNAs, representing the first 93 nt or first 440 nt of KUN RNA sequence, respectively (Fig. 1b), were obtained by SP6 RNA polymerase transcription of plasmid DNA pBSSKcore digested with *Bgl*II (for 5′ UTR) or *Bam*HI (for 5′ core RNA). To obtain the 3′ UTR RNA the sequence corresponding to the last 624 nt of KUN RNA was PCR amplified from the cDNA synthesized from ligated KUN virion RNA as described previously [12], cloned in pBluescript II KS (pBSIIKS) vector and then used for RNA transcription with T7 RNA polymerase. To obtain radiolabelled RNA probes, [^{35}S] UTP was included in these in vitro transcription reactions. All synthesized RNAs were treated with DNAse and purified by phenol-chloroform extraction and two ethanol precipitations with 2 M ammonium acetate. Digoxigenin (DIG)-labelled 5′ core and 3′ UTR RNAs were synthesized in similar in vitro

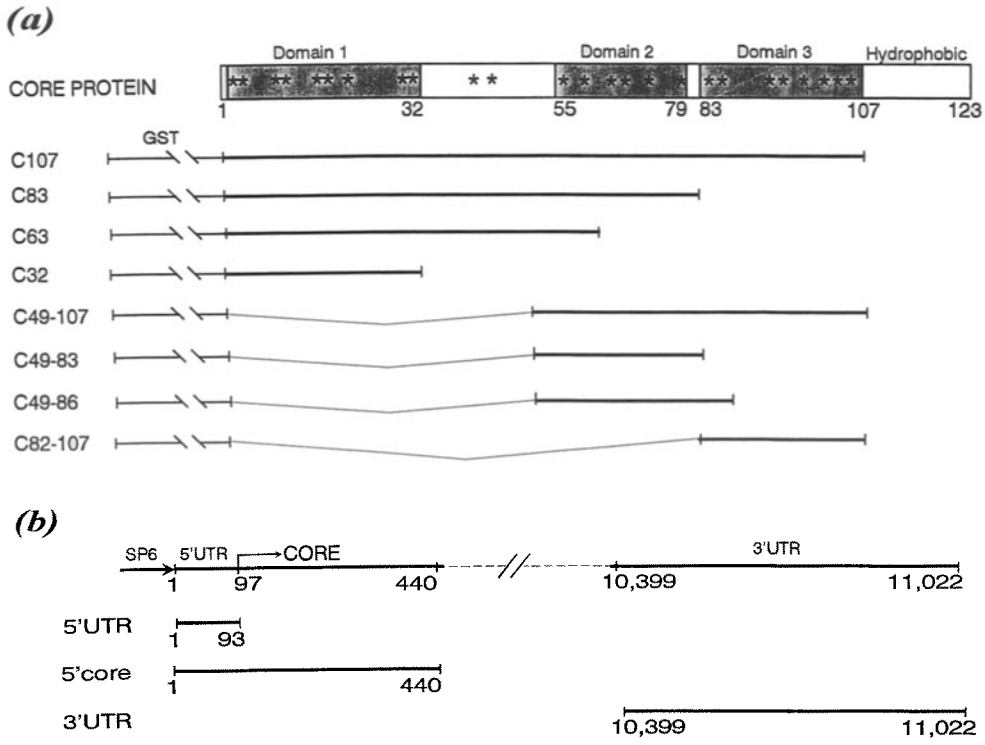

Fig. 1. Schematic representations of **a** KUN virus core proteins fused to glutathione-S-transferase (GST) expressed in *E. coli* DH5α from cDNA constructs, and **b** KUN RNA probes, all prepared as described in Materials and methods. In **a**, the numbers in C107, C83, C63 and C32 represent the numbers of N-terminal amino acids of C protein fused to GST; for the remaining GST-fusion proteins the amino- and carboxy-terminal residues of C are as indicated for C49–83, C49–86, C49–107 and C82–107. Asterisks indicate positions of basic amino acid residues (arginine or lysine); shaded boxes arbitrarily define domains of clustered basic amino acids within the boundaries indicated by amino acid numbers in the C sequence. In **b**, 5′ and 3′ UTRs represent 5′ and 3′ untranslated regions of KUN genomic RNA, respectively; 5′ core includes nearly all the coding sequence of C as well as the 5′ UTR. The nucleotide numbers at the boundaries of each sequence are shown. *SP6* RNA polymerase promoter

transcription reactions with the exceptions that the DIG RNA-labeling mixture (Boehringer Mannheim) was substituted for the rNTPs, and the DIG-labelled RNAs were purified by ethanol precipitation with 0.4 M LiCl.

To obtain ^{32}P-labelled total cell RNA, a confluent monolayer of Vero cells in a T-80 flask (Nunc) was first incubated for 2 h at 37 °C with phosphate-free minimum essential medium (MEM; Gibco BRL) plus 0.2% bovine serum albumin and then for 6 h with the same medium containing 3 mCi of [^{32}P]orthophosphate. Cells were collected and lysed at 4 °C in 400 µl of lysis buffer (50 mM Tris-HCl pH 7.6, 0.1 M NaCl, 5 mM $MgCl_2$, 0.5% NP40) supplemented with 200 units of rRNasin (Promega). The lysate was kept at 4 °C on ice for 10–15 min, and spun in a microfuge for 10 min at maximum speed at 4 °C; the supernatant was adjusted to 0.5% SDS and digested with 200 µg per ml of proteinase K for 30 min at 37 °C. RNA from the treated supernatant was isolated by phenol-chloroform and ethanol precipitation with 0.3 M sodium acetate.

Protein-RNA binding on beads

The quantities of protein and tRNA used in the binding reactions were selected as optimal for detection after using a range of concentrations in titration assays. The binding reactions were thus performed with 2–4 μg (~ 40 pmol) of fusion proteins attached to the Glutathione Sepharose beads (2–8 μl of bead slurry) and 1–2 μg (~ 10 pmol) of in vitro synthesized unlabelled or DIG-labelled RNAs, or 20–28 ng (~ 0.14 pmol) of ^{35}S-labelled RNAs, or 500 000 cpm of ^{32}P-labelled total Vero cell RNA, in 100 μl of binding buffer (BB; 20 mM Tris-HCl pH 7.6, 0.05% NP40, 100 mM NaCl, 1 mM β-mercaptoethanol, 0.75 mM ATP, and 20U of rRNasin) at 4 °C for 1 h. In competition experiments appropriate amounts of yeast tRNA (Boehringer Mannheim) or unlabelled in vitro synthesized RNAs were added to reaction mixtures simultaneously with RNA probes. After the reaction the beads were washed 3 times with 0.5 ml BB to remove residual unbound RNA. RNA bound to the fusion protein on beads were recovered by digestion of the protein with proteinase K (0.5 mg per ml) in the buffer containing 1% SDS, 200 mM Tris-HCl pH 8.0, 10 mM NaCl, 1 mM EDTA. Recovered unlabelled RNA was identified by ethidium bromide staining after electrophoretic separation in a nondenaturing agarose gel in TAE buffer (40 mM Tris-acetate, 1 mM EDTA). Radiolabelled recovered RNA was identified by counting of a sample in a β-counter.

Northwestern blotting

RNA-protein binding reactions using the Northwestern blotting technique were performed with DIG-labelled KUN 5′ core and 3′ UTR RNAs and ^{32}P-labelled cell RNA essentially as described by Stohlman et al. [22]. Briefly, purified proteins were removed from the beads by boiling in gel loading buffer, separated in a 12% Laemmli gel as described previously [21] and transferred electrophoretically to a nitrocellulose membrane. Transferred proteins were allowed to renature on the membrane at room temperature overnight in 30 ml of SBB buffer (10 mM Tris-HCl pH 7.6; 50 mM NaCl; 1 mM EDTA; 0.02% bovine serum albumin; 0.02% Ficoll; 0.02% polyvinyl pyrrolidone). The membranes were then transferred into 50 ml capped tubes, overlaid with 3 ml of SBB buffer containing either 1 μg of DIG-labelled RNA or 500,000 cpm of ^{32}P-labelled total Vero cell RNA and incubated for 2–3 h at room temperature with constant rotation. After incubation the membranes were washed 3 times for 15 min with SBB buffer and were either dried and exposed to X-ray film (for detection of ^{32}P-labelled RNA) or processed for colorimetric detection of DIG-labelled RNA as described in the Manual (The DIG System User's Guide for Filter Hybridization, Boehringer Mannheim, GmbH, Biochemica, 1993).

Results

Reactions of RNAs with GST-core fusion C107

C107 was expressed in *E. coli* and purified by retention on Glutathione Sepharose beads as described in Materials and methods. The three viral RNA probes 3′ UTR, 5′ core and 5′ UTR (Fig. 1b) bound to C107 on the beads in the presence of 100 mM NaCl (Fig. 2a). Most of added ^{35}S-labelled 5′ core RNA (20 ng) was retained when bound to 1 μg C107 on the beads (64% and 51% in two experiments) whereas a smaller proportion (31% and 28% in two experiments) of ^{35}S-labelled 3′ UTR (28 ng added) was retained. Larger amounts (1–2 μg) of unlabelled or DIG-labelled RNAs were used in subsequent experiments, hence an excess of RNA was always added for binding to C107 fusion protein. A measure of the specificity of RNA-protein binding is its resistance to inhibition

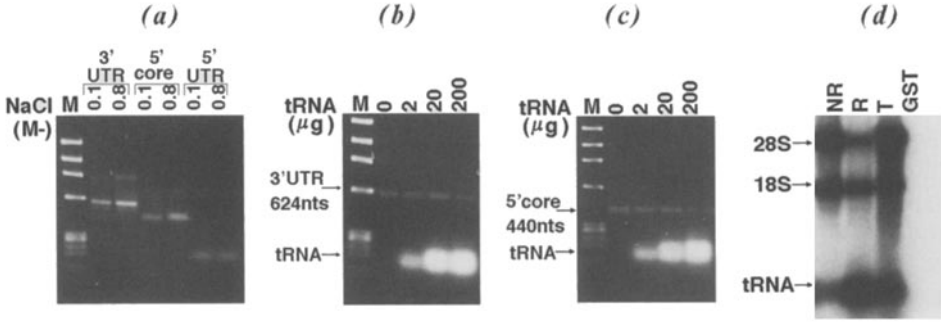

Fig. 2. Binding of KUN RNA probes 3′ UTR (624 nucleotides), 5′ core (440 nucleotides), 5′ UTR (93 nucleotides) and cell RNA to purified GST-core fusion protein C107 attached to Glutathione Sepharose beads. **a** Resistance to high salt concentration, **b**, **c** effects of increasing concentrations of yeast tRNA on binding efficiency. **d** Binding of ^{32}P-labelled Vero cell RNA. All reactions were performed as described in Materials and methods. RNA attached to the fusion proteins was recovered by proteinase K digestion and stained with ethidium bromide after electrophoresis in 1.8% agarose gels. Each *M* lane contained ϕX174 RF DNA digested with *Hae* III, and *M*- refers to molarity (0.1 and 0.8) of NaCl. In **d** T is ^{32}P-labelled Vero cell total RNA before application to the beads, *R* and *NR* represent labelled RNA retained or not retained on beads, respectively. The last lane (*GST*) shows that labelled 28S and 18S ribosomal RNAs, and tRNA, were not retained on beads carrying GST only

by high salt concentration ($\geq$ 600 mM NaCl; [4, 6, 17]). We therefore included 800 mM NaCl in the binding reactions which were thus shown to be unaffected by high salt (Fig. 2a). Reactions of C107 with the viral probes 3′ UTR and 5′ core in 100 mM NaCl were repeated in competition experiments with yeast tRNA. Although tRNA bound strongly to C107 in the presence of both probes (using 2, 20, and 200 μg tRNA), no clear evidence of competition with the KUN RNA probes was observed (Fig. 2b, c). Thus the 5′ and 3′ UTR regions of KUN RNA bound to full length core protein C107 in a relatively specific manner.

In order to analyse possible binding of cellular RNA to C, we prepared ^{32}P-labelled Vero cell RNA (see Materials and methods). Radiolabelled Vero cell RNA bound strongly to C107 in Northwestern blots (results not shown). In binding experiments on Glutathione Sepharose beads, we found by polyacrylamide gel analysis of the bound RNA and autoradiography of dried gels that the major binding component of Vero cell RNA was tRNA (Fig. 2d), which was competed out by an excess of yeast tRNA (results not shown). We therefore continued to use yeast tRNA in competitive RNA binding assays. A minor amount of ^{32}P-labelled 18S and 28S RNA also bound to C-107 on the beads. No labelled RNA was bound to GST alone.

Comparisons of RNA binding reactions of C107 with those of deleted core protein fusion constructs

In attempts to define which regions of the mature form of C were binding to RNA, three regions enriched in basic amino acids were chosen arbitrarily for

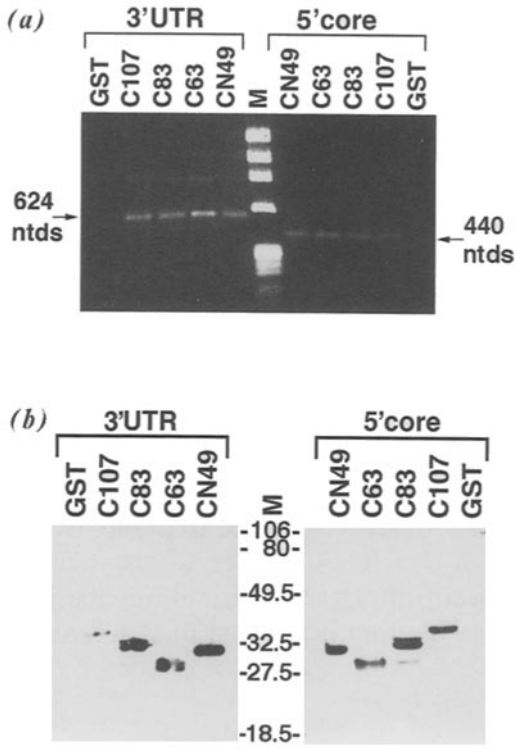

Fig. 3. Binding of KUN RNA probes 3′ UTR and 5′ core to the GST-core fusion proteins C107, C83, C63, and C49–107 (see Fig. 1a). **a** Ethidium bromide-stained 1.8% agarose gel of electrophoresed RNAs recovered from Glutathione Sepharose beads by proteinase K digestion after binding reactions with the fusion proteins. **b** Northwestern blot analysis of DIG-labelled RNAs bound to the fusion proteins on nitrocellulose membranes. Binding reactions were performed as described in Materials and methods. CN49 represents C49–107. In **a**, *M* represents ϕX174 RF DNA, digested with *Hae*III. In **b**, *M* represents the low molecular weight range of prestained proteins (BIO-RAD) as SDS-PAGE standards

expression as domains incorporating six to nine residues of arginine/lysine. Fusion proteins with amino-terminal deletions (designated C49–107) or carboxy-terminal deletions (C63, C83) (see Fig. 1a) were expressed in *E. coli* and purified from cell lysates on beads. All these deleted fusion proteins bound to 3′ UTR and 5′ core probes on beads as efficiently as C107 (Fig. 3a). Similar binding results were obtained with the same reagents using Northwestern blots as the assay system (Fig. 3b). In both binding assays GST alone did not bind to either KUN probe. All the core fusion protein constructs migrated on SDS-polyacrylamide gels according to their predicted molecular weights except for C63 which migrated faster than anticipated (Fig. 3b). The nucleotide sequence from which C63 mRNA was transcribed in *E. coli* was checked and found to be correct. Either C63 migrated anomalously in gels, or the carboxy terminus was degraded, possibly by *E. coli* proteases, reducing the size of C protein moiety by approximately 50%. If C63 was truncated in this manner, seven to nine basic amino acid residues of the amino-terminal region would still remain (Fig. 1a). These results did not identify location of the binding regions in C, but showed that the presence of GST protein fused to the amino terminus of the core protein did not compete with or appear to interfere with binding of the RNA probes. The Northwestern blots established that the RNA probes were binding only to the purified core fusion proteins, and not to any other proteins which could possibly be attached to the Glutathione Sepharose beads either directly from *E. coli* lysates or coprecipitated with GST-core fusion proteins during purification.

Binding of 2 μg tRNA to the deleted fusion proteins C49–107 and C83 was quantitatively similar to binding of 2 μg tRNA to C107, but apparently had reached saturation binding because little or no increase in binding of tRNA was observed when the larger amounts 20 to 200 μg tRNA were added to the assay. 200 μg tRNA competed weakly with binding of the 3′ UTR and 5′ core KUN RNA probes to C49–107 and C83. Furthermore, this binding was not blocked by 800 mM NaCl (results not shown). With C63, tRNA bound weakly; a large excess (200 μg) was required for detection of about 1 μg tRNA bound. However, binding of both KUN RNA probes to C63 was as efficient as that to C107, C49–107 and C83, but in contrast was blocked by 800 mM NaCl and competed out by 200 μg tRNA (results not shown). Thus although C63 bound KUN RNA probes more readily than tRNA, its binding to the KUN RNA probes was weaker than that of C49–107 and C83.

These results showed that binding of KUN 5′ core and 3′ UTR probes to C49–107 and C83 was very similar to that observed with C107, and tRNA competed weakly with binding of these KUN RNA probes to the deleted C polypeptides only when used at high concentrations (200 μg or about 500-fold molar excess). The location of the possible binding domain(s) in C remained undefined.

Binding of UTRs to fusion proteins containing single domains of basic amino acids

Because no clear picture had emerged showing which if any basic amino acid region in C was the major contributor to the apparent specificity of UTR binding to C107, shorter fusion protein constructs were prepared each containing only a single arbitrary domain enriched in basic amino acids (C32, C49–83 and C82–107; Fig. 1a). All of these fusion proteins containing such short sequences of C migrated similarly in SDS-polyacrylamide gels as expected based on their predicted molecular weights (Fig. 4a). C49–83 did not detectably bind 5′ core RNA, but C32 and C82–107 did (Fig. 4b). Despite their small size (containing only 32 or 26 amino acids of C), C32 and C82–107 bound similar amounts of KUN UTRs as C107 did previously under the same conditions. For example, C82–107 bound about 51% (56% and 46% in two experiments) and about 34% (39% and 29% in two experiments) of ^{35}S-labelled 5′ core and 3′ UTR probes, respectively. C32 bound about 42% and 32%, respectively, of the same labelled probes. These results appeared to exclude the central region of C, incorporating the arbitrary domain 2 in C49–83, as a KUN UTR binding region.

Competition binding experiments were extended with C32 and C82–107 using ^{35}S-labelled 5′ core and 3′ UTR RNA probes and excess of the same probes but unlabelled as competitors. Homologous UTR probes competed (excess of unlabelled versus ^{35}S-labelled UTRs) as did the KUN heterologous UTR probes (excess of unlabelled 5′ core versus ^{35}S-labelled 3′ UTR, and excess of unlabelled 3′ UTR versus ^{35}S-labelled 5′core; results not shown). 20 μg tRNA bound weakly to C32 and C82–107 and did not apparently compete with the 5′ core RNA probe

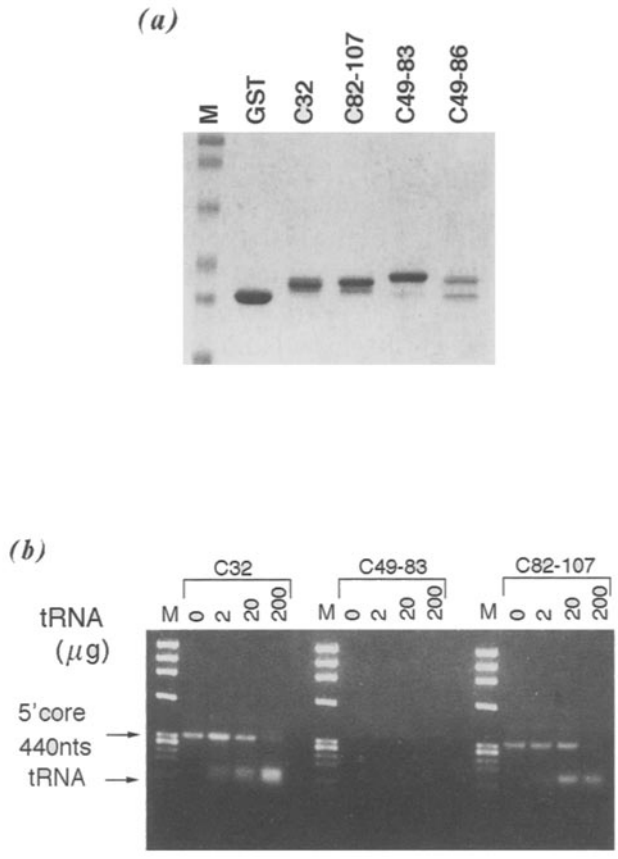

Fig. 4. SDS-PAGE analysis of purified KUN core-fusion proteins (see Fig. 1a) and their use in competition RNA binding experiments between 5′ core RNA and tRNA. In **a**, GST-fusion proteins expressed in *E. coli* were purified with Glutathione Sepharose 4B beads, eluted with glutathione as described in Materials and methods, and identified by SDS-PAGE in a 12% gel stained with Coomassie Blue. *GST* is GST protein expressed and purified as for the others; *M* represents low molecular weight range of prestained proteins (BIO-RAD) as used in Fig. 3b. In **b**, 5′ core RNA (2 μg) was mixed with 0, 2, 20 or 200 μg tRNA (0 to 500 molar excess) and used in RNA competition binding experiments with C32, C49–83 and C82–107 (see Fig. 1a) attached to Glutathione Sepharose beads. RNA bound to the beads was recovered and analysed in nondenaturing agarose gels (see Materials and methods). *M* indicates marker lanes containing ϕX174 RF DNA digested with Hae III

(Fig. 4b). However, 200 μg tRNA (500 molar excess) successfully competed out the binding. In salt block experiments, binding of both 5′ core and 3′ UTR probes to C32 and C82–107 occurred at 200 mM NaCl but inhibition was observed at 400 mM NaCl (results not shown). Thus although the short peptides in fusion proteins C32 and C82–107 bound the KUN UTR probes, their strength of binding was reduced relative to that of the longer peptides in C49–107 and C83.

As a further test for binding capacity of C32 and C82–107, we repeated the binding experiments using homologous RNA and tRNA as competitors in gel shift assays. We also prepared a new construct C49–86 (Fig. 1a) in lieu of C49–83

to ascertain whether the addition of two basic amino acids (total content of 8) would produce binding of UTRs unattainable with C49–83 (total content of 6 basic amino acid residues) in the binding assays on Glutathione Sepharose beads. C49–86 migrated similarly to C49–83 (Fig. 4a). C32, C49–86, and C82–107 were purified from *E. coli* lysates on Glutathione Sepharose beads and eluted off the beads in buffer containing 10 mM glutathione (see Materials and methods). Conditions for binding of the fusion proteins to the ^{35}S-labelled RNA probes were selected after titration experiments using gel shift assays. Increase in concentration of fusion proteins (from 0.1 to 1.0 μg) resulted in increases in mobility shift, and 0.2 to 0.3 μg was selected for further assays. Both C82–107 and C83 caused obvious gel shifts after incubation with ^{35}S-labelled 5′ core and 3′ UTR probes, which were readily competed with 10 to 50-fold molar excess of unlabelled homologous probes (Fig. 5a, b). In contrast, tRNA in 250-fold molar excess failed to completely eliminate the gel shifts observed in reactions of C82–107 with both ^{35}S-labelled 5′ core and 3′ UTR probes, or in reactions of C32 with the 3′ UTR probe. A 10-fold molar excess of tRNA competed, albeit weakly, with gel shift of ^{35}S-labelled 5′ core RNA and C32, and more than a 50-fold molar excess was required to eliminate this gel shift completely. Thus these two non-overlapping regions of KUN C enriched in basic amino acids (C32 or amino-terminal, and C82–107 or carboxy-terminal) individually bound KUN 5′ core and 3′ UTR which were competed by excess of unlabelled homologous probes but only partially by a large excess of tRNA, in both gel shift assays and binding reactions on Sepharose beads.

When gel shift assays with C49–86 were attempted under the same conditions used for C32 and C82–107, the migration of the 3′ UTR and 5′ core probes was unaffected (results not shown), agreeing with the lack of binding observed previously in the bead assay using C49–83 (Fig. 4b). Thus the addition of two more basic amino acids (an increase of content from 17% in C49–83 to 21% in C49–86) was insufficient to induce RNA-protein binding, and the result therefore apparently excludes this region of C in RNA binding activity. Therefore it can be concluded that the binding observed with C32, C63 and C83 was associated with domain 1, and that the stronger binding observed with C49–107 and C82–107 was associated with domain 3 (see Fig. 1a). These two binding domains were thus localized to peptides of 32 amino acid residues (in C32) and 26 residues (in C82–107).

Discussion

This is the first report of RNA binding of the genus flavivirus core protein in vitro. Fusion protein C107 bound strongly to the KUN UTR riboprobes in a relatively specific matter, being resistant to competition by a large excess of tRNA and to binding inhibition by 0.8 M NaCl. C107 contains the virtual equivalent of KUN mature C of 105 amino acids which is the detectable form of C found in virions and in infected cells [16, 21]. Within the core protein sequence, putative KUN UTR binding domains were shown to reside in the amino-terminal and carboxy-terminal sequences, exemplified by the binding reactions

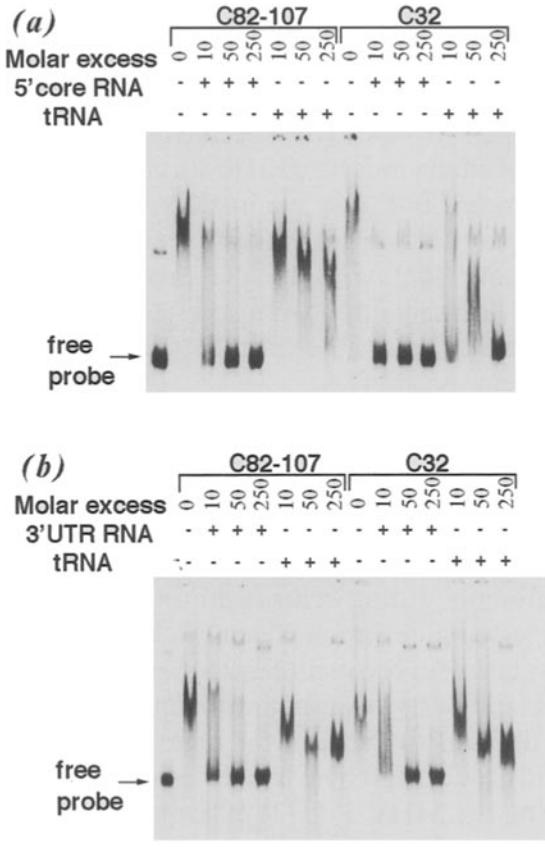

Fig. 5. Gel shift assays with **a** KUN 5′ core and **b** 3′ UTR probes in reactions with GST-core fusion proteins C32 and C82–107. Protein-RNA binding reactions in solution were performed with 0.2–0.3 μg (~ 10 pmol) of eluted purified fusion proteins (see the text) and 10–14 ng (~ 70 fmol) of ^{35}S-labeled 5′ core and 3′ UTR RNAs in 10 μl of reaction buffer containing 20 mM Tris-HCl pH 7.6, 0.05% NP40, 100 mM NaCl, 1 mM β-mercaptoethanol, 5% glycerol, and 10 U of rRNasin for 20 min at 30 °C. The mixtures were then separated on a 4% non-denaturing polyacrylamide gel containing 5% glycerol in 0.5 × TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA) electrophoresed in 0.5 × TBE for 2–3 h at 150 V at 4 °C. The gel was dried and exposed to X-ray film for detection of radiolabelled RNA. In the competition experiments, 10, 50, and 250 molar excess of unlabelled homologous RNAs or yeast tRNA was added to reaction mixtures simultaneously with radiolabelled RNAs

of C32 and C82–107. These relatively short peptides probably act co-operatively in the mature C molecule because individually they lack the strong resistance to inhibition by 800 mM NaCl and are partially competed by large excess of nonspecific RNA (Figs. 4b, 5, and results of salt block experiments) in contrast to C107. With other positive-stranded RNA viruses, nonspecific as well as specific binding to RNA by coat or capsid proteins occurs commonly. Binding of the coat protein (CP) of the tripartite RNA virus brome mosaic virus (BMV) to the 3′ terminal 200 nucleotides of BMV RNA-3 was easily blocked in gel shift assays by only 3- or 4-fold molar excess of nonspecific RNA competitors comprising

reporter (305 nucleotides) or vector (224 nucleotides) sequences [6]. However, BMV CP bound specifically to full length RNA 1 and nonspecifically to vector RNA, based on the observation that only the specific complex was resistant to dissociation by 0.4 to 0.8 M NaCl [6]. Specific binding to CP by the 3′ UTR of RNA-3 and RNA-4 of alfalfa mosaic virus (both containing conserved stem loop structures) was reduced but not eliminated by addition of excess non-homologous competitor RNA in gel shift assays [18]. Binding to the nucleocapsid protein of Puumala hantavirus by the small negative-stranded genomic S RNA segment was strongly inhibited also in gel shift assays by 5 to 50 mass excess of tRNA [9], even though the viral RNA and nucleocapsid protein bind specifically during virus assembly.

RNA binding domains within capsid proteins and involving enrichment in basic amino acids have been analysed for several positive-stranded RNA viruses, including the BMV CP discussed above [6]. The amino-terminal region (amino acids 1–75, 31% enriched in lysine and arginine) of hepatitis C virus core protein bound to HCV genomic RNA and hepatitis B virus RNA in Northwestern blots, and to ribsomal subunits during centrifugation [20]. Masters [13] described binding of the lysine-arginine enriched central region of mouse hepatitis virus (MHV) nucleocapsid protein N to cellular RNA during in vitro translation of N, detected by gel shift analyses. However, Stohlman's group [15, 22] also identified an RNA-binding domain in the central region of MHV native N protein (between amino acids 169 and 308) by Northwestern blotting, using ^{32}P-labelled riboprobe containing the MHV 5′ UTR or leader RNA sequence which could not be competed by a large excess of total cellular RNA. Sindbis virus capsid protein C (264 amino acids) bound to cellular RNA during in vitro translation; subsequently this cellular RNA which had bound to C was successfully competed in a gel migration assay by a 849 nucleotide Sindbis RNA probe containing the 572 nucleotide packaging signal in ongoing studies on encapsidation by S. Schlesinger and colleagues [8]. Despite acknowledged considerable variation between experiments, the authors established by deletion of amino-terminal and carboxy-terminal sequences in Sindbis C that the regions between amino acids 76–116 (39% enriched in lysine and arginine) were the most important for binding of a 68 amino acid peptide Δ17 to the nucleotide packaging signal. Similar results with another alphavirus (Semliki Forest) were reported by Forsell et al. [7] who showed by deletions in genomic RNA transcripts that most of the basic amino acids (27% enriched) in the amino- terminal 112 residues of the capsid protein were required for capsidation in vivo and infectivity, and concluded that most of the region is involved in both specific and nonspecific interactions with encapsidated RNA. Interestingly, it was shown that the Sindbis fusion protein GST- Δ17 or free Δ17 caused an (incomplete) gel shift of a "highly structured" ^{35}S-labelled 132 nucleotide viral RNA fragment (containing four stem loops predicted by computer folding) with the capsid binding activity [25]. Bacterial RNA and tRNA at about 25-fold mass excess were effective competitors (60 to 70% inhibition) only when the Sindbis 132-mer RNA probe was reduced to a 67-mer still able to bind to the active 68 amino acid sequence

in GST-Δ17. Thus nonspecific binding became more apparent when the 132-mer Sindbis RNA probe was reduced to 67 nucleotides. In similar experiments, the short KUN 5′ UTR riboprobe of 93 nucleotides was still relatively specific; it bound C107 and resisted inhibition by 0.8 M NaCl (Fig. 2a). In most cases tRNA appeared to be able to bind independently of the presence of the KUN riboprobes, and Vero cell tRNA was retained on C107 in greater proportion than 28S and 18S ribosomal RNA, both many times larger than tRNA (Fig. 2d).

Clearly our results are analogous to those obtained with other virus coat or capsid proteins involving both specific and nonspecific RNA binding. All flavivirus C proteins have a similar distribution of basic amino acids (see [5]) which may be arbitrarily grouped for KUN virus as in Fig. 1: domain 1 comprising amino acids 3–32 (9 basic residues, 30% enrichment), domain 2 comprising amino acids 55–79 [6, 24%] and domain 3 comprising amino acids 84–107 (8, 33%). C107 and all the shorter fusion proteins except C49–83 and C49–86 bound the KUN UTR riboprobes, thus excluding the arbitrary domain 2 from further consideration. Binding of the KUN riboprobes to the isolated amino-terminal domain 1 (in C32) and to the isolated carboxy-terminal domain 3 (in C82–107) was inhibited by high salt, but was strongly or partially resistant to competition by large excess of tRNA (Figs. 4b, 5). Thus our results show that RNA binding to KUN core fusion proteins is easily demonstrable, and that relatively specific binding appears to occur between C107, the virtual equivalent of mature core protein C, and the 5′ and 3′ UTR regions of KUN RNA. The binding specificity appears to reside in domains 1 and 3, probably acting in concert.

While the criteria for binding of viral coat proteins to RNA including viral UTRs cannot as yet be precisely defined, several general principles are emerging. Enrichment in basic amino acids obviously facilitates binding and helps localize the binding domains experimentally [6, 7, 8, 13, 15, 20, 22]. The binding domains may be relatively short if the arginine/lysine content is in the vicinity of 25% or greater e.g. the Δ17 peptide (68 residues) of Sindbis C [8], in the capsid protein of Semliki Forest virus [7], and the 26 and 32 residue peptides in C82–107 and C32, respectively, of KUN C. Not surprisingly, binding was stronger for longer peptides and in some cases for longer RNA sequences ([8, 25], the KUN binding data). Secondary structures within viral RNA may also enhance specificity [25], especially the stem loop structures in UTRs of flaviviruses [1, 2, 10, 14, 19, 26] and in UTRs of several other positive-stranded viruses for which binding has been reported (see Introduction). As noted above, the ordered structure of tRNA bound more strongly than the longer ribosomal RNA to the KUN core fusion protein C107 (Fig. 2d), but was able to partially compete in binding of the KUN 5′ and 3′ UTR probes with the short core fusion proteins C32 and C82–107 only when in large excess (Figs. 4b, 5). A further criterion for binding is resistance to inhibition by high salt, as observed in the present study and in other reports [4, 6, 17].

It seems reasonable that some specific association must occur in infected cells similar to that observed in vitro with C107 and the KUN UTRs and that the stem

loop structures in the KUN 5′ and 3′ UTRs acting together in virion RNA may contribute to the strength of binding to mature C. It is not possible from this initial study to conclude what role the observed binding reactions play in vivo. Some additional binding may occur with longer (coding) regions of KUN RNA, but such reactions were beyond the scope of the present study, and will be the subject of future work. Whether nonspecific RNA interactions also play a role in flavivirus replication is uncertain. There have been suggestions that a lack of RNA specificity may aid in virion assembly [6, 7, 23]. Further work on the role or extent of RNA-protein binding in replication of flaviviruses is needed because of the dearth of information in this area.

Acknowledgement

This work was supported by a grant from the National Health and Medical Research Council of Australia.

References

1. Brinton MA, Dispoto JH (1988) Sequence and secondary structure analysis of the 5′-terminal region of flavivirus genome RNA. Virology 162: 290–299
2. Brinton MA, Fernandez V, Dispoto JH (1986) The 3′-nucleotides of flavivirus genomic RNA from a conserved secondary structure. Virology 153: 113–121
3. Chambers TJ, Hahn CS, Galler R, Rice CM (1990) Flavivirus genome organization, expression, and replication. Annu Rev Microbiol 44: 649–688
4. Citovsky V, Knorr D, Schuster G, Zambryski P (1990) The P30 movement protein of tobacco mosaic virus is a single-strand nucleic acid binding protein. Cell 60: 637–647
5. Coia G, Parker MD, Speight G, Byrne ME, Westaway EG (1988) Nucleotide and complete amino acid sequences of Kunjin virus: definitive gene order and characteristics of the virus-specified proteins. J Gen Virol 69: 1–21
6. Duggal R, Hall TC (1993) Identification of domains in brome mosaic virus RNA-1 and coat protein necessary for specific interaction and encapsidation. J Virol 67: 6406–6412
7. Forsell K, Suomalainen M, Garoff H (1995) Structure-function of the NH_2-terminal domain of the Semliki Forest virus capsid protein. J Virol 69: 1556–1563
8. Geigenmüller-Gnirke U, Nitschko H, Schlesinger S (1993) Deletion analysis of the capsid protein of Sindbis virus: identification of the binding region. J Virol 67: 1620–1626
9. Gött P, Stohwasser R, Schnitzler P, Darai G, Bautz EKF (1993) RNA binding of recombinant nucleocapsid proteins of hantaviruses. Virology 194: 332–337
10. Hahn CS, Hahn YS, Rice CM, Lee E, Dalgarno L, Strauss EG, Strauss JH (1987) Conserved elements in the 3′ untranslated region of flavivirus RNAs and potential cyclization sequences. J Mol Biol 198: 33–41
11. Hase T, Summers PL, Eckels KH, Putnak JR (1989) Morphogenesis of flaviviruses. In: Harris JR (ed) Subcellular biochemistry, vol 15. Plenum, New York London, pp 275–305
12. Khromykh AA, Westaway EG (1994) Completion of Kunjin virus RNA sequence and recovery of an infectious RNA transcribed from stably cloned full-length cDNA. J Virol 68: 4580–4588
13. Masters PS (1992) Localization of an RNA-binding domain in the nucleocapsid protein of the coronavirus mouse hepatitis virus. Arch Virol 125: 141–160

14. Mohan PM, Padmanabhan R (1991) Detection of stable secondary structure at the 3′ terminus of dengue virus type 2 RNA. Gene 108: 185–191
15. Nelson GW, Stohlman SA (1993) Localization of the RNA-binding domain of mouse hepatitis virus nucleocapsid protein. J Gen Virol 74: 1975–1979
16. Nowak T, Färber PM, Wengler G, Wengler G (1988) Analysis of the terminal sequences of West Nile virus structural proteins and of the in vitro translation of these proteins allow the proposal of a complete scheme of the proteolytic cleavages involved in their synthesis. Virology 169: 365–376
17. Pardigon N, Strauss JH (1991) Cellular proteins bind to the 3′ end of Sindbis virus minus-strand RNA. J Virol 66: 1007–1015
18. Reusken CBEM, Neelman L, Bol JF (1994) The 3′-untranslated region of alfalfa mosaic virus RNA 3 contains at least two independent binding sites for viral coat protein. Nucleic Acids Res 22: 1346–1353
19. Rice CM, Lenches EM, Eddy SR, Shin SJ, Sheets RL, Strauss JH (1985) Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science 229: 726–733
20. Santolini E, Migliaccio G, Monica NL (1994) Biosynthesis and biochemical properties of the hepatitis C virus core protein. J Virol 68: 3631–3646
21. Speight G, Westaway EG (1989) Carboxy-terminal analysis of nine proteins specified by the flavivirus Kunjin: evidence that only the intracellular core protein is truncated. J Gen Virol 70: 2209–2214
22. Stohlman SA, Baric RS, Nelson GN, Soc LH, Welter LM, Deans RJ (1988) Specific interactions between coronavirus leader RNA and nucleocapsid protein. J Virol 62: 4288–4295
23. Wei N, Morris TJ (1991) Interactions between viral coat protein and a specific binding region on turnip crinkle virus RNA. J Mol Biol 222: 437–443
24. Wei N, Hacker DL, Morris TJ (1992) Characterization of an internal element in turnip crinkle virus RNA involved in both coat protein binding and replication. Virology 190: 346–355
25. Weiss B, Geigenmüller-Gnirke U, Schlesinger S (1994) Interactions between Sindbis virus RNAs and a 68 amino acid derivative of the viral capsid protein further defines the capsid binding site. Nucleic Acids Res 22: 780–786
26. Wengler G, Castle E (1986) Analysis of structural properties which possibly are characteristic for the 3′-terminal sequence of the genome RNA of flaviviruses. J Gen Virol 67: 1189–1194
27. Westaway EG (1980) Replication of flaviviruses. In: Schlesinger RW (ed) The togaviruses. Biology, structure, replication. Academic Press, New York, pp 531–581
28. Westaway EG (1987) Flavivirus replication strategy. Adv Virus Res 33: 45–90

Authors' address: Dr. E. G. Westaway, Sir Albert Sakzewski Virus Research Centre, Royal Children's Hospital, Herston Road, Herston, Brisbane, Qld. 4029, Australia.

Received June 29, 1995