Identification of Sam68 Arginine Glycine-rich Sequences Capable of Conferring Nonspecific RNA Binding to the GSG Domain*

Taiping Chen‡, Jocelyn Côté§§, Héctor Valderrama Carvajal†, and Stéphane Richard¶

From the Terry Fox Molecular Oncology Group and the Bloomfield Center for Research on Aging, Lady Davis Institute for Medical Research, Sir Mortimer B. Davis Jewish General Hospital, and Departments of Oncology, Medicine, Microbiology and Immunology, McGill University, Montreal, Quebec H3T 1E2, Canada

Sam68 is an RNA-binding protein that contains a heterogeneous nuclear ribonucleoprotein K homology domain embedded in a larger RNA binding domain called the GSG (GRP33, Sam68, GLD-1) domain. This family of proteins is often referred to as the STAR (signal transduction and activators of RNA metabolism) proteins. It is not known whether Sam68 is a general nonspecific RNA-binding protein or whether it recognizes specific response elements in mRNAs with high affinity. Sam68 has been shown to bind homopolymeric RNA and a synthetic RNA sequence called G8–5 that has a core UAAA motif. Here we performed a structure function analysis of Sam68 and identified two arginine glycine (RG)-rich regions that confer nonspecific RNA binding to the Sam68 GSG domain. In addition, by using chimeric proteins between Sam68 and QKI-7, we demonstrated that one of the Sam68 RG-rich sequences of 26 amino acids was sufficient to confer homopolymeric RNA binding to the GSG domain of QKI-7, another STAR protein. Furthermore, that minimal sequence can also give QKI-7 the ability (as Sam68) to functionally substitute for HIV-1 REV to facilitate the nuclear export of RNAs. Our studies suggest that neighboring RG-rich sequences may impose nonspecific RNA binding to GSG domains. Because the Sam68 RNA binding activity is negatively regulated by tyrosine phosphorylation, our data lead us to propose that Sam68 might be a specific RNA-binding protein when tyrosine phosphorylated.

Sam68 (Src substrate associated during mitosis of 68 kDa)† is a substrate for tyrosine kinases including Src family kinases p60-src (1–4), p59fyn (5), p56lck (6), BRK/SIK (7), and Zap70 (8). Sam68 has been shown to bind numerous Src homology 3, Src homology 2, and WW domain-containing proteins, leading several groups to suggest that Sam68 may be an adaptor protein for tyrosine kinases (5, 9). Sam68 is an RNA-binding protein that contains a KH domain embedded in a larger domain of ~200 amino acids, the GSG (GRP33, Sam68, and GLD-1) domain. Sam68 has been shown to bind homopolymeric RNA poly(U) and poly(A) (2, 10). The tyrosine phosphorylation of Sam68 by p59fyn severely inhibits its ability to bind poly(U)-Sepharose (11). Sam68 has also been shown to bind synthetic RNA sequences with a core UAAA with high affinity (12). The function of Sam68 is unknown, but recent findings suggest that it may be involved in the regulation of splicing and/or RNA transport (13–16).

The GSG domain is an evolutionarily conserved protein module initially identified by aligning the first three members of this family (17, 18). In addition to the KH domain, the GSG domain contains ~75 amino acids N-terminal and ~25 amino acids C-terminal to the KH domain called the NK (N-terminal of KH) and CK (C-terminal of KH) regions, respectively (schematically represented in Fig. 1). Several properties have been ascribed to the GSG domain including RNA binding (10, 12, 18–21), self-association (10, 18, 19, 22), heterodimerization (10, 18, 19, 22), and protein localization (13).

GSG domain-containing proteins are called STAR proteins for signal transduction and activators of RNA metabolism (24, 25), and Sam68 is the prototype because of its links to signaling proteins (5, 9). The GSG domain is found in a rapidly growing family of RNA-binding proteins (25). Genetic and biochemical evidence has demonstrated that STAR proteins are involved in many essential processes such as splicing (26, 27), tumorogenesis (17, 28), apoptosis (18, 19, 29), cell cycle progression (30), translation (31, 32), and development (17, 22, 33, 34).

The physiological importance of the GSG domain is demonstrated by the fact that many genetic mutations that result in growth or developmental defects have been identified in this protein module. In the nematode Caenorhabditis elegans, the GSG protein GLD-1 functions as a tumor suppressor that is required for normal oocyte development (35, 36). Thirty-two gld-1 mutations have been identified that fall into six phenotypic classes (17). In mice, a missense mutation in the quaking gene (qk) has been identified (24) that is known to be embryonic-lethal (37). This mutation, altering glutamic acid 48 to glycine (24), occurs in the NK region of the GSG domain and has been shown to prevent QKI dimerization (19). In Drosophila melanogaster, HOW plays a critical role in skeletal muscle development, because weak alleles result in the “held-out-wings” phenotype (33, 34).

The phenotype of the quaking viable and lethal mice suggests that the QKI proteins are involved in myelination and...
early embryogenesis (24, 38). The mouse qk gene expresses at least five alternatively spliced mRNAs including QKI-5, QKI-6, and QKI-7 that differ in their C-terminal 30 amino acids (24). The quaking viable mutation, which prevents the expression of QKI-6 and QKI-7 isoforms in oligodendrocytes (39), severely impairs myelination, and as a result the mice develop a characteristic tremor after birth (40). The specific RNA targets of QKI have not been identified, but the sequence identity characteristic tremor after birth (40). The specific RNA targets of QKI-6 and QKI-7 isoforms in oligodendrocytes (39), severely impairs myelination, and as a result the mice develop a characteristic tremor after birth (40). The specific RNA targets of QKI have not been identified, but the sequence identity between C. elegans GLD-1 and mouse QKI proteins has led Goodwin and co-workers (41) to examine whether the introduction of QKI-6, similar to GLD-1, translationally suppressed the expression of TRA-2 and bound with high affinity to the “tra-2 and GLI elements” (TGEs) (41).

Previously we have shown that the Sam68 GSG domain in addition to at least 50 amino acids at its C terminus are necessary and sufficient for RNA binding (10). To examine the role of the C-terminal amino acids in RNA binding, we generated chimeric proteins between two STAR proteins, Sam68 and QKI. Because the primary amino acid sequence of QKI-5, -6, and -7 isoforms are identical except for the last 6–30 amino acids (depending on the isoform) and are predicted to have identical RNA binding specificity, we chose QKI-7 for our analysis. Here we identified two small regions harboring arginine-glycine repeats in Sam68 that can confer nonspecific RNA binding activity to an adjacent GSG domain. In addition, a novel Sam68 dimerization region has been identified in its C-terminal sequences.

MATERIALS AND METHODS

DNA Constructions—The constructs encoding Myc-QKI-7, Myc-QKI-7:E48G, Myc-Sam68, and Myc-Sam68:3RG were described previously (5, 10). Myc-S-Q codes for the C-terminal 180 amino acids of QKI-7. The DNA sequence encoding the C-terminal 180 amino acids of Myc-Sam68 was generated by inverse PCR using Myc-Q-S as the DNA template and T7 promoter primer and 5′-CTT GGT ACC AGA TAT GAT GAG TAC CCC GCC TCG AGC CAA G-3′ and 5′-ATT TGG AGG AGC TTT AGC ACC AGC ATG TAA-3′ as primers (underlined nucleotides denote changes introduced). Myc-Q-S:E was constructed using the same strategy as that of Myc-Q-S except that Myc-QKI-7:E48G, instead of Myc-QKI-7, was used to subclone the Sam68 DNA fragment. The construct encoding Myc-Q-S:A:315 was generated by inverse PCR using Myc-Q-S as the DNA template and 5′-ATT CAA CTT GAA GCA GAA AGC AAG GGA-3′ and 5′-ATT TGT AAG TCC TCT AGG TAC CGG TCT TTT AAT GCT ATA-3′ as primers (underlined nucleotides denote changes introduced). The Myc-Q-S:315RG construct was created by site-directed mutagenesis of Myc-Q-S:A:315, which prevents the expression of 3′-UTR, (108) 3′-UTR, which was kindly provided by Tim Schedl (Washington University).

The identities of all the above plasmid constructs were verified by digestion with oligonucleotide sequencing.

Protein Expression and Analysis—Proteins were expressed in HeLa cells, using the vaccinia virus T7 expression system as described previously (5). HeLa cells were lysed in lysis buffer (1% Triton X-100, 150 mM NaCl, and 20 mM Tris-Cl (pH 8.0)), 50 mM NaF, 100 mM sodium vanadate, 0.01% phenylmethanesulfonyl fluoride, 1 μM of aprotinin/ml, and 1 μg of leupeptin/ml, and the cellular debris and nuclei were removed by centrifugation. For immunoprecipitation, the supernatant was subjected to antibody precipitation and the resulting immune complexes were eluted with glutathione beads. The eluted polypeptides were analyzed by SDS-PAGE and transferred to nitrocellulose. The immune complexes were analyzed by Western blotting.
was incubated on ice with the specified antibody for 1 h. Then 20 μl of a 50% protein A-Sepharose slurry was added and incubated at 4 °C for 30 min with constant end-over-end mixing. The beads were washed twice with lysis buffer and once with PBS. Protein samples were analyzed on SDS-polyacrylamide gels and transferred to nitrocellulose membranes. Immunoblotting was performed using the anti-Myc (9E10), anti-hemagglutinin (HA), or anti-p59 \(^{\text{fyn}}\) antibodies. The rabbit anti-p59 \(^{\text{fyn}}\) antibody was provided kindly by André Veillette (Institut de Recherche Clinique de Montréal, Université de Montréal, McGill University). The designated primary antibody was followed by goat anti-mouse or goat anti-rabbit antibodies conjugated to horse radish peroxidase (ICN), and chemiluminescence was used for protein detection (DuPont).

In Vitro Transcription—a\(^{32}\)P-labeled G8–5 RNA and tra-2 3'-UTR RNA were transcribed in vitro with the T7 RNA polymerase following the protocols recommended by the manufacturer (Promega). After in vitro transcription, the template DNA was digested with DNase I (Promega). The designated primary antibody was followed by goat anti-Myc antibodies. For G8–5 or tra-2 3'-UTR RNA binding, Myc-tagged proteins expressed in HeLa cells were immunoprecipitated with an anti-Myc antibody or mouse IgG (control), and the immunoprecipitates were incubated at 4 °C for 30 min with 1 μl (10\(^6\) cpm) of a\(^{32}\)P-labeled RNA in lysis buffer supplemented with 2 mg/ml heparin. The beads were washed twice with lysis buffer and once with PBS, and the bound radioactivity was quantitated by scintillation counting. To verify the identity of the radiolabeled RNA bound to beads, the bound RNA was eluted with sample buffer and analyzed with non-denaturing polyacrylamide gel electrophoresis and autoradiography. To verify protein expression, the immunoprecipitates were analyzed by immunoblotting with anti-Myc antibody. REV assays were performed as described previously (7).

RESULTS

STAR proteins contain a GSG domain, which is a tripartite protein module containing from N to C terminus the NK region, the KH domain, and the CK region (Fig. 1B). The Sam68 KH domain is necessary for RNA binding because its deletion prevents RNA binding (10, 12, 30). To investigate whether the RNA binding specificity of the STAR proteins resides only in the KH domain or whether neighboring regions can regulate RNA binding, we constructed chimeric proteins between QKI-7 and Sam68. These two proteins show a high degree of homology in their GSG domain (Fig. 1A) but possess distinct RNA binding specificities. Sam68 has been shown to bind homopolymeric RNA poly(U) and poly(A) (2, 10) as well as a synthetic RNA (G8–5) amplified by using systematic evolution of ligands by exponential enrichment (12). The QKI proteins bind C. elegans GLD-1 target, tra-2 (41), but not homopolymeric RNA (10). Chimeric proteins were generated between Sam68 and QKI-7 in such a way that the CK region and the C terminus of one protein was replaced by the corresponding region of the other protein (Fig. 1B). Q-S (QKI-7-Sam68 chimeric protein) contains the NR region and the KH domain of QKI-7 and the CK region and the C terminus of Sam68 (Fig. 1B). S-Q (Sam68-QKI-7 chimeric protein) contains the Sam68 N-terminal portion, the NR region and the Sam68 KH domain, and the QKI-7 C terminus including the CK region.

The C-terminal Portion of Sam68 Confers Poly(U) Binding to QKI-7—To examine the RNA binding specificity of these chimeric proteins, Myc-tagged Q-S and S-Q were expressed in HeLa cells and tested for their ability to bind poly(U)-Sepharose. Q-S bound poly(U)-Sepharose, whereas S-Q did not (Fig. 2A, lanes 9 and 12, upper panel). This difference in binding was also observed if the Q-S and S-Q proteins were incubated together with the same poly(U)-Sepharose beads, hence eliminating the possibility of a recovery problem (Fig. 2A, lanes 22–24). As positive and negative controls, respectively, Sam68A1–67 (herein renamed S\(\alpha\)N, Fig. 1B) bound poly(U) and QKI-7 did not (Fig. 2A, upper panel, lanes 6 and 3, respectively). We used S\(\alpha\)N in this assay because it has been shown to bind poly(U) as well as full-length Sam68, and its poly(U) binding activity is regulated by p59 \(^{\text{fyn}}\) (11). We have shown previously that the introduction of QKI-7 glutamic acid 48 to glycine in the NR region prevents QKI-7 dimerization (19). To examine whether dimerization via the QKI-7 NR domain was required for Q-S to bind poly(U)-Sepharose, we introduced the E48G amino acid substitution in Q-S. Q-S:E\(\text{G}\) was similar to that observed for Q-S, truncating the proteins at Sam68 amino acids 330 and 294. The chimeric protein S-Q contains the N-terminal portion of S\(\alpha\)N and the C-terminal portion of QKI-7. S-Q205 and S-Q285 are truncated proteins of S-Q, deleting the QKI-7 amino acids at 205 and 285, respectively.

**FIG. 1. Schematic representation of protein constructs.** A, the amino acid sequence homology between Sam68 and QKI-7 is shown. An asterisk denotes identity, two dots are representative of semi-conservative amino acids, and a single dot signifies lower conservation. B, the QKI-7 protein sequence is shown in black, and the Sam68 sequence is shown in white. The GSG domain consists of a KH domain flanked by the NK and CK regions. The chimeric protein Q-S contains the N-terminal portion of QKI-7 and the C-terminal portion of Sam68. Q-S:E\(\text{G}\) with Q-S:E\(\text{G}\) is identical to Q-S, or that the lethal point mutation E48G or a point mutation (A110N) that inactivates the KH domain (equivalent to the fragile X syndrome protein I304N) was introduced in the GSG domain (indicated by a vertical line). Q-S330 and Q-S294 are C-terminal deletion mutants of Q-S, truncating the proteins at Sam68 amino acids 330 and 294. The chimeric protein S-Q contains the Sam68 amino acids 330 and 294. The chimeric protein S-Q contains the QKI-7 amino acids at 205 and 285, respectively.
immunoblotted with anti-p59fyn antibodies (Fig. 2A, lane 6, lower panel). These findings indicated that the poly(U) binding activity of Q-S and Q-S:E–G was regulated by p59fyn and further demonstrated that Q-S behaved like Sam68. The poly(U) binding of Q-S:330 was not affected by p59fyn (Fig. 2A, lane 18, lower panel). This finding was expected, because Q-S:330 does not contain the phosphorylation sites for p59fyn that reside in the C terminus of Sam68 (5). The RNA binding activity of Sam68 was inhibited by the expression of p59fyn and served as a positive control for the assay (Fig. 2A, lane 6, lower panel). The expression of p59fyn was confirmed by immunoblotting an aliquot of total cell lysate corresponding to Fig. 2A with anti-p59fyn antibodies (Fig. 2B). Because the C terminus of Sam68 harbors a regulatory domain that can abrogate RNA binding when phosphorylated by p59fyn on tyrosine residues (11), it was conceivable that the C-terminal region of QKI-7 in S-Q may be inhibiting the ability of the Sam68 KH domain from binding poly(U). To eliminate this possibility, we made C-terminal deletions in S-Q (Fig. 1B, S-Q:284, and S-Q:205) and measured their ability to interact with poly(U)-Sepharose. None of these chimeric proteins bound poly(U)-Sepharose (Fig. 2C: individually, lanes 1–12; mixed, lanes 13–15), demonstrating that the C-terminal region of QKI-7 does not harbor a sequence that inhibits RNA binding.

An isoleucine to asparagine substitution in the second KH domain of the fragile X mental retardation gene product (FMRP) is sufficient to severely impair RNA binding (42). The equivalent amino acid substitution in QKI-7, alanine 110 to asparagine (A→N), was introduced in Q-S to examine the contribution of the QKI-7 KH domain in the poly(U) binding observed with Q-S. Q-S:A→N had impaired poly(U) binding compared with wild-type Q-S (Fig. 2D), suggesting that the QKI-7 KH domain is required for Q-S poly(U) binding. These findings suggested that the C-terminal sequences of Sam68 are able to confer a new RNA binding activity to the QKI-7 KH domain.

The Sam68 C-terminal Region Confers G8–5 RNA Binding—

The physiological RNA targets for Sam68 are unknown, but a degenerate RNA sequence containing a UAAA motif called G8–5 has been identified by systematic evolution of ligands by exponential enrichment that binds Sam68 with high affinity (12). Myc-Sam68, -QKI-7, -S-Q, and -Q-S expressed in HeLa cells were immunoprecipitated with anti-Myc antibodies or control mouse IgG, and the immunoprecipitates were incubated with in vitro transcribed 32P-labeled G8–5 RNA. The immunoprecipitates were washed, and the amount of bound RNA was quantitated and expressed as counts per minute (Fig. 3A). The radioactivity bound by Sam68 and Q-S anti-Myc immunoprecipitates was 15–20 times higher than control immunoprecipitates, whereas there was only a 2–3-fold difference between anti-Myc and control immunoprecipitates of QKI-7 and S-Q (Fig. 3A). The bound RNAs were analyzed with non-denaturing polyacrylamide electrophoresis and visualized by autoradiography to verify that the radioactivity correlated with 32P-labeled G8–5. The G8–5 RNA was observed in Sam68 and Q-S Myc immunoprecipitates (Fig. 3B), confirming that Sam68 and Q-S bound G8–5. The absence of G8–5 binding with QKI-7 and S-Q was not caused by a lower expression of these proteins, because anti-Myc immunoblotting of the immunoprecipitates showed comparable expression of the Myc-tagged proteins (Fig. 3C). These data are consistent with the poly(U) binding results shown in Fig. 2, confirming that the Q-S chimeric protein has an RNA binding specificity similar to Sam68.

G8–5 RNA Binding Activity of Sam68 and Q-S Is Regulated by p59fyn—We investigated whether the ability of Sam68 and Q-S to bind G8–5 was regulated by p59fyn. Although G8–5 is a known high affinity RNA target for Sam68, it is not known...
whether G8–5 RNA binding is regulated by tyrosine phosphorylation. Myc-Sam68, -ΔN, and -Q-S were transfected in HeLa cells with or without p59\textsuperscript{fyn}, the cells were lysed, and the lysates were immunoprecipitated with anti-Myc or control IgG antibodies. The immunoprecipitated proteins were subsequently incubated with \textsuperscript{32}P-labeled G8–5 RNA, and the beads were washed, counted in a scintillation counter, and expressed in counts per minute. The G8–5 binding activity of full-length Sam68, ΔN, and Q-S was inhibited by co-expression of p59\textsuperscript{fyn} (Fig. 4A). The expression of p59\textsuperscript{fyn} nearly abolished G8–5 binding to ΔN, whereas G8–5 binding to Sam68 and Q-S was reduced by \textasciitilde 50\% (Fig. 4A). These findings suggested that the Sam68 N-terminal 67 amino acids regulate the ability of Src kinases to negatively abrogate RNA binding. Equivalent Myc and p59\textsuperscript{fyn} expression were observed in the different samples (Fig. 4, B and C). These data further demonstrated that Q-S behaved like Sam68.

The RG Repeats in Sam68 Are Necessary for Poly(U) Binding—The minimal region of Sam68 required to confer poly(U) RNA binding to QKI-7 resided in Sam68 amino acids 256–330, which included the CK region (Fig. 1B, Q-S330). To verify whether the Sam68 sequence harboring the CK region or the RG repeats within the additional 50 amino acids was responsible for the new specificity of the Q-S chimera, a new chimeric protein was constructed that extended the QKI-7 sequences to include its CK region. This chimeric protein named Q(GSG)-S, which now contained the entire QKI-7 GSG domain, was tested for its ability to bind poly(U)-Sepharose. Q(GSG)-S retained the ability to bind poly(U) to the same extent as ΔN or Q-S (Fig. 5A). Thus the sequences C-terminal of the Sam68 GSG domain, and not the Sam68 CK region, were capable of conferring poly(U) binding to the QKI-7 GSG domain. The minimal region essential to confer poly(U) binding specificity to QKI-7 was mapped by engineering chimeric proteins where the junction between QKI-7 and Sam68 sequences was gradually displaced toward the C terminus. The chimeric proteins were named according to the number of RG repeats that were deleted (Fig. 5A: Q(GSG)-S\textsuperscript{4RG}, S\textsuperscript{3RG}; Δ6RG, and Δ11RG). These chimeras were expressed in HeLa cells and tested for their ability to bind poly(U)-Sepharose. The deletion of four or six RG repeats had little or no effect on the ability of the QKI-7-Sam68 chimeras to bind poly(U)-Sepharose (Fig. 5A: Q(GSG)-S\textsuperscript{4RG} and Δ6RG). In contrast, a larger truncation deleting 11 RG repeats had little or no effect on the ability of the QKI-7-Sam68 chimeras to bind poly(U)-Sepharose (Fig. 5A: Q(GSG)-S\textsuperscript{11RG}). The minimal Sam68 sequence capable of changing the RNA binding specificity of the QKI-7 GSG domain was located between amino acids 308–333, which harbors five RG repeats. This 26-amino acid sequence from Sam68 was
Sam68 binds very strongly to the poly(U)-Sepharose resin. The upper portion of the nitrocellulose membrane was independently probed with an antiserum against Sam68, showing that under the conditions used, endogenous Sam68 binds very strongly to the poly(U)-Sepharose resin.

introduced at a similar position in QKI-7. The plasmid expressing this QKI-5RG chimeric protein was transfected in HeLa cells and examined for its ability to bind poly(U)-Sepharose. The chimeric protein QKI-5RG bound poly(U)-Sepharose (Fig. 5A). These findings demonstrate that the Sam68 26 amino acids spanning amino acids 308–333 are sufficient for conferring homopolymeric RNA binding to the QKI-7 GSG domain. As a control for loading and recovery, a representative poly(U) binding reaction with Myc-tagged QKI-7 was reimmunoblotted using anti-Sam68 antibodies, confirming that endogenous Sam68 bound poly(U) and not the epitope-tagged Myc-QKI-7 (Fig. 5B, upper half).

Three of the five RG repeats in the Sam68 sequences spanning amino acids 308–333 were replaced with alanine and serine residues, respectively. The Sam68 NG→AS mutant proteins were expressed in HeLa cells, and their capacity to bind poly(U) was examined (Fig. 6). If the RG repeats participate in conferring homopolymeric RNA binding to the Sam68 GSG domain, the removal by amino acid substitution or deletion should prevent poly(U) binding. The substitution of the individual RGs at position 315, 320, or 325 had a comparable effect such that poly(U) binding was reduced by more than 50% (Fig. 6: Sam68 3RG). When the double substitution of 315 and 325 or the deletion of amino acids 315–325 was performed, poly(U) binding was completely abrogated (Fig. 6: 315,325RG). These data demonstrate that the RG repeats at positions 315 and 325 are necessary for Sam68 poly(U) binding.

The deletion of amino acids 315–325 was performed next in the context of the full-length Sam68 protein, and its ability to bind poly(U)-Sepharose was examined (Fig. 6: Sam683RG). Sam683RG bound poly(U)-Sepharose with wild-type affinities unlike Sam68 3RG (Fig. 6). These findings suggested that the N-terminal sequences of Sam68 function in a redundant manner with the residues located between 315 and 325 to confer poly(U) binding to the Sam68 GSG domain. The Sam68 N-terminal 67 amino acids (1, 5) harbor several individual arginine residues and a motif that matches the consensus of an RGG box, a type of RNA binding motif (43). To identify the N-terminal sequence required to confer poly(U) binding to the Sam68 GSG domain, a series of mutant proteins was engineered in the Sam683RG “background.” Sam683RG was chosen to eliminate any contribution from the C-terminal 315–325 amino acids. Alteration of arginines 10, 13, and 17 to alamines had a minor effect on poly(U) RNA binding (Fig. 6: R→A:10, 13, 17). In contrast, substitution of arginines at position 43 (just outside the RGG box) and 45 (eliminating the first RGG repeat) to alanines was sufficient to severely impair the interaction with poly(U) RNA (Fig. 6: R→A:43, 45). Removing the second RGG sequence as well as the adjacent arginine had little or no effect on poly(U) binding (R→A:52, 56), suggesting that the first RGG sequence plays a major role in conferring poly(U) specificity to the Sam68 GSG domain.

The RG repeats in Sam68 are necessary for poly(U) binding. A schematic diagram representing the various Myc-tagged proteins used is shown on the left. RG repeats at positions 315, 320, and 325 were substituted with alanine serine, as indicated in the sequence below the Sam68 representation. Similarly, arginine residues in the N-terminal domain of Sam68 were replaced by alanine residues, in the context of a protein in which amino acids 315–325 have been deleted (Sam68 3RG). Dashes represent amino acids that have been deleted as in Sam68 3RG. Each plasmid construct was transfected in HeLa cells, the cells were lysed, and the lysates were incubated with Sepharose control (C) or poly(U)-Sepharose (pU). The bound proteins were separated by SDS-PAGE, transferred, and immunoblotted with anti-Myc antibodies shown on the right.

The RG repeats in Sam68 are necessary for poly(U) binding. A schematic diagram representing the various Myc-tagged proteins used is shown on the left. Each plasmid construct was transfected in HeLa cells, the cells were lysed, and the lysates were incubated with control (C) or poly(U)-Sepharose (pU). The bound proteins were separated by SDS-PAGE, transferred, and immunoblotted with anti-Myc antibodies shown on the right. The upper portion of the nitrocellulose membrane was independently probed with an antiserum against Sam68, showing that under the conditions used, endogenous Sam68 binds very strongly to the poly(U)-Sepharose resin. The upper portion of the nitrocellulose membrane was independently probed with an antiserum against Sam68, showing that under the conditions used, endogenous Sam68 binds very strongly to the poly(U)-Sepharose resin.
expression vector. CAT activity was normalized for 

\[
\text{CAT activity} = \frac{\text{Transfection efficiency}}{\text{pCH110 cated expression vectors}} \times \frac{\text{pCH110 activity}}{\text{Transfection efficiency}}
\]

tution for Rev in an HIV RNA export assay. Above calculated from at least five distinct experiments. Shown is a rep-

\[\text{Fig. 8,} \quad 18\]

Surprisingly, Q-S associated with both Sam68 and QKI-7

\[\text{lane 15}\]

consistent with the Sam68 associated with Sam68 (Fig.

\[\text{lane 33}\]

association with Sam68 (Fig. 8, lane 33) but maintained the associa-

\[\text{Q-S} (\text{Q-S}294, \text{data not shown})\]

The chimeric protein harboring the C-terminal 149 amino acids of Sam68 were deleted from

\[\text{QKI-7 lethal mutation Q-S:E} \rightarrow \text{G} \]

did not associate with QKI-7, but the association with Sam68 remained intact (lane 30), consistent with the idea that the NK region of QKI-7 mediates the association with QKI-7. In summary, the chimeric proteins associated with QKI-7 in a predicted fashion: if the NK region of QKI-7 was present, there was association with QKI-7, and if the NK region was absent or if it contained the lethal point mutation E48G, there was no association with QKI-7. The association with Sam68 was more complex. Both the known region for self-association, namely the KH domain, and a newly identified region located in the C-terminal 113 amino acids were involved in association with Sam68. Thus the usage of chimeric proteins has permitted the discovery of a region in the C terminus of Sam68 that is involved in multimerization that was not observed by deletion and/or mutation analysis (10).

\[\text{Tra-2 3'-UTR RNA Binding Activity of Sam68, QKI-7, and Chimeric Proteins—One question that remained unanswered was whether the S-Q chimera had gained a QKI-like RNA binding specificity. In the course of these studies it was shown by Goodwin and co-workers (41) that QKI-7 bound the 3'-UTR of C. elegans tra-2. To test whether the chimeric proteins associated with an RNA target bound by QKI-7, Myc-tagged Sam68/QKI-7 chimeric proteins were expressed in HeLa cells and immunoprecipitated with anti-Myc or control IgG, and immunoprecipitates were incubated with \[32\]P-labeled tra-2 3'-UTR RNA or a mutant RNA with a deletion of 108 nucleotides (~108 3'-UTR, Fig. 9C). The amount of bound RNA after several washes was quantitated and expressed as counts per minute. Anti-Myc immunoprecipitates of QKI-7 bound tra-2 (~70,000 cpm, Fig. 9A) but not the ~108 mutant RNA, which is consistent with previous studies (41). In contrast, Sam68, Q-S, and S-Q had negligible tra-2 binding (<1000 cpm, Fig. 9A). The absence of binding with Sam68, Q-S, and S-Q was not caused by a lower expression of these proteins, because anti-Myc immunoblotting of the immunoprecipitates showed comparable expression of all Myc-tagged proteins (Fig.

\[\text{Sam68 Sequences Required for Nonspecific RNA Binding}\]

\[\text{Fig. 7. The RG repeats can confer QKI-7 the ability to substitute for Rev in an HIV RNA export assay. COS-7 cells were transfected with an RRE-CAT reporter plasmid in the presence of the indicated expression vectors as well as the pCH110 β-galactosidase expression vector. CAT activity was normalized for β-galactosidase activity to eliminate transfection efficiency variations. Shown is a rep-}\]

\[\text{Fig. 9. The C-terminal portion of Sam68 harbors a region that mediates self-association. Myc-tagged Sam68, QKI-7, S-Q, Q-S, Q-S:S-E} \rightarrow \text{G}, \text{or Q-S330 were co-transfected with HA-tagged Sam68 or HA-}\]

\[\text{Fig. 8,}\]

FIG. 8.
of Sam68 may contain regulatory sequences. The mouse Sam68 N-terminal 67 amino acids contain two RGG repeats interspaced by four residues (5) and may be a bona fide RGG box, a type of RNA binding motif (43). The deletion of the Sam68 N-terminal 67 amino acids had no effect on poly(U) and G8–5 RNA binding, suggesting that the RGG repeats do not play a major role in RNA binding or that other sequences in Sam68 function in a redundant manner. Indeed, RG sequences on either side of the GSG domain function in a redundant manner to confer poly(U) binding to the GSG domain. Deletion of the N-terminal RGG boxes or deletion of amino acids 315–325 had no effect on poly(U) binding. However, a Sam68 protein containing the double deletion/mutation was unable to associate with poly(U)-Sepharose. The tyrosine phosphorylation of Sam68 by p59fyn has been shown to abrogate poly(U) binding (11), suggesting that the phosphorylated C terminus of Sam68 can negatively regulate the nonspecific RNA binding contributions from RG-rich sequences located from 315–325. The fact that the full-length protein was less affected by phosphorylation by p59fyn suggests several possibilities: 1) the N-terminal RG sequences are not regulated by p59fyn and may require additional signals such as arginine methylation (15), and 2) a tyrosine kinase may phosphorylate different tyrosines on Sam68 that may now regulate the N-terminal RG region. In summary, our analysis has uncovered a role for the N-terminal RGG sequences that was not obvious using simple deletion strategies.

The Q-S chimeric protein has lost its ability to interact with a QKI-7-specific RNA target, the 3' UTR of tra-2 (see Fig. 9). Previous studies demonstrated that the GSG domain was the minimal region required for RNA binding and that GSG proteins devoid of a CK region had impaired RNA binding (19, 21). Our results are consistent with this notion, because we show that if the complete GSG domain of QKI-7 is included in the chimeras (Q(GSG)-S) we now observe specific binding to the tra-2 RNA. Moreover, the Q(GSG)-S chimeric protein still displays Sam68-like RNA binding specificity, namely interaction with poly(U) and G8–5 RNAs. Thus the GSG domain is the only RNA binding region in the QKI-7 protein required for specific high affinity RNA binding.

The construction of chimeric QKI-7/Sam68 proteins has also permitted us to find an oligomerization region located in the Sam68 C-terminal 113 amino acids. Our previous studies demonstrate that the Sam68 GSG domain (Sam68:103–269) is able to associate with a wild-type Sam68 protein and that a deletion in the KH domain of Sam68 (Sam68AKH) abolished self-association (10). These data showed that the GSG domain was necessary and sufficient for self-association and that the Sam68 C-terminal ~200 amino acids in Sam68AKH were not sufficient, without an intact KH domain, to mediate self-association (10). Using QKI-7/Sam68 chimeras, it is evident that the C-terminal region of Sam68 harbors a region required for self-association. In this situation the QKI-7 KH domain seems to compensate for the loss of the Sam68 KH domain. The coiled coil in the NK region of QKI-7 did not participate in the Sam68 association, because the introduction of the E48G substitution (Q-S:E→G), known to abolish the association with QKI-7 (19), did not affect the association with Sam68. The presence of two regions in Sam68 that mediate self-association suggests that Sam68 may be forming head-to-tail multimers and/or that

Fig. 9. tra-2 3' UTR RNA binding activity of Sam68, QKI-7, and chimeric proteins. A. Myc-tagged proteins were expressed in HeLa cells and immunoprecipitated with control mouse IgG or anti-Myc antibodies. The immunoprecipitates were incubated with 32P-labeled wild-type or mutant (−108) tra-2 3'UTR RNA, and the bound radioactivity was quantitated. Each bar represents the mean ± standard deviation of data from more than six independent experiments normalized for the background counts obtained with control mouse IgG. B, a typical representation of the expression of Myc-tagged proteins is shown. C, 32P-labeled wild-type or mutant (−108) tra-2 3'UTR RNAs were produced in vitro using T7 RNA polymerase and resolved on nondenaturing polyacrylamide gels. Molecular size markers are indicated in nucleotides.

DISCUSSION

In the present study, we demonstrate that an RG-rich sequence C-terminal to the Sam68 GSG domain spanning amino acids 508–333 is necessary for poly(U) binding, G8–5 binding, and functionally substituting for REV in the transport of HIV RNAs. Using chimeric proteins, we also demonstrate that these 26 amino acids of Sam68 are sufficient to confer to another unrelated STAR protein, QKI-7, the ability to bind poly(U)-Sepharose and G8–5 and functionally substitute for REV in the transport of HIV RNAs. Because a functional KH domain is required for these three activities of Sam68 (10, 12, 14, 30), it is clear that the RG sequences themselves do not possess intrinsic RNA binding activity. The most likely explanation is that the RG repeats confer to neighboring GSG domains the ability to bind certain RNAs. The collaboration between the RG-rich regions of Sam68 with the GSG domain is reminiscent of a study by Rosbach and co-workers (44) on the GSG domain of BBP/SF1, another STAR protein. BBP/SF1 contains both a GSG domain and a Zn knuckle RNA binding motif. The GSG domain of these proteins is involved in specific recognition of the pre-mRNA branchpoint sequence, but one or more accessory modules is required to achieve efficient binding (44).

The inhibitory effect of p59fyn on Sam68 was more severe than on full-length Sam68, suggesting that the N-terminal region of Sam68 may contain regulatory sequences. The mouse Sam68 N-terminal 67 amino acids contain two RGG repeats interspaced by four residues (5) and may be a bona fide RGG box, a type of RNA binding motif (43). The deletion of the Sam68 N-terminal 67 amino acids had no effect on poly(U) and G8–5 RNA binding, suggesting that the RGG repeats do not play a major role in RNA binding or that other sequences in Sam68 function in a redundant manner. Indeed, RG sequences on either side of the GSG domain function in a redundant manner to confer poly(U) binding to the GSG domain. Deletion of the N-terminal RGG boxes or deletion of amino acids 315–325 had no effect on poly(U) binding. However, a Sam68 protein containing the double deletion/mutation was unable to associate with poly(U)-Sepharose. The tyrosine phosphorylation of Sam68 by p59fyn has been shown to abrogate poly(U) binding (11), suggesting that the phosphorylated C terminus of Sam68 can negatively regulate the nonspecific RNA binding contributions from RG-rich sequences located from 315–325. The fact that the full-length protein was less affected by phosphorylation by p59fyn suggests several possibilities: 1) the N-terminal RG sequences are not regulated by p59fyn and may require additional signals such as arginine methylation (15), and 2) a tyrosine kinase may phosphorylate different tyrosines on Sam68 that may now regulate the N-terminal RG region. In summary, our analysis has uncovered a role for the N-terminal RGG sequences that was not obvious using simple deletion strategies.

The Q-S chimeric protein has lost its ability to interact with a QKI-7-specific RNA target, the 3' UTR of tra-2 (see Fig. 9). Previous studies demonstrated that the GSG domain was the minimal region required for RNA binding and that GSG proteins devoid of a CK region had impaired RNA binding (19, 21). Our results are consistent with this notion, because we show that if the complete GSG domain of QKI-7 is included in the chimeras (Q(GSG)-S) we now observe specific binding to the tra-2 RNA. Moreover, the Q(GSG)-S chimeric protein still displays Sam68-like RNA binding specificity, namely interaction with poly(U) and G8–5 RNAs. Thus the GSG domain is the only RNA binding region in the QKI-7 protein required for specific high affinity RNA binding.

The construction of chimeric QKI-7/Sam68 proteins has also permitted us to find an oligomerization region located in the Sam68 C-terminal 113 amino acids. Our previous studies demonstrate that the Sam68 GSG domain (Sam68:103–269) is able to associate with a wild-type Sam68 protein and that a deletion in the KH domain of Sam68 (Sam68AKH) abolished self-association (10). These data showed that the GSG domain was necessary and sufficient for self-association and that the Sam68 C-terminal ~200 amino acids in Sam68AKH were not sufficient, without an intact KH domain, to mediate self-association (10). Using QKI-7/Sam68 chimeras, it is evident that the C-terminal region of Sam68 harbors a region required for self-association. In this situation the QKI-7 KH domain seems to compensate for the loss of the Sam68 KH domain. The coiled coil in the NK region of QKI-7 did not participate in the Sam68 association, because the introduction of the E48G substitution (Q-S:E→G), known to abolish the association with QKI-7 (19), did not affect the association with Sam68. The presence of two regions in Sam68 that mediate self-association suggests that Sam68 may be forming head-to-tail multimers and/or that
Sam68 is involved in intramolecular interactions. We have shown previously that the tyrosine phosphorylation of Sam68 by p59/56 prevents self-association (10). The mechanism by which tyrosine phosphorylation regulates self-association is unknown. Now with the discovery of a new region that resides in the C-terminal tyrosine-rich region of Sam68, it is possible that the GSG domain associates with a region in the C-terminal tyrosine region and that the phosphorylation of this region by Src kinases would interfere with this self-association, which is consistent with our previous observations (10).

Based on the results presented in this study, we propose the following model: Sam68 in the resting state would be nonphosphorylated and exhibit nonspecific RNA binding mediated by its flanking RG-rich regions. The tyrosine phosphorylation of Sam68 would render it a specific RNA-binding protein. Supporting this model is the fact that poly(U) binding, the UAAA RNA target G8–5, and REV-like function of Sam68 were identified by using unphosphorylated Sam68 (2, 12, 14). Moreover, Sam68 poly(U) binding (11, 45), G8–5 (this study), and REV-like function (7) are negatively regulated by tyrosine kinases. The absence of secondary structure in poly(U) and the systematic evolution of ligands by exponential enrichment RNAs such as G8–5 and the lack of defined sequence/structure in the RRE recognized by Sam68 suggest that specific RNA binding, mediated by only the Sam68 GSG domain, may only be observed in the context of the whole protein when tyrosine-phosphorylated.

REFERENCES
1. Wong, G., Muller, O., Clark, R., Conrey, L., Moran, M. F., Polakis, P., and McCormick, F. (1992) Cell 69, 551–558
2. Taylor, S. J., and Shalloway, D. (1994) Nature 368, 867–871
3. Fumagalli, S., Totty, N. F., Hsuan, J. J., and Courtneidge, S. A. (1994) Nature 368, 871–874
4. Weng, Z., Thomas, S. M., Rickles, R. J., Taylor, J. A., Brauer, A. W., Seidel-Dugan, C., Michael, W. M., Dreyfuss, G., and Brugge, J. S. (1994) Mol. Cell. Biol. 14, 4509–4521
5. Richard, S., Yu, D., Blumer, K. J., Hausladen, D., Olazowy, M. W., Connely, P. A., and Shaw, A. S. (1995) Mol. Cell. Biol. 15, 186–197
6. Vogel, L. B., and Fujita, D. J. (1995) J. Biol. Chem. 270, 2506–2511
7. Derry, J. J., Richard, S., Valderrama Carvajal, H., Ye, X., Vasioukhin, V., Cochrane, A. W., Chen, T., and Tyner, A. L. (2000) Mol. Cell. Biol. 20, 6114–6126
8. Lang, V., Mege, D., Semichon, M., Gary-Gooux, H., and Bismuth, G. (1997) Eur. J. Immunol. 27, 3360–3367
9. Taylor, S. J., Anafi, M., Pawson, T., and Shalloway, D. (1995) J. Biol. Chem. 270, 10120–10124
10. Chen, T., Damaj, B. B., Herrera, C., Lasko, P., and Richard, S. (1997) Mol. Cell. Biol. 17, 5797–5718
11. Wang, L. L., Richard, S., and Shaw, A. S. (1995) J. Biol. Chem. 270, 2010–2013
12. Lin, Q., Taylor, S. J., and Shalloway, D. (1997) J. Biol. Chem. 272, 27274–27280
13. Chen, T., Boisvert, F. M., Bazett-Jones, D. P., and Richard, S. (1999) Mol. Biol. Cell 10, 3015–3033
14. Reddy, T. B., Xu, W., Mau, J. K., Goodwin, C. D., Suhasini, M., Tang, H., Frimpong, K., Rose, D. W., and Wong-Staal, F. (1999) Nat. Med. 5, 635–642
15. Bedford, M. T., Frankel, A., Yaffe, M. B., Clarke, S., Leder, P., and Richard, S. (2000) J. Biol. Chem. 275, 16030–16038
16. Hartmann, A. M., Nayler, O., Schaiger, F. W., Omer, A., and Stamm, S. (1999) Mol. Biol. Cell 10, 3909–3926
17. Jones, A. R., and Shedl, T. (1995) Genes Dev. 9, 1491–1504
18. Di Fuscoio, M., Chen, T., Bonyadi, S., Lasko, P., and Richard, S. (1998) J. Biol. Chem. 273, 30122–30130
19. Chen, T., and Richard, S. (2000) Mol. Cell. Biol. 18, 4863–4871
20. Di Fuscoio, M., Chen, T., and Richard, S. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 2710–2715
21. Rain, J. C., Raff, Z., Rhani, Z., LeGrain, P., and Kramer, A. (1998) RNA (N. Y.) 4, 551–565
22. Zorn, A. M., and Krieg, P. A. (1996) Trends Genet. 13, 2176–2190
23. Wu, J., Zou, L., Tonissen, K., Tee, R., and Artzt, K. (1999) J. Biol. Chem. 274, 29202–29210
24. Ebersole, T. A., Chen, Q., Justice, M. J., and Artzt, K. (1996) Nat. Genet. 12, 260–265
25. Vernet, C., and Artzt, K. (1997) Trends Genet. 13, 479–484
26. Arning, S., Gruter, P., Bilbe, G., and Kramer, A. (1996) RNA (N.Y.) 2, 794–810
27. Berglund, J. A., Abovich, N., and Rosbash, M. (1998) Genes Dev. 12, 658–687
28. Liu, K., Li, L., Nisson, P. E., Gruber, C., Jessee, J., and Cohen, S. N. (2000) J. Biol. Chem. 275, 40195–40201
29. Plitte, J., Larroque, D., and Richard, S. (2001) Genes Dev. 15, 845–858
30. Barlat, I., Maurier, F., Duchesne, M., Guitard, E., Tocque, B., and Schweighoffer, F. (1997) J. Biol. Chem. 272, 3132–3132
31. Jan, E., Mototzny, C. K., Graves, L. E., and Goodwin, E. B. (1999) EMBO J. 18, 288–298
32. Clifford, R., Lee, M. H., Nayak, S., Ohmachi, M., Giorgini, F., and Shedl, T. (2000) Development 127, 5385–5376
33. Raebereek, E. H. (1997) Development 124, 1333–1332
34. Zaffran, S., Astier, M., Gratecos, D., and Semeriva, M. (1997) Development 124, 2097–2098
35. Francis, R., Maine, E., and Shedl, T. (1995) Genetics 143, 667–670
36. Francis, R., Bartk, M. K., Kimble, J., and Shedl, T. (1995) Genetics 143, 579–606
37. Bode, V. C. (1984) Genetics 108, 457–470
38. Hogan, B. L., and Greenfield, S. (1984) Myelin, Plenum Publishing Corp., New York
39. Hardy, R. J., Loshin, C. L., Friedrich, V. L., Chen, Q., Ebersole, T. A., Lazzarini, R. A., and Artzt, K. (1996) J. Neurosci. 16, 7941–7949
40. Sidman, R. L., Dickie, M. M., and Appel, S. H. (1964) Science 144, 309–311
41. Saccomanno, L., Loshin, C., Jan, E., Punkay, E., Artzt, K., and Goodwin, E. B. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 12605–12610
42. Suiii, H., Chai, M., Suiii, M. C., Nussbaum, B. L., and Dreyfuss, G. (1994)
43. Burd, C. G., and Dreyfuss, G. (1994) Science 265, 615–621
44. Berglund, J. A., Fleming, M. L., and Rosbash, M. (1996) RNA 4, 998–1006
45. Gilbert, C., Barabe, F., Bollet-Labelle, E., Bourgois, S. G., McColl, S. R., Damaj, B. B., and Naccache, P. H. (2001) J. Immunol. 166, 4664–4671
Identification of Sam68 Arginine Glycine-rich Sequences Capable of Conferring Nonspecific RNA Binding to the GSG Domain
Taiping Chen, Jocelyn Côté, Héctor Valderrama Carvajal and Stéphane Richard

J. Biol. Chem. 2001, 276:30803-30811.
doi: 10.1074/jbc.M102247200 originally published online June 6, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M102247200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 44 references, 36 of which can be accessed free at http://www.jbc.org/content/276/33/30803.full.html#ref-list-1