The Identification of Two Drosophila K Homology Domain Proteins

KEP1 AND SAM ARE MEMBERS OF THE SAM68 FAMILY OF GSG DOMAIN PROTEINS*

(Received for publication, January 15, 1998, and in revised form, August 25, 1998)

Marco Di Fruscio‡§, Taiping Chen‡¶, Sheida Bonyadi*, Paul Lasko**†, and Stéphane Richard‡‡

From the §Terry Fox Molecular Oncology Group, Lady Davis Institute for Medical Research, Sir Mortimer B. Davis Jewish General Hospital, and Departments of Oncology, Medicine and Microbiology and Immunology, McGill University, Montreal, Quebec H3T 1E2 and the ‡Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada

Sam68 is a member of a growing family of RNA-binding proteins that contains an extended K homology (KH) domain embedded in a larger domain called the GSG (GRP33, Sam68, GLD1) domain. To identify GSG domain family members, we searched data bases for expressed sequence tags encoding related portions of the Sam68 KH domain. Here we report the identification of two novel Drosophila KH domain proteins, which we termed KEP1 (KH encompassing protein) and SAM. SAM bears sequence identity with mammalian Sam68 and may be the Drosophila Sam68 homolog. We demonstrate that SAM, KEP1, and the recently identified Drosophila Wh/How are RNA-binding proteins that are able to self-associate into homomultimers. The GSG domain of KEP1 and SAM was necessary to mediate the RNA binding and self-association. To elucidate the cellular roles of these proteins, SAM, KEP1, and Wh/How were expressed in mammalian and Drosophila S2 cells. KEP1 and Wh/How were nuclear and SAM was cytoplasmic. The expression of KEP1 and SAM, but not Wh/How, activated apoptotic pathways in Drosophila S2 cells. The identification of KEP1 and SAM implies that a large GSG domain protein family exists and helps redefine the boundaries of the GSG domain. Taken together, our data suggest that KEP1 and SAM may play a role in the activation or regulation of apoptosis and further implicate the GSG domain in RNA binding and oligomerization.

The K homology (KH)† domain is a protein module consisting of 70–100 amino acids that was originally identified as a repeated sequence in heterogeneous nuclear ribonucleoprotein K (1). The KH domain is an RNA binding motif that is thought to make direct protein-RNA contacts with a three-dimensional βαβαβα-βα-fold (2). Alignment of KH domains from various proteins (2, 3) reveals a subfamily of KH domains including mammalian Sam68 (4, 5), Artemia salina GRP33 (6), Caenorhabditis elegans GLD-1 (7), mouse Qk1 (8), Xenopus Qk1 (9), mammalian SF1 (10), and the Drosophila Wh/How (11–13). All these proteins contain a single extended KH domain included within a larger protein domain of ~200 residues called the GSG (GRP33, Sam68, GLD-1) domain (7). This domain is also called STAR for signal transduction and activator of RNA (14). The KH domain embedded in the GSG domain is approximately 26 amino acids longer than other KH domains with extra amino acids located in loops 1 and 4 (2). Although the function of the GSG domains remains unknown, this protein module has been shown to be required for self-association and RNA binding (15–18).

There is considerable genetic evidence from various species supporting the physiological role of the KH domain. In humans, gene lesions that prevent expression of the KH protein FMR1 result in the fragile X mental retardation syndrome (19, 20), the most common form of heritable mental retardation (21). The particular significance of the KH domain was implicated by a point mutation altering a conserved isoleucine 304 to asparagine in the second KH domain of FMR1 (22). The point mutation alters the structure of the KH domain (2) and severely impairs RNA binding activity (23). In C. elegans, GLD-1 is a cytoplasmic protein required for germ cell differentiation (24–26) and alteration of glycine 227, within the KH domain, results in a recessive tumorous germ line phenotype (7). The structure of the KH domain predicts that this conserved glycine forms part of the RNA binding surface (2) and mutation of the corresponding residue in Sam68 abolishes RNA binding (16). In mice, the quaking viable mutation severely impairs myelination and as a result, mice develop a rapid tremor at postnatal day 10 (27). A missense mutation in the GSG domain of Qk1 is embryonic lethal (8). This point mutation has been shown to prevent homodimerization and may be the reason for the lethality observed in mice (18). The Drosophila Bicaudal C (Bic-C) contains five KH domains and gene lesions that truncate the Bic-C protein or a point mutation that replaces glycine 295 with an arginine in the third KH domain are strong alleles, leading to defects in RNA binding, oogenesis, and anterior-posterior embryonic patterning (28, 29). Drosophila Wh/How is expressed in muscle and animals that possess weak mutations either die or survive as adults with defects in wing position (11–13). A point mutation in the KH domain of Wh/How has been identified and results in flies with “wings held out” (11).

In this study, we report the cloning of two new Drosophila GSG domain family members that we have called SAM and

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF096272 (SAM) and AF096273 (KEP1).

‡ Supported in part by funds from the Gertrude and Charles Clark Cancer Research Fellowship.

¶ Supported by a studentship from the Cancer Research Society.

** Supported by grants from the National Cancer Institute of Canada (NCIC) and the Cancer Research Society; Research Scientist of the NCIC.

†† Supported by grants from the Cancer Research Society and Medical Research Council of Canada; Scholar of the Medical Research Council of Canada. To whom correspondence should be addressed: Molecular Oncology Group, Lady Davis Institute, 3755 Côte Ste-Catherine Rd., Montreal, Quebec H3T 1E2, Canada. Tel.: 514-340-8260; Fax: 514-340-7576; E-mail: mcrd@musica.mcgill.ca.

† The abbreviations used are: KH, homology; kb, kilobase pair(s); GFP, green fluorescent protein; PCR, polymerase chain reaction; PAGE, polyacrylamide gel electrophoresis; HA, hemagglutinin; DAPI, 4,6-diamidino-2-phenylindole.
KEP1. SAM bears homology to the mammalian Sam68 and may be its homolog, whereas KEP1 is a unique family member. SAM is a cytoplasmic protein expressed in early embryos and female flies. KEP1 is a nuclear protein that is expressed at all stages of Drosophila development. KEP1 and SAM are RNA-binding proteins that bind homopolymeric RNA in vitro and self-associate into multimers. We further show that extracts from KEP1- and SAM-transfected Drosophila S2 cells trigger chromatin condensation and apoptotic body formation, in vitro events known to be caspase-dependent (30, 31).

EXPERIMENTAL PROCEDURES

DNA Constructs and Cloning of KEP1, SAM, and Who/How—To obtain a full-length KEP1 cDNA clone, we screened a 0.2–h embryonic Drosophila library constructed in the λ ZAP vector. A 600-base pair BamHI fragment isolated from expressed sequence tag LD14468 (GenBank accession AA439802, Genome Systems) was used as a probe. The longest clone, 9–12, was obtained and contained a 1.4-kb insert. The Bank accession AA439802, Genome Systems) was used as a probe. The Bam KEP1 and SAM32P probes were generated by PCR using the above tems). The oligonucleotides utilized for the PCR were 5'-AAA GGA TCC TTC ATG ACC GAG AAG TAC GAC GCC C-3' and 5'-TTC GAC ATT GTT GCC TCG CTG CAT GTG-3' promoter primer. The DNA fragment was digested with EcoRI and XhoI (located at the 3' of the insert) and subcloned in the corresponding sites in HA- and Myc-Bluescript (16, 32). The expression sequence tag LD08190 (GenBank accession AA264447, Genome Systems) contained the entire coding region of SAM. This region was PCR-amplified using the 5'-GGC TCA GAT AAA AAT GGA AAC CCC AAG C-3' and 5'-TTC GAC ATT GTT GCC TCG CTG CAT GTG-3' promoter primer. The DNA fragment was digested with EcoRI and XhoI and subcloned in HA- and Myc-Bluescript (16, 32). The entire open reading frame of Who/How was amplified by PCR using an expressed sequence tag (GenBank accession AA439173, Genome Systems). The oligonucleotides utilized for the PCR were 5'-GGG CAA TTC TTC TAT GCA GAA GAG AAG TAC GAC GCC C-3' and 5'-GGG CAA TTC TTC TAT GCA GAA GAG AAG TAC GAC GCC C-3', respectively. The DNA fragment was digested with EcoRI and subcloned in HA- or Myc-Bluescript (16, 32). The DNA inserts were sequenced manually with Sequenase (U.S. Biochemical Corp.) or by the Sheldon Biotechnology Institute automated sequencing facility (McGill University). All sequence data were compilations of multiple reads on both strands.

GSS constructs were generated by PCR using the following oligonucleotides. The forward oligonucleotide for all the SAM constructs was 5'-AAA GAA TTC ATG ACC GAG AAG TAC GAC GCC C-3', whereas the reverse oligonucleotides used to generate deletions were 5'-GGG CAA TTC TTC TAT GCA GAA GAG AAG TAC GAC GCC C-3', 5'-GGG CAA TTC TTC TAT GCA GAA GAG AAG TAC GAC GCC C-3', and 5'-GGG CAA TTC TTC TAT GCA GAA GAG AAG TAC GAC GCC C-3'. The amplified DNA fragment was digested with EcoRI and subcloned in the corresponding site in Myc-Bluescript. The DNA for SAM 233–428 was PCR-amplified with T7 promoter primer and 5'-GGG CAA TTC TTC TAT GCA GAA GAG AAG TAC GAC GCC C-3' using LD08190 as a DNA template. The amplified DNA fragment was digested with EcoRI and XhoI and subcloned in Myc-Bluescript.

Mammalian green fluorescent protein (GFP) fusion proteins were generated as follows. Myc-SAM and Myc-KEP1 plasmids were digested with EcoRI and XhoI. The DNA fragments encoding SAM and KEP1 were subcloned in the EcoRI and SalI site of pEGFP-C1 (CLONTECH). GFP-Who/How was generated by digesting Myc-Who/How with EcoRI, and the DNA fragment was subcloned into the EcoRI site of pEGFP-C1. The Myc-GFP plasmids were generated by digesting the mammalian GFP constructs with Nhel and SmalI. The ends of the DNA fragments were made blunt with the Klenow fragment of DNA polymerase I. The DNA fragments were subcloned into the SmalI site of pRHM Ha3 (33).

Northern Blot Analysis—RNA was prepared from early (0–4 h), mid to late (4–20 h) embryos, first instar larvae, adult males and females, and ovaries, by phenol/chloroform extraction and ethanol precipitation as described previously (34). Poly(A)^+ RNA was isolated by oligo(dT)-cellulose chromatography (type III, Sigma) and quantitated by measuring absorbance at 260 nm. For preparation of the Northern blots, 10 μg of poly(A)^+ RNA denatured with glyoxal, electrophoretically separated on a 1% agarose gel, and transferred to Gene Screen filters. The KEP1 and SAM 32P probes were generated by PCR using the above described oligonucleotides with KEP1-Myc-Bluescript and SAM-Myc-Bluescript as DNA templates. Sizes of the hybridized bands were estimated by comparison to the migration distances of commercial marker RNAs (0.24- to 9.5-kb RNA ladder, Life Technologies, Inc.).

Protein Expression and Analysis—HeLa cells were transfected with the vaccinia virus T7 expression system and lysed as described previously (30). The samples were analyzed on SDS-polyacrylamide gels and transferred to nitrocellulose membranes. Immunoblotting was performed using the following monoclonal antibodies: anti-Myc 9E10 (35) and anti-hemagglutinin (HA; BabCO). The secondary antibody was goat anti-mouse conjugated to horseradish peroxidase (Organon Teknika-Cappell, Durham, NC), and chemiluminescence was used for protein detection (NEN Life Science Products).

RNA Binding Analysis—Homopolymeric RNA binding was performed using poly(U) (American Pharmaceutical, Biozyme), poly(A), poly(C), and poly(G) Sepharose beads (Sigma) in lysis buffer supplemented with 2 mg/ml heparin (Sigma) for 1 h at 4 °C. The beads were washed and analyzed as described (36). Chemical cross-linking studies were performed as described previously (16).

Transfection in NIH 3T3 Cells and Apoptosis—NIH 3T3 cells were plated 12 h before transfection typically at a density of 10^5 cells/22-mm coverslip (Fisher Scientific Co.). Cells were transfected with DNA constructs encoding GFP alone, GFP-SAM, GFP-KEP1, or GFP-Who/How using LipofectAMINE PLUS reagent (Life Technologies, Inc.). Twelve, 24, or 36 h after transfection, the cells were fixed with 4% paraformaldehyde in phosphate-buffered saline for 15 min and permeabilized with 1% Triton X-100 in phosphate-buffered saline for 5 min. The nuclei were stained with 3 μg/ml 4,6-diamidino-2-phenylindole (DAPI). The morphology of transfected cells was examined with fluorescence microscopy, and areas with morphological features such as nuclear condensation and fragmentation were considered apoptotic.

Transfection in Drosophila Schneider 2 Cells and Condensation of HeLa Nuclei—Drosophila Schneider 2 (S2) (Invitrogen) cells were maintained at room temperature in Schneider media (Life Technologies, Inc.) supplemented with 10% fetal calf serum. Prior to transfection, 2 × 10^6 S2 cells were plated and grown for 6 h. The cells were transfected using the calcium phosphate precipitation method, as suggested by the supplier. In brief, cells were incubated with DNA for 4 h and washed twice with complete S2 media prior to induction by addition of CuSO_4 to a final concentration of 0.5 mM. For chromatin condensation assay, HeLa nuclei were freshly isolated essentially as described (30) with the following modification; the buffer composition was 10 mM Hepes, pH 7.9, 1.5 mM MgCl_2, 10 mM KCl, 0.5 mM dithiothreitol, 10 mM ATP, 0.5 mM phenylmethylsulfonyl fluoride, and 5 mM Na_2 EDTA, and proteinase K, 50 μg/ml, was added as suggested by the supplier. The cells were collected, washed twice with ice-cold phosphate-buffered saline, and resuspended in 125 μl of 0.25 M Tris, pH 7.4. Cells were lysed by three freeze/thaw cycles (dry ice-ethanol/37 °C). Cell debris was removed by centrifugation for 10 min at 10,000 rpm at 4 °C. One hundred μl of each lysate was added to 30 μl of isolated HeLa nuclei (~2.0 × 10^6 nuclei), and incubated at 37 °C for up to 3 h. Aliquots of nuclei were removed, streaked onto coverslips, and stained with DAPI. The presence of condensed nuclei and apoptotic bodies were scored as nuclei undergoing apoptosis.

RESULTS

Cloning of Two Novel Drosophila KH Domain Proteins—To identify GSS related family members, the expressed sequence tag data base was searched with the BLAST program (37) using the amino acid sequence of the KH domain of Sam68 (32). Two Drosophila sequence tags were identified that contained conserved regions similar to the Sam68 KH domain. One protein we called KEP1 for KH domain encompassing protein 1. The mRNA for this protein has an open reading frame of 960 nucleotides and encodes a protein of 320 amino acids (Fig. 1A). The amino acid sequence of this protein does not resemble any of the known GSS domain proteins. In contrast, the other protein was called SAM because it has a high sequence similarity with mammalian Sam68 and therefore potentially represents the Drosophila Sam68 homolog. The mRNA for SAM encodes a protein of 428 amino acids (Fig. 1B). The primary amino acid sequence of this protein does not resemble any of the known GSS domain proteins. In contrast, the other protein was called SAM because it has a high sequence similarity with mammalian Sam68 and therefore potentially represents the Drosophila Sam68 homolog. The mRNA for SAM encodes a protein of 428 amino acids (Fig. 1B). The primary amino acid sequence of this protein does not resemble any of the known GSS domain proteins. In contrast, the other protein was called SAM because it has a high sequence similarity with mammalian Sam68 and therefore potentially represents the Drosophila Sam68 homolog. The mRNA for SAM encodes a protein of 428 amino acids (Fig. 1B).
sequence identity in the GSG domain (Fig. 1C), whereas the closest other Sam68 family member, GRP33, has a 38% sequence identity with Sam68. In addition, SAM has a cluster of 12 tyrosines at its carboxyl terminus like Sam68. In addition to the KH domain, Sam68 has two RGG sequences at its amino terminus that might function as an RGG box and may be the site of arginine methylation (38). The function of this potential Sam68 RNA binding motif is unknown (16). SAM also has an RGG sequence like Sam68; however, it is localized at the carboxyl terminus of the molecule. In contrast to Sam68 (32, 39–41), SAM does not contain the multiple proline-rich motifs that serve as binding sites for SH3 domains.

**Alignment of KEP1 and SAM with Other GSG Domain Family Members**—The KH domain boundaries used for this alignment were taken from Musco and co-workers (2). The GSG domain amino-terminal boundary begins with the lysine at position 4 of GRP33 as defined by Jones and Schedl (7). Defining the COOH terminus of the GSG domain associated region (7). Now with the presence of other family members, it is clear that the GSG domain extends to the COOH terminus of the previously called CGA domain. Based on the data presented here, we define the whole region including the CGA as the GSG domain. Alignment of KEP1 and SAM with Who/How, Qk1, GLD-1, Sam68, GRP33, and SF1 demonstrates that 14 amino acids in the GSG domain are identical between all members (Fig. 2, *). Twenty-one amino acids are conservative substitutions (:), 39 residues (∧) are found in at least five of the seven members, and 30 residues are found in four of the seven members (1). This indicates that the GSG domain and the embedded KH domain are highly conserved between species. Of particular note in the KH domain, the GXXXG sequence predicted to form the RNA binding site (2), is fully conserved between all members. In addition, the lengths of the KH domain loop 1 and 4 are identical between members except for GLD-1, Qk1, and Who/How, which have a shorter loop 4 lacking two residues (Fig. 2).

**Expression of the KEP1 and SAM Transcripts**—The pattern of expression of KEP1 and SAM was investigated by Northern blot analysis (Fig. 3). The SAM gene produces four identifiable transcripts: a major transcript of 2.5 kb, and less abundant mRNAs of 2.8, 4.0, and 4.6 kb. The longest transcript is the...
least abundant, but like Sam68, uniformly expressed throughout development. The three shorter transcripts are detectable only in adult females, ovaries, and early (0–4 h) embryos (Fig. 3, lanes 1–6), suggesting that these SAM mRNAs are synthesized during oogenesis and are contributed maternally to the embryo. The presence of multiple SAM transcripts suggests that the SAM gene is alternatively spliced, as it has recently been demonstrated for Sam68 (42). The KEP1 RNA is expressed at high levels throughout all stages of development except for larvae, in which its relative level is reduced. The major transcript from KEP1 is 1.6 kb, although a 2.0-kb form is also observed. Two smaller transcripts (1.3 and 0.6 kb) are specific to males (Fig. 3, lanes 7–12). These findings show that KEP1 and SAM have a different pattern of expression and most likely perform completely separate functions in the flies.

KEP1 and SAM Are RNA-binding Proteins—To determine whether KEP1 and SAM are RNA-binding proteins, we epitope tagged each protein at the amino terminus with a short sequence from c-Myc that is recognized by the hybridoma antibody 9E10 (35). Since specific RNA target sequences for these members are unknown, we utilized homopolymeric RNA. The plasmids encoding these proteins were transfected in HeLa cells, and the cells were lysed in lysis buffer supplemented with 2 mg/ml heparin. The cell lysates were incubated with control, poly(G), poly(U), poly(C), and poly(A) Sepharose beads. The bound proteins were separated by SDS-PAGE, transferred to nitrocellulose, and probed with 9E10. The results show that KEP1 and SAM are indeed RNA-binding proteins, and their binding patterns are consistent with their known RNA-binding domains (Fig. 4).

Identification of KEP1 and SAM: Novel KH Domain Proteins
Northern blots of Drosophila containing amino acids 223–428 was unable to associate with COOH-terminal to the GSG domain (Fig. 5). This homopolymeric RNA binding specificity was similar to that observed for Sam68 and GRP33 (16, 40). KEP1 migrated with an approximate molecular mass of 40 kDa and bound poly(U) Sepharose (Fig. 4, lanes 1–6). The GSG domain is necessary but not sufficient for RNA binding. In contrast to SAM, the GSG domain of KEP1 was sufficient to mediate RNA binding. A KEP1 truncation protein encompassing the entire GSG domain bound poly(U) homopolymeric RNA (Fig. 5A, KEP:1–224).

We investigated whether KEP1, SAM, and Who/How could self-associate by co-immunoprecipitation studies. HeLa cells transfected with HA-SAM and Myc-SAM, HA-KEP1 and Myc-KEP1, HA-Who/How and Myc-Who/How were immunoprecipitated with control or anti-Myc antibodies. The bound proteins were separated by SDS-PAGE, transferred to nitrocellulose, and immunoblotted with anti-HA antibodies. SAM, KEP1, and Who/How all self-associated into homomultimers (data not shown). The deletion constructs for KEP1 and SAM were tested for their ability to self-associate with wild-type KEP1 or SAM in co-immunoprecipitation assays in transfected HeLa cells. SAM:1–306 and SAM:1–345 were able to associate with wild-type SAM (Fig. 5B, lanes 9 and 12), whereas SAM:1–264 and SAM:233–428 were unable to associate with SAM (lanes 6 and 15). These data suggest that the GSG domain is required for self-association and show that the minimal region for SAM self-association resides in amino acids 1–306. The GSG domain of KEP1 was sufficient to associate with wild-type KEP1, demonstrating that the KEP1 GSG domain can mediate the self-association (Fig. 5B, lane 21).

**The Induction of Apoptosis by SAM and KEP1**—To determine the cellular function of SAM, KEP1, and Who/How, we first expressed these proteins as GFP fusions in mammalian fibroblast cells. GFP alone was localized both in the cytoplasm and nucleus, GFP-Who/How and GFP-KEP1 were found exclusively in the nucleus, and GFP-SAM was cytoplasmic (Fig. 6A, GFP panels). Twelve hours after transfection, GFP-expressing cells exhibited morphological changes characteristic of apoptosis, including cell shrinkage, cytoplasm condensation, membrane blebbing, and membrane blebbing, not observed with GFP alone, GFP-SAM, or GFP-Who/How (Fig. 6A). The apoptotic cells were scored based on the presence of chromatin condensation and fragmentation as observed by DAPI staining (Fig. 6A). Nuclear morphological changes were scored by DAPI staining prior to terminal deoxynucleotidyltransferase-mediated fluorescein-dUTP nick end labeling (TUNEL, Ref. 18).

The levels of apoptosis induced by KEP1 protein was quantitated by randomly counting cells and expressing the number of apoptotic cells as a percentage of transfected (green) cells. NIH 3T3 cells were transfected with plasmids expressing GFP, GFP-KEP1, GFP-SAM, or GFP-Who/How. A small fraction of GFP, GFP-SAM, or GFP-Who/How expressing cells were apoptotic, and this fraction remained steady (Fig. 6B, ~10–15%). In contrast, ~70% of the KEP1-transfected cells were apoptotic at 36 h (Fig. 6B). These data demonstrate that the expression of KEP1 in fibroblast cells induces apoptosis.

We next examined the ability of Who/How, SAM, and KEP1 to induce apoptosis in Drosophila S2 cells. Although the S2 cells have never been demonstrated to undergo the typical morphological changes associated with apoptosis such as nuclear condensation, nuclear bodies, and membrane blebbing, activation of apoptosis in these cells can be measured indirectly by mixing cell extracts with isolated HeLa nuclei (30). This assay has been utilized to show that S2 cells treated with cycloheximide or transfected with reaper have increased caspase activity (31). To determine whether Who/How, SAM, and KEP1 activated the apoptotic pathway, S2 cells were mock-transfected or transfected with GFP-Who/How, GFP-SAM, and GFP-KEP1. As visualized by using the GFP moiety, approximately 10–15% of the S2 cells were transfected; by using a
specific KEP1 antibody, we have estimated the amount of KEP1 protein to be 5–6 times higher than endogenous KEP1 in transfected cells (data not shown).

The transfection of SAM in Drosophila S2 cells demonstrated that it was cytoplasmic, whereas KEP1 and Who/How were both nuclear, consistent with our previous data in NIH 3T3 cells (Fig. 6B and data not shown). Cellular extracts were isolated from the different transfected S2 cells and mixed with isolated HeLa nuclei for 3 h. In the presence of extracts from mock- and GFP-Who/How-transfected cells, HeLa nuclei did not exhibit significant morphological changes over the duration of the experiments (Fig. 7A). In contrast, extracts from KEP1- and SAM-transfected S2 cells induced chromatin condensation and formation of apoptotic bodies (Fig. 7A). The levels of apoptosis induced by SAM and KEP1 was quantitated by randomly counting HeLa nuclei and expressing the number of apoptotic nuclei as a percentage of total nuclei. Approximately 5% of the HeLa nuclei were undergoing apoptosis when mixed with cellular extracts from S2 cells mock- or GFP-Who/How-transfected and 25–40% of the HeLa nuclei were undergoing ap-
optosis when mixed with cellular extracts from S2 cells transfected with SAM or KEP1. These data suggest that the apoptotic pathways are activated in S2 cells by the overexpression of SAM or KEP1.

**DISCUSSION**

We have identified two GSG domain protein family members in *Drosophila*, SAM and KEP1. SAM is a cytoplasmic protein that has 50% sequence identity with Sam68 in the GSG domain and may represent the *Drosophila* homolog of Sam68. In addition to amino acid similarities in the GSG domain, SAM also contains a tyrosine-rich COOH terminus that may serve as a substrate for tyrosine kinases as has been shown for Sam68 (4, 32, 39–41). SAM is an RNA-binding protein, and it bound both poly(U) and poly(A) homopolymeric RNA in vitro. This binding specificity is similar to that observed with Sam68 and GRP33 (16). SAM self-associated and deletion studies demonstrated that the SAM GSG domain was required for both RNA binding and self-association like Sam68 (16). Although SAM has many similarities with Sam68 it is clearly different. SAM is cytoplasmic, whereas Sam68 is predominantly nuclear (data not shown and Ref. 4). The localization of SAM suggests that it may be involved in some aspect of translation or mRNA stability. The presence of SAM in ovaries is consistent with this hypothesis because translation control is an important mechanism of gene regulation in that tissue (44).

KEP1 is a 40-kDa RNA-binding protein that bound poly(U)
homopolymeric RNA. It is a ubiquitously expressed nuclear protein that most likely self-associates into homotrimers, as determined from the molecular mass of the chemical cross-linked complex in HeLa cells (data not shown). KEPI cross-linked complexes were able to bind homopolymeric RNA in vitro, supporting the idea that GSG domain-containing proteins bind RNA as multimers (16). The self-association property is now shared by eight different GSG domain proteins including Sam68, GRP33, mouse Qk1–7, GLD-1 (16, 18), Xqua (17), and SAM, KEPI, and Who/How (this report). All GSG domain proteins identified thus far have been shown to associate with RNA in vitro (4, 9, 10, 16). The multimer formation may be required for high affinity binding to their target RNA sequences. The GSG-domain-containing family of proteins is growing rapidly (for a review, see Ref. 14). Drosophila partial sequences of 27 amino acids in the KH domain were cloned for quaking-related proteins (12). The sequence of these four proteins has been recently reported, and the proteins were named QKR58E1, E2, E3, and B (45). QKR54B and QKR58E3 sequences of 27 amino acids in the KH domain were cloned for homopolymeric RNA. It is a ubiquitously expressed nuclear in vitro physiologically relevant cell line for SAM. By using an domain-containing family members. Taken together, these named QKR58E1, E2, E3, and B (45). QKR54B and QKR58E3 may be required for high affinity binding to their target RNA domain proteins identified thus far have been shown to associate with RNA in vitro. The apoptosis induced by the overexpression of KEPI and SAM did not require additional manipulations such as the reduction in serum as observed with c-Myc (47). In this regard, SAM and KEPI are similar to Reaper, HID, and GRIM, where the overexpression alone is sufficient to induce apoptosis in insect cells (48, 49).

The identification of two new Drosophila GSG domain family members that we called SAM and KEPI demonstrates the existence of a large family of GSG domain proteins. KEPI and SAM are RNA-binding proteins that bind homopolymeric RNA in vitro and self-associate into multimers. Although our results strongly indicate a putative role for KEPI and SAM in the regulation apoptosis, it remains to be determined whether these proteins have a toxic effect such as perturbing the normal cellular processes or are directly involved in caspase-mediated apoptosis. Since the name SAM is used for a protein domain (sterile alpha motif), we propose the name Sam50 for the Drosophila SAM protein.

Acknowledgments—We thank Tim Scheid and John Hiscott for valuable comments and for critically reading the manuscript. We are grateful to Matthieu Miron for performing the CLUSTER alignments. We thank Hugh Chou for performing the CLUSTER alignments.

REFERENCES

1. Sioni, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993) Nucleic Acids Res. 21, 1193–1198
2. Musco, G., Stier, G., Joseph, C., Morelli, M. A., and Pastore, A. (1996) Cell 85, 237–245
3. Gibson, T. J., Thompson, J. D., and Heringa, J. (1993) FEBS Lett. 326, 361–366
4. Gong, W., Muller, O., Clark, R., Conroy, L., Moran, M. F., Polakis, P., and McCormick, F. (1992) Cell 70, 551–558
5. Lock, P., Fumagalli, S., Polakis, P., McCormick, F., and Courtneidge, S. A. (1996) Cell 84, 23–24
6. Cruz-Alvarens, M., and Pelliser, A. (1997) J. Biol. Chem. 262, 13337–13380
7. Jones, A. R., and Schedl, T. (1995) Genes Dev. 9, 1491–1504
8. Ebersole, T. A., Chen, Q., Justice, M. J., and Artzt, K. (1996) Nat. Genet. 12, 260–265
9. Zorn, A. M., Grow, M., Patterson, K. D., Ebersole, T. A., Chen, Q., Artzt, K., and Krieg, P. A. (1997) Gene (Amst.) 188, 199–206
10. Arning, S., Gruter, P., Bilbe, G., and Kramer, A. (1996) RNA 2, 794–810
11. Bachebre, E. H. (1997) Development 124, 1323–1332
12. Fyrborg, C., Becker, J., and Olszowy, M. W., Mahaffey, J., and Fyrborg, E. (1997) Gene (Amst.) 197, 315–323
13. Zaffran, S., Astier, M., Graeteos, D., and Semeriva, M. (1997) Development 124, 2087–2098
14. Vernet, C., and Artzt, K. (1997) Trends Genet. 13, 479–484
15. Lin, Q., Taylor, S. J., and Shalawey, D. (1997) J. Biol. Chem. 272, 27274–27290
16. Chen, T., Damaj, B., Herrerra, C., Lasko, P., and Richard, S. (1997) Mol. Cell. Biol. 17, 5507–5718
17. Zorn, A. M., and Krieg, P. A. (1997) Genetics 11, 2176–2190
18. Chen, T., and Richard, S. (1998) Mol. Cell. Biol. 18, 4863–4871
19. Verkerk, A. J. M. H., Piereetti, M., Sutcliffe, J. S., Fu, Y.-H., Kuhi, D. P., Pizzi, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F., Eussen, B. E., van Ommen, G. J. B., Blonden, L. A. J., Bigga, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A., and Warren, S. T. (1991) Cell 65, 905–914
20. Piereetti, M., Zhang, F., Fu, Y.-H., Warren, S. T., Oostra, B. A., Caskey, C. T., and Nelson, D. L. (1991) EMBO J. 10, 817–822
21. Nussbaum, R. L., and Ledbetter, D. H. (1995) in Metabolic Basis of Inherited Disease (Servier, C. R., Beaudet, A., Sly, W. S., and Valle, D., eds.) pp. 795–810, MacGraw-Hill, New York
22. DeBoulle, K., Verkerk, A. J. M. H., Reyniers, E., Vits, L., Hendrickch, J., Roy, B. V., Bos, F. V. D., DeGraff, E., Oostra, B. A., and Willems, P. J. (1993) Nat. Genet. 3, 31–35
23. Sioni, H., Choi, M., Sioni, M. C., Nussbaum, R. L., and Dreyfuss, G. (1994) Cell 77, 3–39
24. Francis, R., Barton, M. K., Kimbel, J., and Schedl, T. (1995) Genetics 139, 579–606
25. Jones, A. R., Francis, R., and Schedl, T. (1996) Dev. Biol. 180, 165–183
26. Ikenberry, S. O., Less, R. I., Dickie, M. M., and Angell, S. H. (1984) Science 217, 309–311
27. Mahone, M., Saffman, E. E., and Lasko, P. F. (1995) EMBO J. 14, 2043–55
28. Saffman, E. E., Styhier, S., Rother, K., Li, W., Richard, S., and Lasko, P. (1998) Mol. Cell. Biol. 18, 4855–4862
29. Lazzarini, V. A., Cole, S., Cooke, C. A., Nelson, W. G., and Earman, W. C. (1993) J. Cell Biol. 123, 7–22
30. Fraser, A. G., McCarthy, N. J., and Evans, G. I. (1997) EMBO J. 16, 6192–6199
31. Richard, S., Yu, D., Blumer, K. J., Hausladen, D., Oleszewy, M. W., Connelly, P. A., and Shaw, A. S. (1995) Mol. Cell. Biol. 15, 186–197
32. Busek, T. A., Grinblat, Y., and Goldstein, L. S. (1988) Nucleic Acids Res. 16, 1043–1061
33. Doering, R. E., Buter, B., Krzywy, M. E., and Kuli, B. (1988) EMBO J. 7, 2579–84
34. Evan, G. I., and Bishop, J. M. (1985) Mol. Cell. Biol. 5, 2843–2850
35. Wang, L. L., Richard, S., and Shaw, A. S. (1995) J. Biol. Chem. 270, 2010–2013
36. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
37. Gary, J. D., and Clarke, S. (1998) Prog. Nucleic Acid Res. Mol. Biol. 61, 65–131

Identification of KEPI and SAM: Novel KH Domain Proteins 30129
30130  Identification of KEPI and SAM: Novel KH Domain Proteins

39. Fumagalli, S., Totty, N. F., Hsuan, J. J., and Courtneidge, S. A. (1994) Nature 368, 871–874
40. Taylor, S. J., and Shalloway, D. (1994) Nature 368, 867–871
41. Weng, A., 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
42. Barlat, I., Maurier, F., Duchesne, M., Guitard, E., Toque, B., and Schweighoffer, F. (1997) J. Biol. Chem. 272, 3129–3132
43. Burd, C. G., and Dreyfuss, G. (1994) Science 265, 615–621
44. Rongo, C., and Lehmann, R. (1996) Trends Genet. 12, 102–109
45. Fyrberg, C., Becker, J., Barthmaier, P., Mahaffey, J., and Fyrberg, E. (1998) Biochem. Genet. 36, 51–64
46. Song, Z., McCall, K., and Stellar, H. (1997) Science 275, 536–539
47. Harrington, E. A., Bennett, M. R., Fanidi, A., and Evan, G. I. (1994) EMBO J. 13, 3296–3295
48. Vucic, D., Kaiser, W. J., and Miller, L. K. (1998) Mol. Cell. Biol. 18, 330–3309
49. Vucic, D., Kaiser, W. J., Harvey, A. J., and Miller, L. K. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 10183–10188