INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

BIOCHEMICAL AND FUNCTIONAL CHARACTERIZATION OF THE GSG (GRP33, Sam68, GLD-1) DOMAIN

by

Taiping Chen

Division of Experimental Medicine, Department of Medicine

McGill University, Montreal

May 2000

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Doctor of Philosophy

To my wife, Cuiju, and children, Jinjian and Michael.

ABSTRACT

The GSG (GRP33, Sam68, GLD-1) domain is a tripartite protein module of ~200 amino acids. It consists of an hnRNP K homology (KH) domain and two flanking regions N- and C-terminal to the KH domain called the NK region and CK region, respectively. The KH domain embedded in the GSG domain has longer loops 1 and 4 compared to other KH domains. The physiological significance of the GSG domain is demonstrated by the fact that genetic mutations in the GSG domain result in developmental defects in various species. The primary goal of this thesis is to characterize the biochemical properties and functions of the GSG domain. We demonstrated that the GSG domain proteins Sam68, Qk1, GRP33, and GLD-1 are RNA-binding proteins that self-associate into multimers. Sam68 complexes bound homopolymeric RNA and the SH3 domains of $p59^{fyn}$ and phospholipase Cγ1 in vitro, indicating that Sam68 associates with RNA and signaling molecules as a multimer. The formation of Sam68 complex was inhibited by $p59^{fyn}$, suggesting that Sam68 oligomerization is regulated by tyrosine phosphorylation. Deletion studies in Sam68 and Qk1 demonstrated that the GSG domain is required for both self-association and RNA binding. Sam68 oligomerization requires the extended loops 1 and 4 of the KH domain and the Qk1 dimerization domain is mapped to the NK region. A Qk1 lethal point mutation, altering glutamic acid 48 to a glycine in the NK region, abolishes Qk1 self-association but has no effect on its ability to bind total cellular RNA. The mutant Qk1 protein is more potent than wild-type Qk1 in inducing apoptosis when expressed in NIH 3T3 cells, suggesting that failure to dimerize may be the molecular mechanism for the embryonic lethality. In addition to mediating self-association and RNA binding, the Sam68 GSG domain plays a role in protein localization. We demonstrated by indirect immunofluorescence studies that Sam68 is concentrated in novel nuclear structures, termed Sam68/SLM nuclear bodies (SNBs). Mutations and deletions within the Sam68 GSG domain result in altered localization patterns including exclusive accumulation into SNBs, cytoplasmic localization, and microtubule association. The localization patterns do not

correlate with the ability of the Sam68 proteins to bind RNA and self-associate, suggesting that protein localization is a separate property of the GSG domain. Taken together, these findings demonstrate that the GSG domain mediates protein-protein and protein-RNA interactions and directs cellular localization of ribonucleoprotein complexes.

RÉSUMÉ

Le domaine GSG (GRP33, Sam68, GLD-1) est une région d'environ 200 acides aminés. Ce dernier contient trois sous-domaines distincts: le domaine KH (homologue à la ribonucléoprotéine hnRNP K) et les régions NK et CK bordant le domaine KH. Contrairement aux autres domaines KH, celui appartenant au domaine GSG possède de plus longues boucles 1 et 4. L'importance physiologique du domaine GSG est démontrée par le fait que des mutations génétiques dans ce domaine résultent en des anomalies au niveau du développement chez plusieurs espèces. L'objectif majeur de cette thèse est de caractériser biochimiquement les propriétés et les fonctions du domaine GSG. Dans un premier temps, nous avons démontré que le domaine GSG des protéines Sam68, Qk1, GRP33 et GLD-1 est responsable de la fixation à l'ARN. De plus, ces protéines peuvent s'associer entre elles tout en liant l'ARN. En plus de lier l'ARN homopolymérique, la protéine Sam68 se lie aux domaines SH3 de $p59^{fyn}$ et de la phospholipase Cγ1. Ceci suggère que Sam68 s'associe en multimère avec l'ARN et des protéines de signalisation. Cependant, la formation de ce complexe est inhibé par 59^{fyn}, donc ce phénomène serait régulé par les tyrosines kinases. Des études de délétion de Sam68 et Qk1 ont démontré que le domaine GSG est requit pour l'oligomérisation et la fixation à l'ARN. L'oligomérisation de Sam68 nécessite les boucles 1 et 4 du domaine KH. De plus, la région NK de Qk1 a été démontrée comme étant celle impliquée dans la dimérisation de Qk1. Il a été démontré qu'une mutation ponctuelle léthale E48G dans la région NK de la protéine Qk1 abolissait la dimérisation mais n'affectait pas la capacité de Qk1 de se lier à l'ARN cellulaire. Lorsque exprimé dans les cellules NIH 3T3, ce mutant est plus susceptible que la protéine Qk1 de type sauvage à induire l'apoptose. L'incapacité de dimérisation pourrait expliquer l'origine léthale embryonnaire de cette mutation. Le domaine GSG de Sam68 serait également impliqué dans la localisation cellulaire de la protéine. Des études d'immunofluorescence ont démontré que Sam68 est localisé dans des structures nucléaires spécifiques appelées SNBs

(Sam68/SLM nuclear bodies). Des études de mutagenèse dans le domaine GSG de Sam68 ont provoqué plusieurs types d'altération dans la localisation. Certaines mutations provoquent une accumulation exclusive soit au niveau des SNBs, du cytoplasme ou des microtubules. La propriété de localisation de Sam68 serait indépendante de sa capacité à lier l'ARN et d'homodimérisation. L'ensemble de ces observations, indique que le domaine GSG est important dans les interactions protéine-protéine, protéine-ARN de même que dans la localisation cellulaire des complexes ribonucléoprotéiques.

PREFACE

This Ph.D. thesis was written in accordance with the Guidelines for Thesis Preparation from the Faculty of Graduate Studies and Research at McGill University. I have exercised the option of writing the thesis as a manuscript-based thesis. The guidelines state: ".....Candidates have the option of including, as part of the thesis, the text of one or more papers submitted, or to be submitted, for publication, or the clearly-duplicated text (not the reprints) of one or more published papers. These texts must conform to the 'Guidelines for Thesis Preparation' with respect to font size, line spacing and margin sizes and must be bound together as an integral part of the thesis. .....The thesis must be more than a collection of manuscripts. All components must be integrated into a cohesive unit with a logical progression from one chapter to the next. In order to ensure that the thesis has continuity, connecting texts that provide logical bridges between the different papers are mandatory. .....The thesis must include the following: (a) a table of contents; (b) an abstract in English and French; (c) an introduction which clearly states the rational and objectives of the research; (d) a comprehensive review of the literature (in addition to that covered in the introduction to each paper); (e) a final conclusion and summary. .....As manuscripts for publication are frequently very concise documents, where appropriate, additional material must be provided (e.g., in appendices) in sufficient detail to allow a clear and precise judgement to be made of the importance and originality of the research reported in the thesis. .....In general, when co-authored papers are included in a thesis the candidate must have made a substantial contribution to all papers included in the thesis. In addition, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. This statement should appear in a single section entitled 'Contributions of Authors' as a preface to the thesis. .....".

I have included, as chapters of this thesis, the texts of three original papers which have been published. Each of these chapters (Chapters 2, 3 and 4) contains its own summary,

introduction, materials and methods, results, discussion, and references sections. Additionally, in order to bridge the papers with connecting texts, a preface is added at the beginning of each chapter. A general introduction and literature review is given in Chapter 1 and a final discussion is presented in Chapter 5. The references for chapters 1 and 5 are included at the end of the thesis.

Papers included in this thesis:

Chapter 2 **Chen, T.**, Damaj, B. B., Herrerra, C., Lasko, P. and Richard, S. (1997). Self-association of the single-KH domain family members Sam68, GRP33, GLD-1 and Qk1: Role of the KH domain. Mol. Cell. Biol. 17, 5707-5718.

Chapter 3 **Chen, T.** and Richard, S. (1998). Structure-function analysis of Qk1: A lethal point mutation in mouse *quaking* prevents homodimerization. Mol. Cell. Biol. 18, 4863-4871.

Chapter 4 **Chen, T.**, Boisvert, F. M., Bazett-Jones, D. P. and Richard, S. (1999). A role for the GSG domain in localizing Sam68 to novel nuclear structures in cancer cell lines. Mol. Biol. Cell 10, 3015-3033.

Contributions of Authors:

The candidate performed most of the research included in this thesis. Contributions of other authors to this work are described below:

In Chapter 2, the experiments described in Figs. 2-1, 2-5, part of 2-4, and part of 2-7 (A and B) were performed by Stéphane Richard. In situ chemical cross-linking experiments shown in Fig. 2-2 and Fig. 2-8B were done by Bassam B. Damaj. The yeast two-hybrid experiments shown in Table 2-1 were performed by Constance Herrera. Paul Lasko provided *Drosophila* BicC cDNA and antibody used in this study.

In Chapter 4, electron microscopy (Fig. 4-9) was performed by Francois-Michel Boisvert and David P. Bazett-Jones.

All the studies were conducted under the supervision of Dr. Stéphane Richard.

In addition to the papers included in this thesis, the candidate contributed to the following studies, which have been published or submitted:

Di Fruscio, M., **Chen, T.**, Bonyadi, S., Lasko, P., and Richard, S. (1998). The identification of two *Drosophila* KH domain proteins: KEP1 and SAM are members of the Sam68 family of GSG domain proteins. J. Biol. Chem. 273, 30122-30130.

*Di Fruscio, M., ***Chen, T.**, and *Richard, S. (1999). Two novel Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis. Proc. Natl. Acad. Sci. USA 96, 2710-2715. (* these authors contributed equally to this work).

Aarts, M. M., Levy, D., He, B., Stregger, S., **Chen, T.**, Richard, S., and Henderson, J. E. (1999). Parathyroid hormone-related protein interacts with RNA. J. Biol. Chem. 274, 4832-4838.

*Derry, J J., *Richard, S., **Chen, T.**, Ye, X., Vasioukhin, V., and Tyner, A. L. Sik/BRK phosphorylates Sam68 in the nucleus and negatively regulates its RNA binding ability. Submitted. (* these authors contributed equally to this work).

ACKNOWLEDGMENTS

I would like to express my great gratitude to my supervisor, Dr. Stéphane Richard, for his guidance, encouragement, and support during the course of my Ph. D. study. His enthusiasm together with the time and effort he invested in the supervision of my work has made this experience a true success.

I would also like to thank my fellow graduate students, postdoctoral fellows and technicians who contributed to the friendly and collaborative atmosphere in the Richard laboratory. In particular, I wish to thank Dr. Marco Di Fruscio for helpful discussions and Daniel Larocque for translating the abstract of my thesis into French.

I am grateful to members of the Molecular Oncology Group at the Lady Davis Institute for stimulating discussions and criticisms, which have made me more mature.

Special thanks to my wife, Cuiju, for her patience, understanding, and support throughout these stressful years.

I would like to acknowledge the following organizations for financial support during the course of my study: The Medical Research Council of Canada (MRC) for Doctoral Research Award (1998-2000); The National Cancer Institute of Canada (NCIC) for Terry Fox Research Studentship (1998, declined) and Student Travel Award (1999); The Cancer Research Society Inc. (CRS) for CRS Studentship (1997-1998); and McGill Cancer Center for Canderel Studentship (1997).

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter 4

Chapter 5

LIST OF ABBREVIATIONS

BBP	branchpoint binding protein
BMH	bis(maleimido)hexane
BPS	branchpoint sequence
BrdU	bromodeoxyuridine
Br-UTP	bromo-uridine triphosphate
CBC	cap-binding complex
CBP	cap-binding protein
CK region	C-terminal-to-KH region (of the GSG domain)
CPSF	cleavage and polyadenylation specificity factor
CstF	cleavage stimulation factor
CTD	C-terminal domain (of RNA polymerase II)
CTE	constitutive transport element (of simple retroviruses)
DAPI	4, 6-diamidino-2-phenylindole
DSS	disuccinimidyl suberate
eIF	eukaryotic translation initiation factor
ENU	N-ethyl-N-nitrosourea
GAP	GTPase-activating protein
GFP	green fluorescent protein
GSG domain	GRP33, Sam68, GLD-1 domain
GST	glutathione S-transferase
hnRNP	heterogeneous nuclear ribonucleoprotein
IBB domain	importin β-binding domain
IGC	interchromatin granule cluster
IRE	iron response element
IRES	internal ribosome entry site
KH	hnRNP K homology
KNS domain	hnRNP K nuclear shuttling domain
LMB	leptomycin B
MAG	myelin-associated glycoprotein
MBP	myelin basic protein
NES	nuclear export signal
NK region	N-terminal-to-KH region (of the GSG domain)
NLS	nuclear localization signal
NMR	nuclear magnetic resonance
NPC	nuclear pore complex

PAB	poly(A)-binding protein
PAP	poly(A) polymerase
PF	perichromatin fibril
PLC	phospholipase C
PNC	perinucleolar compartment
POD	PML oncogenic domain
PTB	polypyrimidine tract-binding protein
RBD	RNA-binding domain
RNP	ribonucleoprotein particle
RRE	Rev response element
RRM	RNA recognition motif
RS domain	arginine/serine-rich domain
SH2	Src homology 2
SH3	Src homology 3
SNB	Sam68/SLM nuclear body
snRNP	small nuclear ribonucleoprotein
SR protein	serine/arginine-rich protein
STAR	signal transduction and activation of RNA
TGE	tra-2 and GLI element
TUNEL	terminal deoxynucleotidyl transferase-mediated fluorescein-dUTP nick end labeling
U2AF	U2 snRNP auxiliary factor
UTR	untranslated region

Chapter 1
INTRODUCTION AND LITERATURE REVIEW

1.1 GENERAL INTRODUCTION

The transfer of genetic information from DNA to protein is mediated by RNAs. In eukaryotic cells, RNAs are produced in the nucleus by RNA polymerases I, II, or III, but a series of modifications are usually required before the transcripts can be used for protein synthesis. For messenger RNAs, the modifications typically include addition of 5' cap, pre-mRNA splicing, and 3' polyadenylation. Certain transcripts also undergo RNA editing, which results in sequence changes at the level of mRNA. Most of the RNA processing reactions take place in the nucleus. The finished product, an RNA molecule linked to proteins in a ribonucleoprotein particle (RNP), is then transported through the nuclear envelope to the cytoplasm where protein synthesis occurs.

Gene expression is regulated during development and in response to environmental stimuli. For a long time, transcriptional regulation (selective activation and repression of certain transcriptional promoters) had been thought to be the predominant regulatory mechanism. Studies in recent years, however, have revealed a variety of post-transcriptional mechanisms that are also essential in gene expression. In fact, some developmental events are primarily or exclusively controlled by post-transcriptional mechanisms. Regulatory controls may be imposed during all major steps of RNA metabolism, including RNA processing, transport, and translation (Hazelrigg, 1998). Post-transcriptional regulation of gene expression allows cells to respond to environmental cues more quickly than mechanisms acting at the transcriptional level. Moreover, post-transcriptional mechanisms, such as alternative splicing, alternative polyadenylation site selection, and alternative translational start site utilization, can generate functionally different protein products from a single gene.

The processes of RNA metabolism are mediated by *cis*-acting elements on RNA

molecules and *trans*-acting factors, including RNA-binding proteins. RNA-binding proteins are characterized by the presence of RNA-binding domains, which directly interact with RNA sequences or structures. Many of these proteins also contain other functional regions, such as catalytic domains, protein-protein interaction motifs, and localization signals. RNA-binding proteins play key roles in post-transcriptional processes. They can affect RNA conformation, mediate the formation of ribonucleoprotein complexes, and target these complexes to specific intracellular locations.

One group of RNA-binding proteins contains the heterogeneous nuclear ribonuleoprotein (hnRNP) K homology (KH) domain. This evolutionarily conserved protein module has been found in proteins from both prokaryotes and eukaryotes. Studies over the last several years have established that the KH domain mediates RNA binding and that KH domain-containing proteins are involved in various aspects of RNA metabolism. The GSG (GRP33, Sam68, GLD-1) protein family represents a special subgroup of KH domain-containing proteins. In contrast to other KH domain proteins, which usually contain multiple copies of KH domains, GSG family members each contain a single extended KH domain, which is embedded in a larger protein module called the GSG domain. The GSG domain may be derived from the KH domain during evolution, and it may gain additional properties with the inclusion of new sequence elements within and flanking the KH domain. Although their precise roles in many cases remain unknown, GSG proteins have been implicated in a number of eukaryotic RNA processes and functions, including splicing (Arning et al., 1996; Abovich and Rosbash, 1997), nuclear export (Nabel-Rosen et al., 1999), and translation (Jan et al., 1999).

The primary goal of this thesis is to characterize the biochemical properties and functions of the GSG domain. A better understanding of the GSG domain should help us elucidate the functions of GSG domain-containing proteins. In this chapter, I will first give a literature review on the study of the KH and GSG domains. Then, I will describe the major processes of RNA metabolism and several nuclear structures that may be involved in RNA processing.

1.2 KH DOMAIN

RNA-binding proteins play key roles in all processes of RNA metabolism. These proteins contain one or more specialized regions that mediate their direct interaction with their RNA targets (Siomi et al., 1997). One of the well-characterized RNA-binding domains is the KH domain. The KH domain was originally defined as a conserved region of approximately 50 amino acids that is repeated three times in the pre-mRNA-binding protein hnRNP K (Gibson et al., 1993; Siomi et al., 1993a). On the basis of structural data, the length of the KH domain was later revised to approximately 70 residues, with the exception of the GSG protein family which contains a longer KH domain (see below) (Musco et al., 1996). KH domains have been found in a wide variety of proteins from both prokaryotic and eukaryotic organisms (Gibson et al., 1993). Many of them have been shown to physically or functionally associate with RNA, suggesting a role for the KH domain in RNA binding. Notable examples of KH domain-containing proteins include the *Escherichia coli* (E. coli) transcription elongation factor NusA (Das, 1993; Gibson et al., 1993), the bacterial ribosomal protein S3 (Siomi et al., 1993a), the yeast meiosis-specific splicing regulator MER1 (Nandabalan et al., 1993; Siomi et al., 1993a), the *Drosophila melanogaster* P element pre-mRNA splicing regulator PSI (Siebel et al., 1995), the α-globin messenger RNP stability complex-associated proteins αCP-1 and αCP-2 (Kiledjian et al., 1995), the human fragile X mental retardation gene product FMR1 (Verkerk et al., 1991), and the fragile X related proteins FXR1 and FXR2 (Zhang et al., 1995; Khandjian et al., 1998). Over the last few years, a convergence of biochemical, genetic and structural data has demonstrated the direct involvement of the KH domain in RNA binding and the various functions of KH domain-containing proteins in RNA metabolism.

1.2.1 Three-Dimensional Structure of the KH Domain

The solution structures of the sixth KH domain of human vigilin (Musco et al., 1996), the first KH domain of human FMR1 (Musco et al., 1997), and the third KH domain of

hnRNP K (Baber et al., 1999) have been solved by nuclear magnetic resonance (NMR) spectroscopy. The crystal structures of the third KH domains of Nova-1 and Nova-2 have also been determined using X-ray crystallography (Lewis et al., 1999). The KH domain folds into a compact globular structure with an overall topology of βααββα. The three antiparallel β strands form a stable β sheet with the hydrophobic residues clustered on one side, where the β sheet is packed against the three α helices. The interactions between the hydrophobic residues of the β strands and the α helices form a hydrophobic core, which stabilizes the overall structure. The α helices and β strands are connected by five loops. Loop 2, which contains an invariant Gly-X-X-Gly (GXXG) tetrapeptide (X varies for different KH domains but is often a positively charged residue), is a predicted contact site for RNA. Two other loops, loops 1 and 4, which are spatially proximal to each other, protrude from the globular core and extend to the opposite direction as opposed to loop 2, suggesting that they may not be directly involved in RNA binding. Loops 1 and 4 are longer in the KH domain present in the GSG protein family (see below) and they are predicted to mediate protein-protein interactions (Musco et al., 1996). Most recently, the crystal structure of the third KH domain of Nova-2 bound to a high-affinity RNA ligand has been determined. In this cocrystal structure, α1, α2, and the edge of β2 form an aliphatic α/β platform, which is the major RNA recognition site, and the invariant GXXG motif appears to interact with the nucleic acid backbone. Interestingly, two protein-RNA complexes pack against each other and form dimers (Lewis et al., 2000).

1.2.2 Genetic Evidence Supporting a Physiological Role of the KH Domain

Genetic mutations that result in developmental defects have been identified in a number of genes encoding KH domain-containing proteins. The significance of the KH domains in the functions of these proteins is evidenced by the observations that KH domain mutations usually result in severe loss-of-function phenotypes.

Fragile X syndrome, the most common cause of heritable mental retardation in humans, is an X-linked dominant disorder that affects approximately 1 in 1250 males and 1 in 2500

females (Gustavson et al., 1986; Nussbaum and Ledbetter, 1986; Richards and Sutherland, 1992; Nussbaum and Ledbetter, 1995; Rousseau et al., 1995). The name arises from the frequent association of a fragile site (Xq27.3) in the X chromosome (Lubs, 1969; Harrison et al., 1983; Krawczun et al., 1985) with a characteristic spectrum of developmental defects including varying degrees of mental retardation, facial dysmorphia, and enlarged testicles in adult males (Turner et al., 1980a; Turner and Opitz, 1980b; Pembrey et al., 1986). The molecular basis of fragile X syndrome has been attributed to the expansion of a CGG trinucleotide repeat located within the 5' untranslated region (UTR) of the *FMR1* gene (Fu et al., 1991; Verkerk et al., 1991; Warren and Nelson, 1994), whose protein product FMR1 contains two KH domains (Ashley et al., 1993; Gibson et al., 1993; Siomi et al., 1993b). The number of the CGG repeat ranges from 2 to 60 in normal individuals, but is amplified to over 200 copies in affected patients (Fu et al., 1991; Kremer et al., 1991; Rousseau et al., 1991; Yu et al., 1991). Amplification of the CGG repeat is associated with hypermethylation of the CpG island 5' to *FMR1*, resulting in transcriptional silencing of this gene (Bell et al., 1991; Oberle et al., 1991; Sutcliffe et al., 1992; Feng et al., 1995). Deletions encompassing the *FMR1* gene (Gedeon et al., 1992; Wohrle et al., 1992), as well as intragenic deletions and nonsense mutations of *FMR1* (Meijer et al., 1994; Hirst et al., 1995; Lugenbeel et al., 1995) have also been identified in several cases. In one case with extremely severe phenotype, the FMR1 protein is expressed but carries a point mutation (Ile-304→Asn) in the second KH domain (DeBoulle et al., 1993). This highly conserved isoleucine residue is located on α2 helix and it plays a critical role in maintaining the hydrophobic core of the KH domain. Circular dichroic (CD) analysis has shown that introduction of the analogous mutation in the sixth KH domain of vigilin (Ile-32→Asn) completely disrupts the globular structure of the KH domain (Musco et al., 1996). Thus, the Ile-304→Asn mutation is predicted to affect all FMR1-containing complexes. Indeed, the point mutation has been shown to severely impair the ability of FMR1 to associate with RNA and polyribosomes (Siomi et al., 1994; Feng et al., 1997a; Khandjian et al., 1996). These data provide strong evidence supporting the physiological role of the KH domain. The cellular function of

FMR1 is still unknown. FMR1 is predominantly cytoplasmic (Eberhart et al., 1996; Feng et al., 1997a; Feng et al., 1997b) and it associates with polyribosomes in an RNA-dependent manner (Eberhart et al., 1996; Khandjian et al., 1996; Tamanini et al., 1996; Feng et al., 1997a; Corbin et al., 1997), suggesting that it may play a role in protein translation. FMR1 contains both a nuclear localization signal (NLS) (Eberhart et al., 1996; Sittler et al., 1996) and a nuclear export signal (NES) (Eberhart et al., 1996; Fridell et al., 1996) and there is evidence that it shuttles between the the nucleus and the cytoplasm (Feng et al., 1997b). Thus, FMR1 may also function in RNA transport and/or RNA stability.

Caenorhabditis elegans MEX-3, a protein with two KH domains, plays a role in establishing anterior-posterior asymmetry during embryonic development (Draper et al., 1996). At least 10 mutations have been identified in the *mex-3* locus. They all result in fully penetrant, recessive, strict maternal-effect embryonic lethality, and all alleles result in similar terminal phenotypes. Three missense mutations (Gly-153→Asp, Gly-153→Arg, and Gly-156→Asp) are located in the second KH domain, altering the invariant Gly-X-X-Gly motif of the predicted RNA binding site (Loop 2) (Draper et al., 1996). The observation that these missense mutations cause phenotypes indistinguishable from mutations that prevent expression of MEX-3 altogether suggests that RNA binding is essential for normal function of MEX-3. It is likely that MEX-3 contributes to anterior-posterior polarity of the embryo by regulating RNA expression, localization, or metabolism. There is evidence that MEX-3 can bind the 3'UTR of *pal-1* mRNA and repress PAL-1 protein translation (Hunter and Kenyon, 1996). *pal-1*has been shown to play an essential role in posterior development in *C. elegans* (Waring and Kenyon, 1990; Waring and Kenyon, 1991).

In *Drosophila melanogaster*, the *Bicaudal-C* (*Bic-C*) gene is required for a number of processes in oogenesis, including the migration of follicle cells and the establishment of anterior-posterior polarity in the oocyte (Mohler and Wieschaus, 1986; Ashburner et al., 1990; Schupbach and Wieschaus, 1991; Mahone et al., 1995). The Bic-C protein, which contains five KH domains (Mahone et al., 1995), is expressed in ovaries, and both Bic-C RNA and protein accumulate in the developing oocyte early in oogenesis (Saffman et al.,

1998). Females heterozygous for *Bic-C* alleles produce embryos with a range of anterior-posterior patterning defects, including bicaudal embryos (Mohler and Wieschaus, 1986; Mahone et al., 1995). The severity of the alleles is generally correlated with Bic-C protein levels in the ovaries, as strong alleles produce lower levels of protein, while weaker alleles express protein at levels comparable to those of the wild-type (Saffman et al., 1998). Two strong alleles produce normal levels of protein, but both contain mutations in the KH domains: *Bic-C*^{AB79} results in a 53-amino-acid deletion including part of the second and third KH domains, and *Bic-C*^{RU35} results in a point mutation (Gly-296→Arg) in the third KH domain (Mahone et al., 1995; Saffman et al., 1998). Gly-296 is a highly conserved residue located at the end of β3 strand and substitution of this residue by arginine would insert a bulky, charged side chain into the KH hydrophobic core and, hence, severely alter the overall structure of the KH domain (Musco et al., 1996). The Gly-296→Arg mutation has been shown to impair RNA binding in vitro (Saffman et al., 1998). These findings indicate that the KH domain-mediated RNA binding is essential for Bic-C function. One potential target for Bic-C is *oskar* mRNA, whose correct localization to the posterior pole of the developing oocyte is critical to embryonic anterior-posterior patterning (Ephrussi et al., 1991; Kim-Ha et al., 1991; Ephrussi and Lehmann, 1992). It has been reported that lack of Bic-C expression or *Bic-C* mutations that affect the KH domains lead to mislocalization and premature translation of *oskar* mRNA in oocyte (Mahone et al., 1995; Saffman et al., 1998).

1.2.3 Biochemical Analysis of the KH Domain

Direct interactions between RNA and the KH domain have been shown by cross-linking studies. Using UV (mainly crosslinks aromatic residues with RNA) and 2-iminothiolane (mainly crosslinks basic residues with RNA) as cross-linking agents, Urlaub and coworkers demonstrated that the E. coli 30S ribosomal subunit S3, a KH domain-containing protein, binds to the 16S rRNA (Urlaub et al., 1995). Sequencing of the cross-linked peptides after proteolytic digestion of the protein-RNA complexes has revealed that

one of the peptides is identical to a region of the KH domain, demonstrating that the KH domain of S3 directly contacts 16S rRNA (Urlaub et al., 1995). The cross-linked region corresponds to the α2 helix of the KH domain (Musco et al., 1996), which provide strong evidence that the helical side of the KH domain forms part of the RNA binding surface. The E. coli NusA protein has also been shown to contact nascent RNA in transcription complexes (Liu and Hanna, 1995).

The essential role of the KH domain in RNA binding has also been demonstrated by mutagenesis analysis. hnRNP K contains three KH repeats and preferentially binds cytidine-rich sequences (Matunis et al., 1992; Siomi et al., 1993a). It has been shown that deletion of any one of the three KH domains abolishes or strongly reduces poly(C) binding activity of hnRNP K in vitro. Similar effects were observed when the invariant Gly-X-X-Gly motif is altered in each KH domain (Siomi et al., 1994). FMR1 has been shown to bind a selective fraction of human brain RNA including its own message (Ashley et al., 1993), as well as homopolymeric RNAs, such as poly(G) and poly(U) (Siomi et al., 1993b). Introduction of the Ile-304→Asn mutation in the second KH domain of FMR1, which causes severe fragile X mental retardation syndrome, resulted in markedly impaired poly(U) binding in vitro. When the analogous mutation was introduced into the first KH domain of FMR1 (Ile-241→Asn) or any of the three KH domains of hnRNP K (Ile-65→Asn, Ile-167→Asn, or Ile-410→Asn), RNA binding activity was also severely impaired (Siomi et al., 1994). These studies demonstrated that all the KH domains present in hnRNP K and FMR1 contribute to RNA binding.

1.3 GSG DOMAIN

The GSG domain was initially identified by sequence alignment of the single KH domain-containing proteins GRP33 (Cruz-Alvarez and Pellicer, 1987), Sam68 (Wong et al., 1992), and GLD-1 (Jones and Schedl, 1995). The alignment showed that sequence conservation among these proteins is extended beyond the KH domain in both the amino- and carboxyl-terminal directions (Jones and Schedl, 1995). Thus, the GSG domain can be viewed as a tripartite protein module, consisting of a single KH domain and two flanking

regions called NK (for N-terminal to KH) and CK (for C-terminal to KH) regions, respectively. The GSG domain is also called the STAR (for signal transduction and activation of RNA) domain (Vernet and Artzt, 1997). Although some GSG domain-containing proteins have potential signalling motifs including tyrosine-rich and proline-rich regions, their role in signal transduction has not been established. Moreover, some GSG family members, such as GLD-1, do not contain any known signalling motifs, suggesting that the signalling property may not be a common feature for the GSG protein family.

With the discovery of new family members (Arning et al., 1996; Ebersole et al., 1996; Baehrecke, 1997; Fyrberg et al., 1997; Tanaka et al., 1997; Zaffran et al., 1997; Zorn et al., 1997; Di Fruscio et al., 1998; Fyrberg et al., 1998; Di Fruscio et al., 1999), the boundaries of the GSG domain have been redefined (Di Fruscio et al., 1998). The GSG domain consists of ~200 amino acids and the embedded KH domain consists of ~100 amino acids, which is approximately 26 amino acids longer than most other KH domains. The extra amino acids are located specifically in regions corresponding to loops 1 and 4; loop 1 contains six extra amino acids and loop 4 contains an additional 20 (Musco et al., 1996). Moreover, these "extended loops" have sequence conservation among GSG protein family members, unlike other KH domains that usually have highly variable sequences in these loops, implying a functional role specific for GSG proteins. Since loops 1 and 4 are not predicted to directly participate in RNA binding, these flexible loops, for the GSG protein family, may provide a site of interaction with other factors, such as proteins (Musco et al., 1996). However, it is possible that the structure of the "extended KH domain" has major differences from that of a regular KH domain. This issue will be resolved once the three-dimensional structure of the GSG domain is determined. In addition to the KH domain, the GSG domain contains a ~75-amino-acid NK region and a ~25-amino-acid CK region. These regions have been named QUA1 and QUA2 in the mouse GSG protein Qk1 (Ebersole et al., 1996). The function of different regions of the GSG domain is not known.

The physiological significance of the GSG domain has been demonstrated by the fact that genetic mutations in GSG proteins, many of which occur in the GSG domain, result in

growth or developmental defects. In the nematode *C. elegans*, the GSG protein GLD-1 functions as a tumor suppressor that is required for normal oocyte development (Francis et al., 1995a; Francis et al., 1995b). Thirty-two *gld-1* alleles have been identified and 19 of them result in 9 missense mutations in the GSG domain: 1 in the NK region, 6 in the KH domain, and 2 in the CK region (Jones and Schedl, 1995). Substitution of glycine 227 with aspartic acid or serine results in germ-line tumors (Jones and Schedl, 1995). The highly conserved glycine 227 residue is located at the predicted RNA binding site (loop 2) of the KH domain (Musco et al., 1996), suggesting that RNA binding is essential for GLD-1 function. Substitution of the conserved aspartic acid 310 with asparagine in the CK region also displays germ-line tumor phenotype (Jones and Schedl, 1995), indicating that the CK region may also contribute to RNA binding. Another missense mutation (Gly-308→Glu) in the CK region, however, shows a different phenotype: feminization of the germ line (Jones and Schedl, 1995). Substitution of glycine 248 or 250 with arginine in GLD-1 results in masculinization of germline (Mog) (Jones and Schedl, 1995). These mutations occur in the KH domain extended loop 4, which is not predicted to participate in RNA binding (Musco et al., 1996). Interestingly, several missense mutations in the GSG domain reverse the Mog phenotype caused by the Gly-248→Arg mutation. These mutations include Gly-200→Glu in the NK region, Pro-217→Leu in the KH domain extended loop 1, and Ala-294→Thr in the KH domain α3 helix (Jones and Schedl, 1995; Musco et al., 1996). Mouse Qk1 is required for myelination and embryonic survival. Two missense mutations in *quaking* (*qkI*) have been identified that result in embryonic lethality (Ebersole et al., 1996; Cox et al., 1999). These mutations, Glu-48→Gly and Val-157→Glu, occur in the NK region and the β3 strand of the KH domain, respectively. The *Drosophila melanogaster* protein How plays a critical role in muscle development (Baehrecke, 1997; Fyrberg et al., 1997; Zaffran et al., 1997). A missense mutation in How, replacing arginine 185 in the KH domain loop 4 with a cysteine, is a strong allele and results in embryonic lethality (Baehrecke, 1997). The molecular mechanisms by which these mutations cause their phenotypes are largely unknown.

1.4 GSG FAMILY MEMBERS

GSG domain-containing proteins have been identified from a variety of eukaryotic species, ranging from yeast to humans. Although the function of most GSG proteins remains unknown, evidence has been accumulating that GSG proteins associate with RNA and participate in RNA metabolism. Genetic and cellular studies have demonstrated that GSG proteins are involved in many essential processes such as pre-mRNA splicing (Arning et al., 1996), tumorigenesis (Jones and Schedl, 1995), apoptosis (Chen and Richard, 1998; Di Fruscio et al., 1998), cell cycle progression (Barlat et al., 1997), RNA transport (Nabel-Rosen et al., 1999), translational control (Jan et al., 1999), and development (Jones and Schedl, 1995; Baehrecke, 1997; Fyrberg et al., 1997; Zaffran et al., 1997; Zorn and Krieg, 1997) (Table 1-1).

1.4.1 Sam68 and Related Family Members

Sam68, the Src-associated-in-mitosis protein of 68 kDa, is a major substrate for Src-family tyrosine kinases during mitosis (Fumagalli et al., 1994; Taylor and Shalloway, 1994). Its cDNA was cloned and initially thought to encode p62, a tyrosine phosphoprotein that associates with p120 Ras GTPase-activating protein (GAP) and $p190^{rhoGAP}$ (Wong et al., 1992), but was later confirmed to encode Sam68 (Lock et al., 1996). The real GAP-associated protein p62 has been cloned and named $p62^{dok}$ (Carpino et al., 1997; Yamanashi and Baltimore, 1997). The human and mouse Sam68 proteins show over 95% sequence identity (Wong et al., 1992; Richard et al., 1995). The amino acid composition of Sam68 (443 amino acids) predicts a molecular mass of ~45 kDa. Its aberrant migration on SDS-polyacrylamide gels is likely due to the presence of highly negatively charged carboxyl terminus. Features of the Sam68 molecule include a tyrosine-rich region and a putative NLS at the C-terminus, at least five proline-rich motifs, a GSG domain, a potential arginine-glycine-glycine (RGG) box (two RGG repeats interspersed with four amino acids) close to the N-terminus, and a number of arginine-glycine repeats (Wong et al., 1992; Richard et al.,

Table 1-1. GSG family members and their functions

GSG proteins	Species	Possible roles and functions
Sam68	human, mouse	cell cycle progression signal transduction
SLM-1	mouse	cell cycle progression signal transduction
SLM-2	human, mouse	spermatogenesis cell proliferation
Sam50	*Drosophila*	Apoptosis
KEP1	*Drosophila*	Apoptosis
GRP33	*A. salina*	?
Qk1	mouse	myelination embryogenesis apoptosis
	chicken	spermatogenesis
	Zebrafish	?
Xqua	*Xenopus*	notochord development
How	*Drosophila*	muscle development tendon cell differentiation RNA nuclear export
GLD-1	*C. elegans*	germ cell differentiation tumor suppressor translational repressor
SF1	human	pre-mRNA splicing
BBP	yeast	pre-mRNA splicing

1995; Ishidate et al., 1997). Arginine-glycine repeats are potential sites of arginine methylation (Gary and Clarke, 1998). Indeed, amino acid analysis of recombinant Sam68 expressed in insect cells revealed that 9 of the 34 arginine residues in Sam68 are dimethylated (Wong et al., 1992).

Sam68 has been shown to associate with a wide variety of proteins. These proteins can be grouped into two major categories: signalling proteins and proteins that are involved in RNA metabolism.

Signalling molecules that interact with Sam68 include 1) kinases, such as Src-family tyrosine kinases (Fumagalli et al., 1994; Taylor and Shalloway, 1994; Richard et al., 1995), Itk/Tsk (Bunnell et al., 1996), the regulatory subunit of PI3K p85, and Jak3 (Fusaki et al., 1997), 2) other enzymes, such as phospholipase Cγ-1 (PLCγ-1) (Maa et al., 1994; Richard et al., 1995), and 3) adapter proteins, such as Grb2 (Richard et al., 1995), Nck (Lawe et al., 1997), and Grap (Trub et al., 1997). Most of these associations have been shown to occur through SH2- and SH3-mediated interactions. Src-family kinases are thought to associate with Sam68 only during mitosis (Fumagalli et al., 1994; Taylor and Shalloway, 1994). The cooperation between the SH3 and SH2 domains in this context has been demonstrated by mutagenesis experiments using the Src-family kinase $p59^{fyn}$ (Richard et al., 1995). Substitution of proline 134, a conserved residue that is required for SH3 domain function, with leucine in the SH3 domain of $p59^{fyn}$ remarkably reduces but not completely abolishes the ability of $p59^{fyn}$ to bind Sam68. On the other hand, abolition of the SH2 domain function of $p59^{fyn}$ by substituting arginine 176, a critical residue in phosphotyrosine-SH2 domain binding, results in an approximately two-fold reduction in $p59^{fyn}$-Sam68 complex formation. Introduction of both the Pro-134→Leu and the Arg-176→Lys substitutions in $p59^{fyn}$ abolishes its ability to bind to Sam68 (Richard et al., 1995). These results demonstrated that both the SH2 and SH3 domains of $p59^{fyn}$ contribute to the interaction between Sam68 and $p59^{fyn}$. Sam68 contains at least five proline-rich sequences named P1 to P5. Deletion studies suggested that the association of Sam68 with the $p59^{fyn}$ SH3 domain is mediated mainly by P5, but P3 and P4 also contribute to SH3 binding (Richard et

al., 1995). Binding to SH3 domain is a prerequisite for Sam68 tyrosine phosphorylation by Src-family kinases and subsequent interaction with their SH2 domain, since mutations in the SH3 domain of Src-family kinases impair tyrosine phosphorylation of Sam68 (Taylor and Shalloway, 1994; Weng et al., 1994; Richard et al., 1995). The adaptor protein Grb2 associates with Sam68 via its SH2 domain. Although the precise SH2 domain-binding sites have not been determined in Sam68, deletion studies suggested that the SH2 domains of $p59^{fyn}$ and Grb2 interact with distinct phosphotyrosine residues within the C-terminal tyrosine-rich region of Sam68 (Richard et al., 1995). The interactions of Sam68 with a number of signalling molecules support the hypothesis that Sam68 functions as an adaptor protein for tyrosine kinases, recruiting SH2 and/or SH3 domain-containing proteins and relaying signals to downstream effectors (Richard et al., 1995; Taylor et al., 1995).

Sam68 has been shown to associate with several proteins that are known to be involved in RNA metabolism. In poliovirus-infected cells, Sam68 relocalizes from the nucleus to the cytoplasm and associates with the poliovirus RNA-dependent RNA polymerase 3D, suggesting a possible role for Sam68 in poliovirus RNA replication (McBride et al., 1996). Recently, association of Sam68 with the polypyrimidine tract binding protein (PTB) has been observed by immunoprecipitation experiments (Grossman et al., 1998) and GST "pull-down" assays (Chen and Richard, unpublished data). PTB is involved in several aspects of RNA metabolism, including pre-mRNA splicing (Patton et al., 1993; Singh et al., 1995; Ashiya and Grabowski, 1997), splice site selection in alternative pre-mRNA splicing (Lin and Patton, 1995; Perez et al., 1997), 3' polyadenylation (Lou et al., 1996), and internal ribosome entry site (IRES)-mediated translation (Hellen et al., 1994; Kaminski et al., 1995; Witherell et al., 1995). Sam68 may play a role in some of these processes. However, the physiological significance and the mechanisms of these interactions still need to be determined.

The possible involvement of Sam68 in RNA metabolism is further supported by the finding that Sam68 interacts with RNA. Sam68 has been shown to bind single stranded RNA and single stranded and double stranded DNA as well as homopolymeric RNA in

vitro (Wong et al., 1992; Taylor and Shalloway, 1994). The RNA binding activity of Sam68 is regulated by tyrosine phosphorylation. Phosphorylation of Sam68 by the Src-family kinase p59^{fyn} abolishes homopolymeric RNA binding activity (Wang et al., 1995). The RNA binding activity of Sam68 is also modulated by SH3 domain binding, as binding of Src and p85 SH3 domains to Sam68 inhibits the ability of Sam68 to bind homopolymeric RNA (Taylor et al., 1995).

The physiological RNA targets for Sam68 are still unknown. Homopolymeric RNA binding experiments showed that Sam68 preferentially binds poly(U) and poly(A) (Taylor and Shalloway, 1994; Chen et al., 1997). Consistent with these results, Sam68 has been shown to selectively bind A/U-rich sequences using in vitro selection, and sequences with the highest affinity all contain a UAAA motif (Lin et al., 1997). Most recently, it is reported that Sam68 interacts with the Rev response element (RRE) present in HIV-1 mRNA (Reddy et al., 1999). These observations suggest that Sam68 binds RNA with a preference for certain sequences and/or structures.

Although the cellular function of Sam68 remains unknown, several lines of evidence implicate its involvement in cell cycle progression. Sam68 is tyrosine phosphorylated by Src-family kinases during mitosis (Fumagalli et al., 1994; Taylor and Shalloway, 1994). Sam68 is also phosphorylated on threonine residues by the cyclin-dependent protein kinase Cdc2 in a mitosis-specific manner (Resnick et al., 1997). These findings indicate that Sam68 may play a role in mitotic entry and progression. A rare natural isoform of Sam68 has been discovered that contains a 39-amino-acid deletion in the KH domain (called Sam68ΔKH isoform) (Barlat et al., 1997). The RNA-binding capability of Sam68ΔKH is severely reduced as compared to that of Sam68. The Sam68ΔKH isoform is specifically expressed at growth arrest upon confluency in cultured cells. Overexpression of Sam68ΔKH inhibits DNA synthesis and the inhibition can be rescued by coexpression of Sam68 (Barlat et al., 1997). These data suggest that Sam68 isoforms are also involved in the regulation of G1/S transition of the cell cycle. The fact that the two isoforms of Sam68 have opposing biological effects demonstrates the significance of the KH domain in Sam68

function.

Since Sam68 localizes predominantly in the nucleus (Wong et al., 1992; Ishidate et al., 1997), its interaction with Src-family kinases, which are primarily cytoplasmic, during mitosis is presumably facilitated by nuclear envelope breakdown. Alternatively, Sam68 may constantly shuttle between the nucleus and the cytoplasm, like many other nuclear proteins. Consistent with this possibility, Sam68 is observed in the membrane fraction of Src-transformed 3T3 cells (Wong et al., 1992) and it relocalizes from the nucleus to the cytoplasm in poliovirus-infected cells (McBride et al., 1996). Recently, it is shown that overexpressed Sam68 can functionally substitute for HIV-1 Rev in RRE-mediated gene expression and virus replication (Reddy et al., 1999). There is a putative NLS in the C-terminal region of Sam68, as deletion of the region results in cytoplasmic localization of Sam68 (Ishidate et al., 1997). Interestingly, C-terminally deleted Sam68 mutants prevent the nuclear import of Rev and inhibit HIV-1 replication (Reddy et al., 1999). These observations imply that Sam68 may play a role in RNA trafficking.

Two Sam68-like mammalian proteins, SLM-1 and SLM-2, have recently been identified (Di Fruscio et al., 1999). These proteins are similar to Sam68 in structural organization: a GSG domain followed by proline-rich motifs, arginine-glycine repeats, and a tyrosine-rich C-terminus. They have an ~70% sequence identity with Sam68 in their GSG domains. SLM-1 and SLM-2 are both nuclear proteins that heterodimerize with Sam68. They both bind RNA but with different specificities as determined by using homopolymeric RNA agarose. SLM-1, like Sam68, binds poly(U) and poly(A) whereas SLM-2 binds poly(G) and poly(A). SLM-1 shares several properties of Sam68: it binds the SH3 domains of $p59^{fyn}$ and PLCγ-1; it is tyrosine phosphorylated by Src-family kinases during mitosis; and, when phosphorylated, it binds the SH2 domains of $p59^{fyn}$, Grb2, $p120^{rasGAP}$, and PLCγ-1. Thus, SLM-1, like Sam68, may function as an adaptor protein for Src kinases during mitosis. In contrast, SLM-2 is not a substrate for Src-family kinases, and it does not associate with the SH3 domains of $p59^{fyn}$, Grb2, $p120^{rasGAP}$, and PLCγ-1 (Di Fruscio et al., 1999). The mouse *SLM-2* (also called *étoile*) gene maps to chromosome 15, while the

human *SLM-2* (also known as *T-STAR*) gene maps to the syntenic region on chromosome 8 (Venables et al., 1999). SLM-2 is highly expressed in human testis and has been shown to interact with RBM, an RNA binding protein implicated in spermatogenesis (Venables et al., 1999). SLM-2 is also highly expressed in the brain and skeletal muscle (Di Fruscio et al., 1999), suggesting that it may also function in these tissues. Recently, it is shown that overexpression of SLM-2 in chicken embryo fibroblasts down-regulates Sam68 expression and inhibits cell growth (Lee and Burr, 1999).

We have recently identified two *Drosophila* GSG proteins, KEP1 (KH encompassing protein 1) and Sam50 (Di Fruscio et al., 1998). Sam50 has a high sequence similarity with mammalian Sam68 and may therefore represent the *Drosophila* Sam68 homolog. Sam50 has an ~50% sequence identity with Sam68 in the GSG domain and it also has a tyrosine-rich region at its carboxyl terminus (Di Fruscio et al., 1998). Both KEP1 and Sam50 are RNA-binding proteins and their GSG domains are required for RNA binding. In vitro homopolymeric RNA binding experiments revealed that Sam50 has a preference for poly(U) and poly(A), like Sam68, whereas KEP1 binds only poly(U). KEP1 localizes predominantly in the nucleus and Sam50 localizes in the cytoplasm. Overexpression of either KEP1 or Sam50 in *Drosophila* S2 cells results in activation of caspases and apoptosis, suggesting that these proteins may play a role in apoptosis (Di Fruscio et al., 1998).

A Sam68 homolog has been identified from Torpedo californica. It is identified as an interacting partner for rapsyn, a synaptic peripheral membrane protein that mediates clustering of nicotinic acetylcholine receptors on the surface of myotubes in response to neuronally derived factors. This Sam68-like protein also colocalizes with rapsyn at the neuromuscular junction (Fung et al., 1998).

GRP33, glycine-rich protein of 33 kDa (Cruz-Alvarez and Pellicer, 1987), from the brine shrimp *Artemia salina*, is also related to Sam68; it has a 38% sequence identity with Sam68 in the GSG domain (Di Fruscio et al., 1998). Although its biochemical properties and biological functions are unknown, GRP33 shares a number of characteristic features of

core hnRNP proteins: a high content (16.6%) of glycine residues, very few cysteines (2 out of 308 amino acids), and multiple arginine-glycine repeats that are potential sites of arginine methylation (Cruz-Alvarez and Pellicer, 1987).

1.4.2 Qk1 and Related Family Members

Qk1 is encoded by the mouse *quaking* (*qkI*) gene which is involved in early embryogenesis and myelination. Northern blot analysis indicated that transcription of the *qkI* gene produces at least three messages of 5, 6, and 7 kb (Ebersole et al., 1996). An additional transcript, *qkI-G*, has recently been identified (Cox et al., 1999). The resulting proteins, termed Qk1-5, Qk1-6, Qk1-7, and Qk1-G differ only at their C termini (Ebersole et al., 1996; Cox et al., 1999). Genomic structure analysis revealed that the *qkI* gene is composed of six common exons and four alternative ends. The Qk1-7 terminal exon encodes 14 amino acids and a very long 3'UTR. The Qk1-6 terminal exon is wholly contained within the longer C-terminal Qk1-7 exonic sequence and encodes 8 amino acids. The C-terminus of Qk1-5 is encoded by two exons: the first 25 amino acids are encoded by an exon wholly contained within the C-terminal Qk1-7 exonic sequence, and the remaining 5 amino acids and the 3'UTR are encoded by a unique exon. The C terminus of Qk1-G is composed of 8 amino acids and is encoded by a sequence that is intronic in the context of the other isoforms (Cox et al., 1999). Immunocytochemistry studies revealed that the Qk1-5, Qk1-6, and Qk1-7 isoforms are highly expressed in myelinating oligodendrocytes and Schwann cells as well as astrocytes, but not in neurons. In Qk1-expressing cells, individual isoforms displayed distinct intracellular localizations: Qk1-6 and Qk1-7 are localized mainly in the cytoplasm, and Qk1-5 is restricted to the nucleus (Hardy et al., 1996). These findings indicate that the alternative C termini determine Qk1 protein localization. Indeed, the Qk1-5 C terminus has been shown to function as an NLS (Wu et al., 1999). In addition to the GSG domain, Qk1 contains two proline-rich motifs and a cluster of five tyrosine residues close to the C terminus, suggesting a role for Qk1 in the regulation of RNA metabolism and possibly also in signal transduction. While a recent report indicates that

Qk1-6 can functionally substitute for GLD-1 in *C. elegans* as a translational repressor (Saccomanno et al., 1999), experimental evidence for a function of Qk1 in signal transduction is not available.

Two classes of mutations have been identified in the mouse *quaking* locus located on chromosome 17. The *quaking viable* (qk^v) mutation, first described by Sidman and coworkers, is a spontaneous, autosomal recessive mutation in mice characterized by a severe deficiency of myelin in the central nervous system (Sidman et al., 1964) and also a mild dysmyelination in the peripheral nervous system (Samorajski et al., 1970). As a consequence, qk^v homozygotes develop a characteristic tremor (quaking) by postnatal day 10. The molecular defect of the qk^v mutation is a large (>1 Mb) deletion that lies only 1.1 kb away from the start of the *qkI* coding region (Ebersole et al., 1996) and therefore might include some of the *qkI* enhancer/promoter elements. Immunolabelling with isoform-specific antibodies revealed that in qk^v mice Qk1-6 and Qk1-7 are entirely absent in myelin-forming oligodendrocytes and Schwann cells but their expression is normal in astrocytes. Qk1-5 expression in these cells is generally not affected (Hardy et al., 1996). The molecular mechanism by which the qk^v deletion leads to alterations of isoform- and cell type-specific expression of Qk1 proteins is not known.

The second class of mutations in the *quaking* locus is the four N-ethyl-N-nitrosourea (ENU)-induced alleles, qk^{l-1}, qk^{kt1}, qk^{k2}, and $qk^{kt3/kt4}$ (Bode, 1984; Justice and Bode, 1986; Justice and Bode, 1988; Shedlovsky et al., 1988). The qk^{kt3} and qk^{kt4} isolates are suspected to be identical alleles because they were isolated as littermates from the same mutagenized father. These alleles are all recessive embryonic lethal mutations; homozygotes and hemizygotes die at 8 to 9.5 days gestation (Justice and Bode, 1988; Shedlovsky et al., 1988). Embryological studies have shown that there is a developmental block at 15 to 26 somite stage and the mutant embryos are smaller and disorganized (Justice and Bode, 1988). These findings suggest that Qk1 is essential for embryogenesis and may play a role in cell proliferation and differentiation. The qk^v mutation complements the lethality of the ENU-induced alleles but these alleles fail to complement the myelination defect of qk^v

(Justice and Bode, 1988; Shedlovsky et al., 1988), indicating that both phenotypes are controlled by the same gene. Indeed, all the ENU-induced alleles except one have been found to contain single nucleotide alterations that result in missense mutations or a splice site change in Qk1. Both qk^{kt3} and qk^{kt4} change glutamic acid 48 to a glycine residue (Ebersole et al., 1996; Cox et al., 1999), confirming that they are identical alleles. Glutamic acid 48 is a highly conserved residue in the NK (QUA1) region of the GSG domain (Ebersole et al., 1996; Di Fruscio et al., 1998). qk^{k2} also contains a missense mutation in Qk1, altering valine 157 in the KH domain to a glutamic acid (Cox et al., 1999). Valine 157, located at the end of the β3 strand of the KH domain structure, is conserved in the Qk1 homologs of different species and also in Sam68 (Di Fruscio et al., 1998). The Val-157→Glu mutation is analogous to the Gly-296→Arg mutation in Bic-C and is predicted to destabilize the overall structure of the KH domain (Musco et al., 1996). $qk^{l\text{-}1}$ contains a point mutation immediately downstream of the coding sequence of the alternative isoform Qk1-6, resulting in a new splice site that potentially eliminates the Qk1-5 isoform (Cox et al., 1999). The molecular defect of qk^{kt1} is unknown but is not located in the coding region of Qk1 (Ebersole et al., 1996; Cox et al., 1999). The severe effect of missense mutations in the NK region or the KH domain indicates that the GSG domain is critical for the function of Qk1 proteins.

Although genetic evidence suggests that Qk1 is required for myelination and embryonic survival, its biochemical properties and cellular functions are poorly understood. The distinct intracellular localization of individual Qk1 isoforms suggests that they may participate in different cellular processes. The expression of these isoforms is strictly controlled during development. Northern blot analysis showed that the 5 kb *qkI* transcript appears early during embryogenesis and its expression declines after birth. *Qk1-5* is the only isoform expressed at the time homozygotes for the ENU-induced alleles die. In contrast, the 6 kb and 7 kb *qkI* transcripts are expressed late in development when myelination starts and their expression is maintained throughout postnatal development into adulthood (Ebersole et al., 1996; Hardy et al., 1996). It has been speculated that Qk1-5 is involved in early embryogenesis and Qk1-6 and Qk1-7 are essential for myelination (Vernet

and Artzt, 1997). Qk1 is also expressed in organs outside the nervous system including the heart, lung and testes (Ebersole et al., 1996), indicating that Qk1 has a more universal role both during development and in mature animals.

Qk1 homologs have been identified from a number of species (Baehrecke, 1997; Fyrberg et al., 1997; Tanaka et al., 1997; Zaffran et al., 1997; Zorn et al., 1997; Mezquita et al., 1998). The cDNAs encoding *Xenopus*, zebrafish, and chicken Qk1 homologs exhibit striking sequence conservation with that of mouse Qk1 in their coding regions (Tanaka et al., 1997; Zorn et al., 1997; Mezquita et al., 1998). Similar to mouse *qkI*, these *qkI* homologs each produces multiple alternatively spliced isoforms, which are expressed differentially during development (Lo and Frasch, 1997; Tanaka et al., 1997; Zorn et al., 1997; Mezquita et al., 1998). The *Xenopus* Qk1 homolog, Xqua, has been shown to play an essential role in notochord development (Zorn and Krieg, 1997). The chicken Qk1 has been implicated in spermatogenesis (Mezquita et al., 1998).

The *Drosophila qkI* homolog, *How* (held-out wings), was cloned by several independent groups and it is also called *who* (wings held-out), qkr^{93F} (*quaking*-related 93F), and *struthio* (Baehrecke, 1997; Fyrberg et al., 1997; Lo and Frasch, 1997; Zaffran et al., 1997). The *How* gene encodes two alternatively spliced isoforms of proteins, How(L) and How(S), that differ only at their C termini (Lo and Frasch, 1997). How(L) is a nuclear protein and How(S) is present in both the nucleus and the cytoplasm (Nabel-Rosen et al., 1999). *How* is expressed in muscle and muscle attachment site cells during both embryogenesis and metamorphosis (Baehrecke, 1997; Fyrberg et al., 1997; Lo and Frasch, 1997; Zaffran et al., 1997). *How* loss-of-function mutants die late during embryogenesis. On the cellular level, these mutants have differentiated and fused myotubes, but they fail to migrate and attach to the epidermal attachment sites. Weak *How* mutants die during metamorphosis with their head stuck in their thorax. The weakest mutants survive as adults that do not fold their wings properly, resulting in the "held-out-wing" phenotype. One strong allele, who^{e44}, was found to contain a missense mutation in How, altering arginine 185 to a cysteine (Baehrecke, 1997). Arginine 185 is located in the extended loop 4 of the

KH domain (Musco et al., 1996) and it is conserved in most GSG family members (Di Fruscio et al., 1998). These findings indicate that How is involved in muscle development and the GSG domain is required for its normal function. How is an RNA-binding protein and has been shown to bind poly(A) in vitro (Di Fruscio et al., 1998). Most recently, Nabel-Rosen and coworkers showed that How binds *stripe* mRNA and regulates Stripe expression (Nabel-Rosen et al., 1999). Stripe is a transcription factor that plays a key role in the regulation of tendon cell differentiation (Volk, 1999).

C. elegans GLD-1 (defective in germ-line development) is closely related to Qk1, their GSG domain sequences are ~60% identical and ~80% similar (Jones and Schedl, 1995). Genetic studies indicate that the *gld-1* gene functions in several aspects of germ-line development (Francis et al., 1995a; Francis et al., 1995b). In *C. elegans*, animals with a single X chromosome (XO) develop as males that produce sperm in the germ-line, whereas animals with two X chromosomes (XX) develop as hermaphrodites that make both sperm and oocytes in the germ-line (Meyer, 1997). *gld-1* plays an essential role in directing oocyte differentiation and meiotic prophase progression during oogenesis. In *gld-1 null* hermaphrodites, the oocyte germ-line fails to progress through meiosis and inappropriately re-enters mitosis, leading to the overproliferation of germ-line cells and the formation of germ-line tumors. In this sense, *gld-1* acts as a tumor suppressor (Francis et al., 1995a). *gld-1* also has a nonessential function in negatively regulating proliferation of premeiotic germ cells in both males and hermaphrodites (Francis et al., 1995b). In addition, although *gld-1* is not required for sperm formation in either sex, it plays a role in promoting spermatogenesis in the hermaphrodite germ line (Francis et al., 1995b). Thirty-two *gld-1* mutations were grouped into six classes based on the germ-line phenotypes observed in homozygotes (Francis et al., 1995a; Jones and Schedl, 1995). Class A alleles are null or strong loss-of-function mutations. In homozygote hermaphrodites, female germ cells never undergo oogenesis but, instead, form a germ-line tumor (Francis et al., 1995a). Three class A mutations (2 deletions and 1 splice site mutation) result in reading frameshifts near the N terminus of GLD-1. All the other 8 class A alleles contain single nucleotide changes in the

exons encoding the GSG domain, resulting in missense mutations or premature stop codons. Three of the missense mutations affect a single residue (Gly-227→Asp or Gly-227→Ser) and alter the invariant GXXG motif in the KH domain loop 2 (Jones and Schedl, 1995). The Gly-227→Asp mutation is predicted to affect RNA binding (Musco et al., 1996) and, indeed, has been shown to impair the ability of GLD-1 to bind RNA in vitro (Jan et al., 1999), indicating that RNA binding is essential for GLD-1 function. Another missense mutation (Asp-310→Asn) alters a highly conserved residue in the CK region of the GSG domain (Jones and Schedl, 1995), implying that the CK region is also important for the normal function of GLD-1. Class B alleles are partial loss-of-function mutations. Female germ cells in homozygote hermaphrodites arrest in pachytene of meiosis I (Francis et al., 1995a). Most class B alleles are intragenic revertants of one class C allele and contain both the class C mutation (Gly-248→Arg) and an additional change in the GSG domain. These changes include three missense mutations: Gly-200→Glu in the NK region, Pro-217→Leu in the KH domain loop 1, and Ala-294→Thr in the KH domain α3 helix (Jones and Schedl, 1995). Proline 217 and alanine 294 are conserved in almost all GSG proteins, and glycine 200 is conserved in Qk1 and its homologs but not in other GSG family members (Di Fruscio et al., 1998). Class C alleles are both gain-of-function and partial loss-of-function mutations that increase hermaphrodite sperm production but fail to switch to oogenesis, resulting in a masculination of the germ-line (Mog) phenotype (Francis et al., 1995a). Most class C alleles result in Gly-248→Arg or Gly-250→Arg point mutation (Jones and Schedl, 1995). These highly conserved glycine residues (Di Fruscio et al., 1998) are located at the beginning of the KH domain loop 4 and are not predicted to be involved in RNA binding (Musco et al., 1996). The fact that the Mog phenotype of Gly-248→Arg mutation is completely suppressed by other mutations in the GSG domain (resulting in class B phenotype) suggests that the gain-of-function activity of class C alleles is dependent on a functional GSG domain. Class D and class E alleles eliminate hermaphrodite spermatogenesis, resulting in feminization of the germ-line (Fog) phenotypes. The difference between class D and class E is that class D alleles form normal oocytes and they

also feminize the XO male germ line, whereas the class E allele shows an abnormal oocyte phenotype and it does not affect XO males (Francis et al., 1995a). Interestingly, both class D and class E alleles alter amino acids in the CK region of the GSG domain (Jones and Schedl, 1995), further supporting the significance of the CK region in GLD-1 function. The class F allele also displays an abnormal oocyte phenotype but does not have a Fog phenotype (Francis et al., 1995a). Its molecular lesion is the insertion of the transposable element Tc2 in the *gld-1* 3'UTR, possibly affecting the stability of *gld-1* mRNA (Jones and Schedl, 1995). Taken together, these genetic data demonstrated that the GSG domain is essential for GLD-1 function and different regions of the GSG domain may have distinct roles. GLD-1 is a germ-line-specific cytoplasmic protein (Jones et al., 1996). There is recent evidence that GLD-1 represses *tra-2* translation by binding to a regulatory sequence in the 3'UTR of *tra-2* mRNA (Jan et al., 1999). *tra-2* plays a critical role in sex determination in both somatic tissues and the germ line (Hodgkin and Brenner, 1977; Hodgkin, 1980; Kuwabara et al., 1992). Mouse Qk1-6 can replace GLD-1 in repressing *tra-2* translation in *C. elegans* (Saccomanno et al., 1999), suggesting that GLD-1 and Qk1 are also functional homologs.

1.4.3 SF1/BBP

Human SF1 is a splicing factor that plays an important role in spliceosome assembly (Kramer, 1992). It is also known as ZFM1 (zinc finger gene in the *MEN1* [multiple endocrine neoplasia type 1] locus) (Toda et al., 1994). SF1 contains two structural motifs implicated in RNA binding: a GSG domain and a zinc knuckle. The zinc knuckle motif is characteristic for retroviral nucleocapsid proteins (Berg, 1986; Rein, 1994; Darlix et al., 1995) and is also present in some splicing factors, such as yeast Slu7p and the human serine/arginine (SR) protein 9G8 (Frank and Guthrie, 1992; Cavaloc et al., 1997). SF1 was shown to bind RNA in a sequence-independent fashion with a preference for poly(G) and poly(U) (Arning et al., 1996). Deletion studies indicate that the GSG domain is required for RNA binding and is essential for the activity of SF1 in spliceosome assembly (Rain et al.,

1998). The N-terminal region of SF1 mediates the interaction between SF1 and U2AF^{65} (Berglund et al., 1998a; Rain et al., 1998), which binds to the polypyrimidine tract upstream of the 3' splice site (Michaud and Reed, 1991; Bennett et al., 1992; Zamore et al., 1992; Michaud and Reed, 1993). The C-terminal half of SF1 is rich in proline residues (Arning et al., 1996) and has been shown to be the binding sites for the WW domains of formin-binding protein FBP11 and the SH3 domain of the oncoprotein Abl (Bedford et al., 1997). Different isoforms of SF1 that are derived from alternatively spliced mRNAs are expressed in various mammalian cell types (Toda et al., 1994; Arning et al., 1996; Caslini et al., 1997; Wrehlke et al., 1997). These isoforms differ in the length of the proline-rich region and distinct C termini. The yeast (*Saccharomyces cerevisiae*) SF1 homolog, BBP (branchpoint binding protein), was independently identified as an interacting partner for Mud2p, the putative U2AF^{65} homolog in yeast, in a synthetic lethal screen (Abovich and Rosbash, 1997) and in an exhaustive genomic two-hybrid screen (Fromont-Racine et al., 1997). Genetic and biochemical evidence showed that BBP interacts with Mud2p and the yeast U1 small nuclear RNP (snRNP) protein PRP40p (Abovich and Rosbash, 1997). BBP and SF1 have similar structural organizations, but BBP has two zinc knuckles whereas SF1 has only one (Arning et al., 1996; Abovich and Rosbash, 1997). BBP has been shown to interact specifically with the pre-mRNA branchpoint sequence (BPS) UACUAAC. Recombinant SF1 also recognizes BPS but with less specificity than that of BBP (Berglund et al., 1997). In vitro RNA-binding analysis using truncated and chimeric proteins indicated that the GSG domain of SF1/BBP is sufficient for specific recognition of BPS, but the zinc knuckle(s) are also required for optimal RNA binding (Berglund et al., 1998b). These studies demonstrated that SF1/BBP is involved in protein-protein and protein-RNA interactions that bridge the 3' and 5' splice-site ends of the intron during the early steps of pre-mRNA splicing.

1.5 NUCLEAR STRUCTURES

Similar to many molecular processes in the cytoplasm, which take place in or on

subcellular organelles, nuclear functions occur in specialized subnuclear structures. The most prominent and best characterized nuclear structure is the nucleoli, which are the sites where rRNAs are transcribed, processed, and assembled into ribosomal subunits. A number of other morphologically distinct subnuclear structures have also been identified, including several nuclear bodies (Lamond and Earnshaw, 1998). Although the precise functions of these subnuclear structures are largely unknown, their significance is underscored by the fact that alterations of these structures are often associated with human diseases.

1.5.1 Coiled Bodies (Cajal Bodies)

Coiled bodies, also known as Cajal bodies, were first described by Santiago Ramon y Cajal, who termed them nucleolar 'accessory bodies' (Cajal, 1903). They appear as round structures of 0.1-1.0 μm in diameter with coiled threads when viewed with the electron microscope (Monneron and Bernhard, 1969). Coiled bodies are highly enriched with several classes of snRNPs, nucleolar and cell-cycle control proteins, as well as several basal transcription factors (Gall et al., 1999; Matera, 1999). One of the proteins, called p80 coilin, is widely used as a marker for coiled bodies. Immunostaining studies using anti-coilin antibodies indicate that coiled bodies are evolutionarily conserved structures present in both plants and animals. Although mature snRNPs are highly concentrated in coiled bodies, nascent snRNPs do not accumulate within them, suggesting that coiled bodies may be involved in late stages of snRNP biogenesis and trafficking. A recent hypothesis suggests that coiled bodies are assembly centers for transcription and processing machinery (Gall et al., 1999). Each mammalian cell nucleus contains 1 to 10 coiled bodies and this number increases remarkably in cancer cells (Spector et al., 1992), suggesting that coiled bodies may be linked to cell transformation.

1.5.2 Gems

Gemini of coiled bodies, or gems, are nuclear structures frequently found in close proximity with coiled bodies. Gems and coiled bodies are very similar in number, size and

responses to changes in metabolic states (Liu and Dreyfuss, 1996). Recent studies indicate that, in some cell lines, gems and coiled bodies are indistinguishable (Matera and Frey, 1998; Bechade et al., 1999). Gems contain the spinal muscular atrophy gene product SMN (survival of motor neurons) and an associated proteins, SIP1 (SMN-interacting protein 1) (Liu and Dreyfuss, 1996; Liu et al., 1997). In addition to accumulating in gems in the nucleus, SMN and SIP1 are also distributed diffusely in the cytoplasm and appear to play a key role in the cytoplasmic steps of U snRNP biogenesis (Fischer et al., 1997; Liu et al., 1997). It is possible that the SMN-SIP1 complexes mediate the initial phases of snRNP assembly in the cytoplasm and then accompany the assembled snRNPs to the nucleus. Mutational analysis of SMN suggests a nuclear function for this protein, because overexpression of a truncated form of SMN, called SMNΔN27, results in reorganization of gems and coiled bodies. SMNΔN27 also inhibits pre-mRNA splicing in vitro (Pellizzoni et al., 1998). These results suggest that, in addition to a role in snRNP assembly, SMN may also participate in nuclear snRNP function. A third component of gems, Gemin3, has recently been identified as an SMN-interacting protein. Gemin3 contains a DEAD box motif, suggesting that it may function as an RNA helicase (Charroux et al., 1999).

1.5.3 PML Nuclear Bodies

PML nuclear bodies, also known as PODs (PML oncogenic domains), Kr bodies, or ND10 (nuclear domain 10), are nuclear matrix-associated structures of 0.3-1.0 μm in diameter (Dyck et al., 1994; Weis et al., 1994). They comprise a dense ring that contains the promyelocytic leukemia protein, PML, surrounding a central core that does not. In addition to PML, PML nuclear bodies contain a number of other proteins, including the retinoblastoma protein RB1, the autoantigen SP100, and the ubiquitin-like protein SUMO-1 (Lamond and Earnshaw, 1998; Matera, 1999). PML nuclear bodies have been shown or proposed to be involved in many important cellular processes, such as transcription, oncogenesis, and viral replication. Each mammalian cell normally contains 10 to 20 PML nuclear bodies. In acute promyelocytic leukemia (APL), a specific t(15; 17) translocation

fuses the PML gene to the retinoic acid receptor alpha (RARα) gene (de The et al., 1991; Kakizuka et al., 1991). As a consequence, PML nuclear bodies in APL cells are highly disorganized into micropunctate structures. Interestingly, retinoic acid and arsenic trioxide, both of which are used in clinical treatment of APL, induce the reformation of PML nuclear bodies (Dyck et al., 1994; Koken et al., 1994; Weis et al., 1994). PML functions as a negative growth regulator and as a tumor suppressor. *Pml*$^{-/-}$ mice are viable but are more susceptible to tumor formation and viral infections (Wang et al., 1998a). Recently, PML has been shown to mediate apoptotic cell death (Quignon et al., 1998; Wang et al., 1998b). It is proposed that the PML-RARα fusion protein in APL blocks the action of wild-type PML, thus conferring a survival advantage to APL cells and leading to leukemia (Quignon et al., 1998; Wang et al., 1998b).

1.5.4 PNC

The perinucleolar compartment (PNC) is a unique nuclear structure preferentially localized at the periphery of the nucleolus (Ghetti et al., 1992; Matera et al., 1995). It is an irregularly shaped structure with sizes ranging from 0.25 to 1 μm in diameter. Electron microscopic examination of fixed HeLa cells revealed that the PNC is composed of multiple thick, electron-dense strands, each measuring ~80-180 nm in diameter. Some of these strands are in direct contact with the surface of the nucleolus (Huang et al., 1997). Several small RNAs transcribed by RNA polymerase III, including RNase MRP RNA, RNase P RNA, and multiple hY RNAs (Matera et al., 1995; Lee et al., 1996a), as well as two hnRNP proteins, PTB (hnRNP I) (Ghetti et al., 1992) and CUG-BP/hNab50 (Timchenko et al., 1996), have been identified in the PNC. PTB, a 57-kDa RNA-binding protein that specifically binds pyrimidine-rich sequences (Ghetti et al., 1992), has been implicated in multiple cellular functions. CUG-BP/hNab50 has been shown to bind to the CUG triplet repeats of myotonin protein kinase RNA, which is associated with myotonic dystrophy, an autosomal dominant neuromuscular disease (Timchenko et al., 1996). The phosphorylation and intracellular distribution of CUG-BP/hNab50 are altered in patients with myotonic dystrophy and in a myotonin protein kinase knockout mouse (Robert et al., 1997). Using

PTB as a marker for the PNC, Spector and colleagues showed that the PNC is predominantly found in transformed cells and rarely observed in normal cells (Huang et al., 1997), and this nuclear structure coincides with sites of active transcription (Huang et al., 1998).

1.5.5 Speckles

Spliceosomal snRNPs and other pre-mRNA splicing factors are localized in a punctate ("speckled") nuclear distribution pattern when analysed by immunofluorescence microscopy (Lamond and Earnshaw, 1998). The speckled pattern results from the association of snRNPs with several structures previously visualized by electron microscopy, including perichromatin fibrils (PFs) and interchromatin granule clusters (IGCs) (Spector, 1993). PFs are fibrillar structures measuring 3 to 5 nm in diameter and are often found at the periphery of IGCs and distributed throughout the nucleoplasm (Monneron and Bernhard, 1969). Several lines of evidence suggests that PFs represent nascent RNA transcripts (Fakans, 1994). IGCs are the major part of the speckled staining pattern. IGCs measure 0.3 to 1.8 μm along their longest dimension and are composed of granules measuring 20 to 25 nm in diameter that appear to be connected by a 9 to 10 nm fiber (Fakan and Puvion, 1980; Spector, 1993). IGCs are not centers of active transcription, but they are often in close proximity to highly active genes and their transcripts (Huang and Spector, 1997). Recent evidence indicates that IGCs may function as depots supplying splicing factors to sites of active transcription (Lamond and Earnshaw, 1998). The phosphorylation of the RS domain of SR proteins is involved in releasing splicing factors from IGCs and their dephosphorylation is, at least, part of the signal for the return of these factors to the IGCs (Misteli and Spector, 1996; Misteli et al., 1998). Most recently, the C-terminal domain (CTD) of the large subunit of Pol II has been shown to be involved in the intranuclear targeting of splicing factors to transcription sites in vivo (Misteli and Spector, 1999).

1.6 RNA PROCESSING

All major eukaryotic RNA types, messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA), are transcribed initially in precursor forms, which are typically larger than the mature RNA products. The primary transcripts are then processed by several biochemical steps. One step involves clipping and splicing reactions that remove surplus nucleotide sequences from the precursor. Additional processing steps may add nucleotides to one or both ends of the precursor and chemically modify individual bases within the precursor. In this section, I will describe the major steps of mRNA processing. The precursor of mRNA, called heterogeneous nuclear RNA (hnRNA) or pre-mRNA, is transcribed by RNA polymerase II (Pol II). The typical reactions processing pre-mRNA include the addition of 5' cap (capping), removal of introns (splicing), and addition of 3' poly(A) tail (polyadenylation). These sequential and interrelated reactions are closely coupled with transcription, as evidenced by the recent observations that many protein factors essential for RNA processing directly associate with the CTD of the large subunit of Pol II (Neugebauer and Roth, 1997; Steinmetz, 1997).

1.6.1 5' Capping

The 5' cap structure, $m^7G(5')ppp(5')N$, consists of an inverted 7-methyl guanosine linked via a 5'-5' triphosphate bridge to the first transcribed residue, and is characteristic of all Pol II transcripts (Shatkin, 1976). The capping reaction proceeds in three sequential steps (Shuman, 1995). First, the terminal phosphate from the 5'-triphosphate terminus of the nascent transcript is removed by RNA 5'-triphosphatase. Then, RNA guanylyltransferase transfers GMP from GTP to the diphosphate end of RNA to form the GpppN cap. Finally, RNA (guanine-7)-methyltransferase transfers a methyl group from S-adenosylmethionine to the N7 position of the cap guanine. The cap structure is added shortly after transcription initiation (Salditt-Georgieff et al., 1980; Coppola et al., 1983; Jove, 1984; Rasmussen and Lis, 1993) when the capping enzymes are recruited to the transcription complex by binding to the phosphorylated CTD of Pol II (Cho et al., 1997;

McCracken et al., 1997).

The cap structure plays an important role in many aspects of RNA metabolism. It enhances pre-mRNA splicing (Konarska et al., 1984; Edery and Sonenberg, 1985; Izaurralde et al., 1994), polyadenylation (Cooke and Alwine, 1996), RNA export (Hamm and Mattaj, 1990; Izaurralde et al., 1992; Izaurralde et al., 1995a), RNA stability and translation (Shatkin, 1985). These effects are mediated by proteins that recognize and bind the cap structure. Most of the nuclear functions of the cap structure are mediated by a nuclear cap-binding protein complex (CBC) (Lewis and Izaurralde, 1997), a heterodimeric complex comprising a cap-binding protein of 80 kDa, CBP80, and a cap-binding protein of 20 kDa, CBP20 (Izaurralde et al., 1994). CBP80 contains a functional NLS at its N terminus (Kataoka et al., 1994), which is required for active nuclear uptake of CBC (Izaurralde et al., 1995b). CBP20 consists essentially of an RNA-binding domain (RBD) followed by a C-terminal tail rich in glycine and arginine residues (Izaurralde et al., 1995; Kataoka et al., 1995). CBC binding stabilizes the association of U1 snRNP with the cap-proximal 5' splice site to commit the pre-mRNA to the spliceosome assembly pathway and ultimately to splicing (Lewis et al., 1996). CBC also interacts with the 3'-end processing machinery, enhancing the efficiency of poly(A) site cleavage prior to the synthesis of poly(A) tail (Lewis and Izaurralde, 1997). Moreover, CBC may accompany capped RNAs, mRNAs or U snRNAs, to the cytoplasm during nuclear export. CBC binding to the cap structure facilitates mRNA export and is required for U snRNA export (Lewis and Izaurralde, 1997). In the cytoplasm, the cap of the RNAs is bound by the cytoplasmic cap-binding protein, eIF4E (eukaryotic translation initiation factor 4E), which is a subunit of the translation initiation factor eIF4F (Sonenberg and Gingras, 1998).

1.6.2 Pre-mRNA Splicing

Most protein-coding genes in eukaryotes contain intervening sequences, or introns, that are eliminated from pre-mRNAs during gene expression. Splicing is the process by which these introns are precisely removed and the flanking functional sequences (exons) are

joined together. Splicing occurs in a large and dynamic ribonucleoprotein complex called the spliceosome. Many components of the spliceosome, including snRNPs and non-snRNP splicing factors, have been shown to associate with the hyperphosphorylated form of Pol II (Chabot et al., 1995; Mortillaro et al., 1996; Vincent et al., 1996; Kim et al., 1997), which plays an important role in coupling splicing and transcription both spatially and temporally.

The accuracy of splicing depends on multiple *cis*-acting elements on pre-mRNAs. The 5' splice site in higher eukaryotes conforms to the consensus sequence AG|GURAGU and the 3' splice site is characterized by the sequence YAG| (the splice sites are denoted by vertical bars and invariant nucleotides are underlined; R=purine, Y=pyrimidine). Another critical sequence element in the intron is the branchpoint sequence (BPS), which is usually located 15-50 nucleotides upstream of the 3' splice site. The BPS is almost invariant (UACUAAC) in the yeast *Saccharomyces cerevisiae*, but it is more degenerate in mammalian pre-mRNAs, displaying a consensus sequence of YNCURAY (the site of branch formation is underlined; N=any nucleotide). In addition, most introns of higher eukaryotes contain a region of high pyrimidine content (polypyrimidine tract) located between BPS and the 3' splice site. Although pre-mRNA splicing is a fundamentally RNA-catalyzed process, splicing factors participate in each step of the reaction, and complex networks of RNA-RNA, RNA-protein, and protein-protein interactions contribute to the specificity of splicing. Protein factors function in the recognition, selection, and juxtaposition of the splice sites and drive conformational changes during spliceosome assembly and catalysis (Madhani and Guthrie, 1994; Kramer, 1996; Reed, 1996; Will and Luhrmann, 1997).

1.6.2.1 The splicing reaction

Splicing proceeds via two consecutive transesterification steps. In the first stage, a cleavage is made at the 5' splice site, releasing the 5' exon and resulting in the formation of a branched molecule, the lariat intermediate. This intermediate contains an unusual 2', 5'-phosphodiester bond between the branchpoint adenosine and the 5' terminal guanosine of

the intron. In the second stage, a cleavage at the 3' splice site releases the free intron in lariat form, while the 5' exons is ligated to the 3' exon (Ruskin et al., 1984; Konarska et al., 1985; Moore et al., 1993).

1.6.2.2 Spliceosomal snRNP biogenesis

The splicing apparatus consists of several snRNPs. Each spliceosomal snRNP contains one (U1, U2, and U5) or two (U4/U6) snRNA molecules and a number of proteins. The proteins can be divided into two general classes: specific proteins that are uniquely associated with one of the snRNAs and common proteins that associate with all snRNAs (Luhrmann et al., 1990). There are at least 8 common proteins, termed B, B', D1, D2, D3, E, F, and G, and they all contain a related domain, the Sm domain, so named because it is recognized by autoimmune sera of the Sm serotype (Mattaj and Nagai, 1995). While the snRNP-specific proteins have specific functions in the splicing reaction, the Sm proteins mainly function in the biogenesis of spliceosomal snRNPs. Synthesis and processing of U snRNAs and their assembly into snRNPs involve many steps in both the nucleus and the cytoplasm (Mattaj, 1986; Mattaj, 1988; Parry et al., 1989; Luhrmann et al., 1990; Neuman de Vegvar and Dahlberg, 1990). All spliceosomal snRNAs, with the exception of U6, are transcribed in the nucleus by Pol II (Parry et al., 1989). Initially, the transcripts have a 7-methylguanosine (m^7G) cap at the 5' end, like other Pol II transcripts, and 1 to 10 extra nucleotides at the 3' end (Neuman de Vegvar and Dahlberg, 1990). Following transcription, U snRNAs are immediately transported from the nucleus to the cytoplasm (Mattaj, 1988), which requires the nuclear cap binding complex CBC (Lewis and Izaurralde, 1997). This export is mediated by the receptor CRM1 (also called exportin 1) (Fornerod et al., 1997), and CBC is presumably bound either directly or indirectly to CRM1. CBC directly interacts with importin α via the CBP80 NLS and the interaction is destabilized by addition of importin β (Gorlich et al., 1996b), suggesting that importin α is also a component of the export complex and that binding of importin β to importin α in the cytoplasm may cause the release of U snRNA. In the cytoplasm, the Sm proteins associate

with U snRNAs by binding to the Sm site (Mattaj, 1988), a common structure of U1, U2, U4, and U5 snRNAs characterized by a single-stranded U-rich sequence flanked by two hairpins (Branlant et al., 1982). Then, the m^7G cap of the snRNAs is hypermethylated to form the 2,2,7-trimethylguanosine (m_3G) cap (Mattaj, 1986), and the extra nucleotides are trimmed from the 3' end (Neuman de Vegvar and Dahlberg, 1990). These cytoplasmic processing reactions are important for the subsequent nuclear import of the assembled snRNPs (Fischer and Luhrmann, 1990; Hamm et al., 1990; Neuman de Vegvar and Dahlberg, 1990; Fischer et al., 1993). snRNP nuclear import is mediated by importin β and snurportin 1, a functional analogue of importin α (Palacios et al., 1997; Huber et al., 1998). Snurportin 1 recognizes the m_3G cap structure of U snRNAs and binds to importin β through an importin-β-binding (IBB) domain (Huber et al., 1998). Just before and after the nuclear import of U snRNPs, some internal nucleotides are modified and many specific proteins associate with individual snRNP precursors to complete their biogenesis (Mattaj, 1988; Luhrmann et al., 1990; Neuman de Vegvar and Dahlberg, 1990; Zieve and Sauterer, 1990). U6 snRNA differs from the other spliceosomal snRNAs in several aspects. U6 RNA is transcribed by Pol III instead of Pol II (Parry et al., 1989) and contains a γ-monomethyl cap instead of the m_3G cap (Singh and Reddy, 1989). U6 RNA is not exported to the cytoplasm for processing (Vakan et al., 1990) and does not bind directly to Sm proteins due to its lack of an Sm site (Reddy and Busch, 1988). Thus, the formation of U4/U6 snRNP is probably a nuclear event mediated largely by basepairing between U4 and U6 snRNAs. Although the general processes of spliceosomal snRNP biogenesis have been elucidated, the detailed mechanisms of these processes and the factors involved are not clear. Recently, the spinal muscular atrophy gene product SMN and an interacting protein SIP1 have been implicated in snRNP assembly in the cytoplasm (Fischer et al., 1997; Liu et al., 1997) and snRNP recycling in the nucleus (Mattaj, 1998; Pellizzoni et al., 1998).

1.6.2.3 Spliceosome assembly

The spliceosome is formed by the ordered and stepwise interaction of the U1, U2,

U4/6, and U5 snRNPs and numerous other proteins with *cis*-acting elements of the pre-mRNA (Moore et al., 1993; Madhani and Guthrie, 1994; Kramer, 1996) (Fig. 1-1). Spliceosome assembly is initiated by the formation of the commitment or early (E) complex, in which the 5' and 3' splice sites are initially recognized by the U1 snRNP and the splicing factor U2AF (U2 snRNP auxiliary factor), respectively (Michaud and Reed, 1991; Jamison et al., 1992; Michaud and Reed, 1993). Binding of U1 snRNP to the pre-mRNA involves base pairing between U1 snRNA and conserved sequences at the 5' splice site (Zhuang et al., 1986). U2AF is a heterodimer comprising a large subunit, $U2AF^{65}$, and a small subunit, $U2AF^{35}$ (Zamore and Green, 1989; Zamore et al., 1992). $U2AF^{65}$ directly contacts the polypyrimidine tract (Ruskin et al., 1988; Zamore and Green, 1991) and $U2AF^{35}$ recognizes the AG dinucleotides at the 3' splice site (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999). Members of the SR family of splicing factors facilitate these interactions by binding to the pre-mRNA and engaging in protein-protein interactions between the U1 snRNP-specific 70K protein and $U2AF^{35}$ (Wu and Maniatis, 1993; Kohtz et al., 1994; Staknis and Reed, 1994). SR proteins are characterized by the presence of one or two N-terminal RNA recognition motifs (RRMs), which mediate RNA binding, and a C-terminal RS domain, which mediates protein-protein interactions (Fu, 1995). The following step, the formation of pre-splicing complex A, entails the binding of U2 snRNP to the pre-mRNA and involves base pairing between U2 snRNA and the BPS (Hodges and Beggs, 1994; Madhani and Guthrie, 1994; Reed, 1996). The base pairing event results in the formation of a short helix from which the branch point adenosine is bulged out (Parker et al., 1987; Wu and Manley, 1989; Zhuang and Weiner, 1989; Query et al., 1994). Therefore, this interaction defines the branchpoint for 5' splice site cleavage and lariat formation during the first catalytic step (Query et al., 1994). A-complex formation is ATP-dependent and requires multiple protein factors (Kramer, 1996; Will and Luhrmann, 1997). U2AF is essential for the association of U2 snRNP with the BPS (Ruskin et al., 1988). The GSG domain-containing protein SF1/BBP interacts with $U2AF^{65}$ and also has sequence specificity for the BPS (Kramer, 1992; Arning et al., 1996; Abovich and Rosbash,

1997; Berglund et al., 1997). This network of interactions may function in the initial recognition of the BPS. Two multisubunit splicing factors, SF3a and SF3b, convert the inactive 12S U2 snRNP into the active 17S U2 particle (Brosi et al., 1993a; Brosi et al., 1993b). Six of the seven SF3a and SF3b subunits interact directly with the pre-mRNA in the vicinity of the BPS (Champion-Arnaud and Reed, 1994; Gozani et al., 1996) and one SF3b subunit (SAP 155) has been shown to interact with U2AF (Gozani et al., 1998). These findings suggest that SF3a and SF3b may play a role in recruiting and tethering the U2 snRNP to the BPS of the pre-mRNA. Pre-splicing complex A is converted to splicing complex B by association with the U4/U6/U5 tri-snRNP complex (Pikielny et al., 1986; Cheng and Abelson, 1987; Konarska and Sharp, 1987). This step requires the tri-snRNP-specific proteins (Utans et al., 1992; Beggs et al., 1995) and SR proteins (Roscigno and Garcia-Blanco, 1995). The active spliceosome (complex C) is formed after a conformational rearrangement of complex B. Subsequent to tri-snRNP association, the U6 snRNA dissociates from the U4 snRNA and forms new duplexes with the U2 snRNA and the pre-mRNA (Madhani and Guthrie, 1994; Kramer, 1996). The U5 snRNA base pairs with a small number of exon nucleotides at both the 5' and 3' splice sites and the interactions are stabilized by U5 snRNP-associated proteins (Newman and Norman, 1991; Newman and Norman, 1992; Chiara et al., 1997). Therefore, U5 snRNP may play a role in aligning the exons for ligation in the second catalytic step (Sontheimer and Steitz, 1993; Newman et al., 1995; Teigelkamp et al., 1995; O'Keefe et al., 1996). The dynamic changes in RNA base-pairing interactions during spliceosome assembly are catalyzed by proteins belonging to the DEAD-box superfamily of RNA helicases (Hamm and Lamond, 1998; Staley and Guthrie, 1998). DEAD-box proteins are characterized by seven conserved domains, including the tetrapeptide DEAD (consensus sequence D-E-A/C/I-D/H) (Gorbalenya and Koonin, 1993). Some of them have been shown to exhibit RNA-stimulated ATPase activity (Will and Luhrmann, 1997) and to mediate unwinding of RNA duplexes (Laggerbauer et al., 1998; Raghunathan and Guthrie, 1998; Schwer and Gross, 1998; Wagner et al., 1998; Wang et al., 1998).

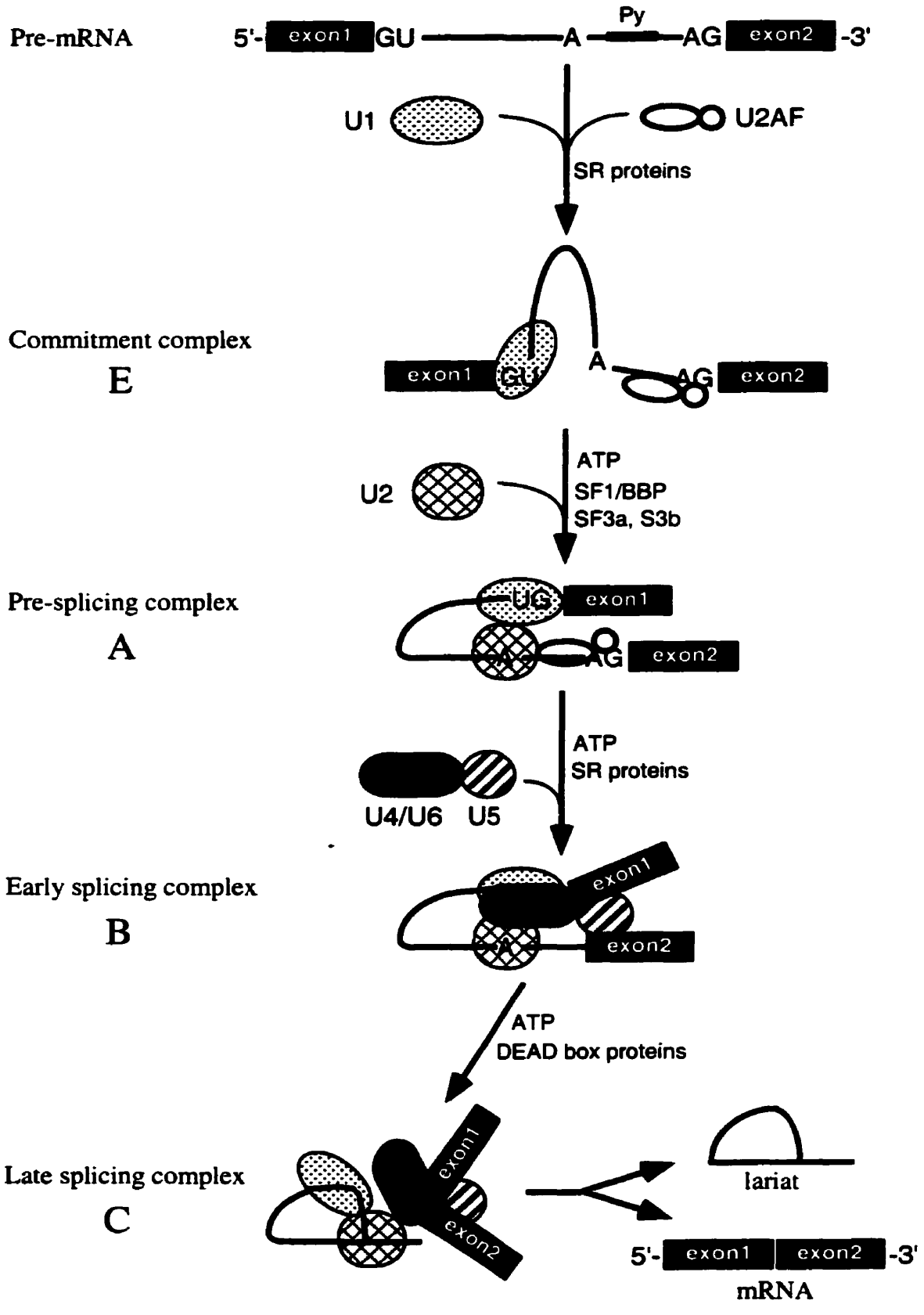


FIG. 1-1. An overview of the spliceosome assembly pathway. U snRNPs are represented by filled symbols and U2AF by open symbols. Important factors required for each step are indicated.

1.6.3 3' Polyadenylation

The 3' end of mRNAs comprises a polyadenosine sequence of 200-250 nucleotides in length, the poly(A) tail, which plays a variety of roles in mRNA metabolism, including transcriptional termination, RNA export, RNA stability and translational regulation (Whitelaw and Proudfoot, 1986; Logan et al., 1987; Connelly and Maley, 1988; Proudfoot, 1989; Jackson and Standart, 1990; Sachs, 1993). The formation of the 3' end is a two-step process in which an endonucleolytic cleavage is followed by addition of poly(A) stretch to the newly-formed 3' end (Wahle, 1995; Zhao et al., 1999).

The sequence elements in the pre-mRNA that determine the precise site of 3'-end cleavage and polyadenylation have been well studied (Wahle and Keller, 1992; Edwalds-Gilbert et al., 1997). The hexanucleotide AAUAAA, located 10-30 nucleotides upstream of the cleavage site, is the central signal for polyadenylation. In mammalian mRNAs, the cleavage site is often found immediately following a CA dinucleotide. Another element is a GU- or U-rich region, usually 10-30 bases downstream of the cleavage site. In addition, some pre-mRNAs (especially viral transcripts) contain auxiliary U-rich sequences, usually located upstream of AAUAAA.

A number of proteins are involved in the 3'-end processing of mRNAs (Keller, 1995; Manley, 1995; Wahle, 1995; Proudfoot, 1996). The two key components responsible for the recognition of the RNA substrate are the cleavage and polyadenylation specificity factor (CPSF) and the cleavage stimulation factor (CstF). CPSF is a large protein complex that recognizes the poly(A) signal AAUAAA (Keller et al., 1991) and is required for both cleavage and polyadenylation (Gilmartin and Nevins, 1989; Bienroth et al., 1991; Murthy and Manley, 1992). CstF is a heterotrimer, with subunits of 77, 64, and 50 kDa (Takagaki et al., 1990; Gilmartin and Nevins, 1991), that is required only for the cleavage event (Gilmartin and Nevins, 1989; Takagaki et al., 1989). It binds to the GU- or U-rich element and contacts the RNA with its 64 kDa subunit via a classical RNP-type RNA-binding domain (Wilusz and Shenk, 1988; Takagaki et al., 1991; Murthy and Manley, 1992; MacDonald et al., 1994). CstF and CPSF mutually stabilize their binding to RNA

(Gilmartin and Nevins, 1989; Wilusz et al., 1990; Gilmartin and Nevins, 1991; Murthy and Manley, 1992). Thus, CstF and CPSF may interact across the cleavage site and direct the endonuclease to its site of action. The 3'-processing complexes that form around the poly(A) sites also contain two cleavage factors (CF Im and CF IIm) that carry out the endonucleolytic cleavage step (Manley, 1995). Poly(A) polymerase (PAP) is also required for efficient cleavage of most mRNAs (Ryner et al., 1989), although its main function is the addition of adenylate residues onto the mRNA 3' end following cleavage. Poly(A) extension is strongly stimulated by an additional component, poly(A)-binding protein II (PAB II). PAB II is a protein of 50 kDa that binds to the growing poly(A) tail in the polyadenylation complex and functions as an elongation factor that allows the efficient synthesis of the full-length poly(A) tail (Keller, 1995; Manley, 1995; Wahle, 1995). In the cytoplasm, the poly(A) tail is bound by a different protein, poly(A)-binding protein I (PAB I), which recruits the 40S ribosomal subunit to the mRNA and promotes translation initiation (Proudfoot, 1996).

1.7 NUCLEOCYTOPLASMIC TRANSPORT

In eukaryotic cells, DNA replication and RNA synthesis take place in the nucleus and protein synthesis occurs in the cytoplasm. This means that proteins required for nuclear functions must be imported into the nucleus, and RNA molecules must be exported to the cytoplasm. The nucleocytoplasmic trafficking of these macromolecules occur through the nuclear pore complex (NPC), a structure of ~125,000 kDa embedded in the nuclear envelope, containing 50-100 distinct polypeptides collectively termed nucleoporins (Doye and Hurt, 1997; Ohno et al., 1998). NPCs provide aqueous channels of 9 nm in diameter, which allow the diffusion of ions, metabolites, and small proteins (less than 40-60 kDa), and mediate the selective transport of particles greater than 25 nm in diameter in an energy-dependent and signal-mediated manner (Davis, 1995; Pante and Aebi, 1995).

Nuclear import and export are tightly coupled processes, and each can be divided into several distinct stages. First, the macromolecule bearing a specific localization signal is

recognized by a soluble receptor protein, and the assembled receptor-macromolecule complex is then targeted to the NPC. Next, the receptor-macromolecule complex is translocated through NPC from one side of the nuclear envelope to the other, which usually requires the GTPase Ran. Lastly, the complex is dissembled, and the receptor proteins are recycled (Corbett and Silver, 1997; Nigg, 1997; Mattaj and Englmeier, 1998).

1.7.1 Nuclear Import

Most nuclear proteins contain NLSs that mediate their active import into the nucleus. The most extensively studied NLS is the classical NLS which typically consists of one or two clusters of basic amino-acid residues (Dingwall and Laskey, 1991). The single cluster in the SV40 large T antigen (PKKKRKV) (Kalderon et al., 1984) and the two in nucleoplasmin (KRPAAIKKAGQAKKKK) (Robbins et al., 1991) are the prototypic examples of a monopartite and a bipartite NLS, respectively. Another well-characterized NLS is the M9 sequence originally found in hnRNP A1 (Siomi and Dreyfuss, 1995; Weighardt et al., 1995), a protein that shuttles rapidly between the nucleus and the cytoplasm (Pinol-Roma and Dreyfuss, 1992). M9, a 38 amino-acid motif rich in glycine and aromatic residues, is sufficient not only for targeting hnRNP A1 into the nucleus but also for its subsequent export out of the nucleus (Michael et al., 1995a). A number of GSG proteins, including Sam68, SLM-1, SLM-2, Qk1-5, and How(L), localizes predominantly in the nucleus and their C termini determine nuclear localization (Hardy et al., 1996; Ishidate et al., 1997; Di Fruscio et al., 1998; Chen et al., 1999; Di Fruscio et al., 1999). Sequence alignment revealed that these proteins all contain an R-X-H-P-Y-X-R motif at their C termini (Di Fruscio et al., 1999; Nabel-Rosen et al., 1999; Wu et al., 1999). Deletions or amino acid substitutions of the motif alter protein localization (Chen et al., 1999; Nabel-Rosen et al., 1999; Wu et al., 1999), suggesting that this motif represents a novel NLS. The import pathways for these GSG proteins are still unknown.

Proteins bearing a classical NLS are recognized and delivered to the NPC by a soluble, heterodimeric complex comprised of importin α (also called karyopherin α) (Gorlich et al.,

1994; Morioianu et al., 1995; Weis et al., 1995) and importin β (karyopherin β) (Chi et al., 1995; Gorlich et al., 1995; Imamoto et al., 1995; Radu et al., 1995). Importin α (60 kDa) recognizes the NLS via a large NLS-binding domain (Conti et al., 1998) and interacts with importin β through a small N-terminal IBB domain (Gorlich et al., 1996a; Weis et al., 1996; Cingolani et al., 1999). Importin β (97 kDa) has Ran-, IBB- and NPC-binding activities and is responsible for docking to the NPC and translocation through the nuclear pore (Chi et al., 1997; Chi and Adam, 1997; Kutay et al., 1997a; Cingolani et al., 1999). Thus, in this nuclear import pathway, importin β is the genuine import mediator or receptor whereas importin α acts as an adaptor that joins importin β and the NLS substrate. Importin β is also involved in the nuclear import of spliceosomal snRNPs, utilizing snurportin-1 as an adaptor that recognizes the m_3G cap structure of U snRNAs (Palacios et al., 1997; Huber et al., 1998). Other transport pathways have also been identified, all involving transport factors with limited sequence identity to importin β. These importin β-like receptors bind their substrates directly rather than via an importin α-like adaptor (Mattaj and Englmeier, 1998; Pemberton et al., 1998). For example, transportin 1 (also known as karyopherin β2) mediates the nuclear import of mRNA-binding proteins, including hnRNP A1 (Nakielny et al., 1996; Pollard et al., 1996; Bonifaci et al., 1997; Fridell et al., 1997) and hnRNP F (Siomi et al., 1997), by directly binding to the M9 NLS (Pollard et al., 1996; Fridell et al., 1997).

Although very little is known about the movement of import substrates through the nuclear pore, a number of players in the translocation step of nuclear import have been identified (Mattaj and Englmeier, 1998; Ohno et al., 1998). The most important breakthrough came with the identification of the small GTPase Ran, a member of the Ras superfamily (Drivas, 1990), as a critical transport factor (Melchior et al., 1993; Moore and Blobel, 1993). This discovery led to the initial hypothesis that the hydrolysis of Ran-bound GTP is the key energy source driving translocation through NPCs (Koepp and Silver, 1996; Gorlich, 1997). However, recent evidence from both in vivo and in vitro studies indicates that GTP hydrolysis is not needed for translocation *per se* (Cole and Hammell, 1998;

Dahlberg and Lund, 1998). The current model suggests that the major role of Ran is to control the directionality of nuclear transport (Izaurralde et al., 1997a; Cole and Hammell, 1998).

Ran is localized primarily within the nucleus, but it is also present in the cytoplasm, suggesting that it may shuttle cross the nuclear envelope. Like other G-proteins, Ran cycles between a GDP-bound form and a GTP-bound form (Bourne et al., 1991). Ran GTPase-activating protein (RanGAP1) catalyses GTP hydrolysis to form Ran-GDP, and a nucleotide exchange factor, regulator of chromosome condensation 1 (RCC1), catalyzes the exchange of GTP for bound GDP to regenerate Ran-GTP (Bischoff and Pontingl, 1995). The Ran regulators are asymmetrically localized on opposite sides of the nuclear envelope. RanGAP1 is located largely in the cytoplasm (Hopper et al., 1990), and it is targeted to the NPC when modified by covalent attachment of SUMO-1, a ubiquitin-like molecule (Matunis et al., 1996; Mahajan et al., 1997; Matunis et al., 1998). RCC1 is exclusively nuclear and chromatin-bound (Ohtsubo et al., 1989). This asymmetric distribution of RanGAP1 and RCC1 predicts that Ran exists primarily in the GTP-bound state in the nucleus and GDP-bound state in the cytoplasm. The steep Ran-GTP/Ran-GDP gradient across the nuclear envelope may play a key role in driving the directional movement of macromolecules through the nuclear pore. In the cytoplasm, Ran-GDP supports the formation of substrate-receptor complexes (Rexach and Blobel, 1995). Once transported into the nucleus, binding of Ran-GTP to import receptors like importin β or transportin causes dissociation of substrate and receptor from each other and from the NPC (Mattaj and Englmeier, 1998; Chook and Blobel, 1999).

1.7.2 Nuclear Export

Macromolecules exported from the nucleus include proteins and different types of RNAs. RNA export occurs in the form of RNPs. In many aspects, nuclear export mirrors nuclear import. Recent studies have demonstrated that each major class of RNAs uses distinct, saturable export pathways, although many of these pathways are still poorly

understood (Izaurralde and Adam, 1998; Stutz and Rosbash, 1998).

Like protein import, nuclear protein export is mediated by specific protein sequences, termed NESs. A general NES has been found in a variety of proteins, including the human immunodeficiency virus type 1 (HIV-1) Rev protein (Fischer et al., 1995), the human T-cell leukemia virus type 1 (HTLV-1) Rex protein (Bogerd et al., 1996), protein kinase A inhibitor (PKI) (Wen et al., 1995), amphibian transcription factor IIIA (Fridell et al., 1996), and the RNA export mediator Gle1p/Rss1p (DelPriore et al., 1996; Murphy and Wente, 1996). It is an ~10-amino-acid sequence that is rich in hydrophobic residues, particularly leucine (Moore, 1996). The M9 domain of hnRNP A1 serves as a combined signal for both import and export (Michael et al., 1995a; Siomi and Dreyfuss, 1995). A similar bi-directional transport domain of ~40 amino acids, termed hnRNP K nuclear shuttling (KNS) domain, is identified from another shuttling protein hnRNP K (Michael et al., 1997). M9 and KNS share no sequence homology with the leucine-rich NES.

RNAs are exported in the form of RNPs, and the export signals are usually provided by RNA-binding proteins that contain NESs (Izaurralde et al., 1995a). Thus, these proteins act as adaptors bridging the RNA substrates and the export receptors. Structural features of RNAs that are recognized by these proteins during export can be viewed as export signals on RNAs. Well-characterized examples include the Rev responsive element (RRE) of HIV-1, a 234 nucleotide region within the HIV-1 *env* gene that is predicted to form a stem-loop structure (Malim et al., 1989b), the constitutive transport element (CTE) of simple retroviruses like Mason-Pfizer monkey virus (MPMV), a 154 nucleotide sequence in MPMV that forms a long, imperfectly paired stem (Bray et al., 1994), and the 5' m^7G cap structure of U snRNAs (Lewis and Izaurralde, 1997). For mRNAs, multiple features including the cap structure, the poly(A) tail, and the "body" seem to contribute to efficient export (Jarmolowski et al., 1994).

Several export receptors, or exportins, have been identified, and they are all members of the importin β family (Ullmann et al., 1997). With the exception of exportin-t, which binds directly to tRNA and mediates its export (Arts et al., 1998; Kutay et al., 1998), all exportins

bind to proteins through recognition of specific NESs. CRM1 (exportin 1) mediates the export of proteins with leucine-rich NESs (Fornerod et al., 1997; Ossareh-Nazari et al., 1997; Stade et al., 1997). CRM1 is a target of leptomycin B (LMB) (Nishi et al., 1994), which inhibits the export of proteins with leucine-rich NESs in many systems (Fornerod et al., 1997; Fukuda et al., 1997; Wolff et al., 1997). CRM1 has been shown to mediate the nuclear export of U snRNAs (Fornerod et al., 1997), probably by interacting with an NES-containing adaptor associated with CBC or other factors that participate in the export of U snRNAs. It is not yet clear whether CRM1 has a direct role in mRNA export. In higher eukaryotes, LMB has no effect on mRNA export, suggesting that mRNA export is independent of CRM1 (Fornerod et al., 1997). The GSG protein Sam68 can replace Rev in the export of HIV-1 RNA (Reddy et al., 1999). But unlike Rev, the effect of Sam68 is not inhibited by LMB (Reddy et al., 1999), suggesting that Sam68 may use a CRM1-independent pathway. Another exportin, CAS, has been reported to mediate the re-export of importin α into the cytoplasm (Kutay et al., 1997b). Although the region of importin α that interacts with CAS is not defined, previous work showed that the N-terminal IBB domain is sufficient to direct nuclear import but not export (Gorlich et al., 1996a; Weis et al., 1996), indicating that CAS binding requires other regions of importin α. Recently, the receptor MsN5 has been shown to mediate the export of the phosphorylated transcription factor Pho4 (Kaffman et al., 1998). An important property common to the exportins is the requirement for Ran-GTP as a component of the stable export complexes. In contrast to import complexes, export complexes are promoted by Ran-GTP and destabilized by Ran-GDP (Dahlberg and Lund, 1998).

1.7.3 Rev-mediated export of viral RNAs

Important insight into export mechanisms has come from studies of the HIV-1 Rev protein (Cullen, 1998). HIV-1 replication requires the transport of unspliced, incompletely spliced and fully spliced RNA transcripts from the nucleus to the cytoplasm. By acting as an export adaptor, the Rev protein plays an essential role in the nuclear export of unspliced

and incompletely spliced viral RNAs (Cullen, 1998). Rev contains an arginine-rich NLS (Malim et al., 1989a) and a leucine-rich NES (Fischer et al., 1995) and has been shown to shuttle between the nucleus and the cytoplasm (Myer and Malim, 1994). The arginine-rich sequence also serves as an RNA-binding motif, which mediates the direct interaction of Rev with its target RNA sequence RRE (Malim et al., 1989b; Zapp and Green, 1989; Malim et al., 1990). The Rev-mediated RNA export proceeds as follows. First, multiple Rev molecules assemble onto the RRE RNA and recruit the export receptor CRM1 in a Ran-GTP-dependent manner. Second, CRM1 targets the resultant RNP complex to the nuclear pore via a direct interaction with nucleoporins. Third, the complex is translocated through the nuclear pore, probably involving the sequential interaction of CRM1 with several specific nucleoporins. Fourth, once the complex reaches the cytoplasm, RanGAP1 and RanBP1 induce the hydrolysis of Ran-GTP to Ran-GDP, thereby inducing the release of HIV-1 RNA cargo (Cullen, 1998). Rev is recycled back to the nucleus by directly interacting with importin β via the arginine-rich NLS (Truant and Cullen, 1999). Cellular proteins that bind Rev (Ruhl et al., 1993; Luo et al., 1994; Tange et al., 1996) and/or RRE (Xu et al., 1996; Powell et al., 1997) have been identified, which modulate Rev activity positively or negatively. Sam68 has been identified as a functional homolog of Rev, suggesting that Sam68 may play a role in RNA export (Reddy et al., 1999).

1.7.4 A role for hnRNP proteins in mRNA export

Nascent pre-mRNA transcripts associate in vivo with a set of more than 20 polypeptides, the hnRNP proteins, with sizes ranging from 34 to 120 kDa (Dreyfuss et al., 1993; Pinol-Roma and Dreyfuss, 1993). hnRNP proteins represent general RNA-binding proteins and bind readily to single-stranded nucleic acids. Individual hnRNP proteins exhibit distinct binding preferences for homopolymeric RNAs (Swanson and Dreyfuss, 1988), and the binding of at least some hnRNP proteins is transcript dependent (Bennett et al., 1992). Because hnRNP proteins are located in the nucleus at steady state, they were not initially considered as potential export mediators. The demonstration that some hnRNP proteins shuttle continuously (Pinol-Roma and Dreyfuss, 1992) between the nucleus and

the cytoplasm changed this perception. The best-studied member of this protein family to date is human hnRNP A1. hnRNP A1 is composed of two RNP motif RNA-binding domains (RBDs) as well as a third RNA-binding domain, the RGG box (Burd and Dreyfuss, 1994). The M9 motif of hnRNP A1 serves as both an NLS (Siomi and Dreyfuss, 1995; Weighardt et al., 1995) and an NES (Michael et al., 1995a). Both the number of hnRNP A1 molecules in HeLa cells (10^8 molecules per nucleus) and the amount of A1 shuttling (10^5 molecules per minute) are enormous (Michael et al., 1995b). Immunoelectron microscopy shows that mRNAs in transit through the NPC to the cytoplasm are associated with hnRNP A1 (Visa et al., 1996). Microinjection of saturating amounts of hnRNP A1 into the nucleus of *Xenopus* oocytes competitively inhibits export of some mRNAs while a deletion mutant which lacks a functional NES has no effect (Izaurralde et al., 1997b), indicating that the hnRNP A1 NES plays a central role in the export of mRNP particles. The yeast hnRNP A1 homolog, Npl3p, also shuttles between the nucleus and the cytoplasm (Flach et al., 1994; Lee et al., 1996b), and its mutation results in mRNA accumulation in the nucleus (Russell and Tollervey, 1995; Singleton et al., 1995). These findings provide strong evidence that hnRNP A1 is involved in mRNA export. Other shuttling hnRNP proteins, including A2, D, E, I and K, are also candidate mRNA export mediators (Pinol-Roma and Dreyfuss, 1992; Pinol-Roma and Dreyfuss, 1993; Michael et al., 1995b). The export receptor that recognizes hnRNP A1 or other shuttling hnRNP proteins has not been identified. Although transportin mediates the nuclear import of hnRNP A1 by binding to the M9 domain (Nakielny et al., 1996; Pollard et al., 1996; Bonifaci et al., 1997; Fridell et al., 1997), it does not appear to mediate the nuclear export of hnRNP A1 (Izaurralde et al., 1997b).

1.8 TRANSLATIONAL REGULATION

Protein production is generally controlled by regulating the rate of translation initiation. During translation initiation, ribosomes are recruited to mRNAs in a sequential, multistep process. In eukaryotes, following the recruitment of the small ribosomal subunit to the

mRNA, the mRNA sequence is scanned and the small subunit is placed at the initiation codon. Then, the joining of the large ribosomal subunit to the mRNA completes the assembly of the ribosome (Hershey et al., 1996).

The rate of translation initiation is determined by *cis*-acting elements in the mRNA molecule and *trans*-acting factors which interact with these elements. The 5' cap structure recruits the 40S small ribosomal subunit to the mRNA during translation initiation. This activity is mediated by a well-characterized translation initiation factor eIF4F (Sonenberg, 1996). eIF4F in all eukaryotic cells consists of two core subunits. These are the cap binding protein eIF4E and the large subunit eIF4G. eIF4G bridges the 5' cap structure and the small ribosomal subunit by interacting with eIF4E and the 40S subunit-associated factor eIF3 (Sachs et al., 1997). Reversible phosphorylation of eIF4E is one of the most frequently used mechanisms for translational control. Phosphorylation of eIF4E at Ser-209 is increased in response to many extracellular stimuli required for cell growth and translation. Phosphorylated eIF4E has a higher binding affinity to the mRNA 5' cap relative to its unphosphorylated counterpart (Sonenberg and Gingras, 1998). The activity of eIF4E is negatively regulated by eIF4E-binding proteins (4E-BPs). 4E-BPs share an eIF4E-binding motif with eIF4G (Mader et al., 1995) and compete with eIF4G for binding to eIF4E (Haghighat et al., 1995). 4E-BPs do not interfere with the binding of eIF4E to the 5' cap structure, but rather prevent the assembly of eIF4F. They inhibit cap-dependent, but not cap-independent translation (Sonenberg and Gingras, 1998).

The polyadenylated 3' end of mRNAs not only protects mRNA from degrading but also regulates the initiation of translation. The function of the poly(A) tail in translation is dependent on the cytoplasmic poly(A)-binding protein, PAB I, which interacts with eIF4G/eIF4E complex, leading to the recruitment of the 40S small ribosomal subunit (Sachs et al., 1997).

Translation of picornavirus mRNAs is cap-independent and mediated by sequence elements in the viral 5' UTR known as internal ribosome entry sites (IRESs) (Jackson and Kaminski, 1995). IRESs have also been identified in some nonpicornaviral and cellular

mRNAs (Jackson and Kaminski, 1995; Johannes and Sarnow, 1998). Except for the presence of a pyrimidine-rich sequence element, IRESs have moderate conservation of primary sequence and more significant conservation of predicted secondary structure. It is not fully understood how IRES elements recruit ribosomal subunits. Both RNA-RNA and protein-RNA complexes may be involved in the recruitment process. Nearly the same set of canonical factors that are important in cap-stimulated translation initiation have been found to be important in IRES-mediated internal initiation. eIF4F binds directly to the EMC virus IRES via eIF4G, which contains a putative RNA-binding site and could mediate the recruitment of the 40S ribosomal subunit to the IRES element. Noncanonical translation initiation factors, such as the La autoantigen and PTB, are also implicated in IRES-mediated translation, but their precise role in translation is not clear (Sachs et al., 1997).

Other elements in the 5'UTR as well as the 3'UTR of mRNAs also play important roles in translational regulation. Regulation of ferritin mRNA translation by an iron response element (IRE)-binding protein is a well-characterized example involving a regulatory sequence in the 5'UTR. In this case translation is repressed because the bound protein interferes with movement of the preinitiation complex from the 5' cap to the start codon (Klausner et al., 1993). Translational regulation mediated by the 3'UTR have been described in many developmental events. For example, elements in the 3'UTR of the *Drosophila hunchback*, *oskar*, and *nanos* mRNAs are necessary for repressing translation and hence controlling anterior-posterior patterning (Wharton and Struhl, 1991; Kim-Ha et al., 1995; Rongo et al., 1995; Dahanukar and Wharton, 1996; Gavis et al., 1996; Smibert et al., 1996).

REFERENCES: See page 191.

Chapter 2

SELF-ASSOCIATION OF THE SINGLE-KH-DOMAIN FAMILY MEMBERS SAM68, GRP33, GLD-1, AND QK1: ROLE OF THE KH DOMAIN

PREFACE

KH domain proteins usually contain two or more KH domains. Previous studies have shown that multiple copies of KH domains are required for optimal RNA binding. This raises the question as to how GSG proteins, which contain a single KH domain, interact with RNA. One possibility is that these proteins self-associate into multimers. To test this possibility, we investigated whether the GSG protein Sam68 and other family members oligomerize. Because the boundaries of the GSG domain were not clear at that time, the embedded KH domain was highlighted in this study,

SUMMARY

Sam68 is a member of a growing family of proteins that contain a single KH domain embedded in a larger conserved domain of ~170 amino acids. Loops 1 and 4 of this KH domain family are longer than the corresponding loops in other KH domains and contain conserved residues. The KH domains are protein motifs that are involved in RNA binding and are often present in multiple copies. Here we demonstrate by coimmunoprecipitation studies that Sam68 self-associated and that cellular RNA was required for the association. Deletion studies demonstrated that the Sam68 KH domain loops 1 and 4 were required for self-association. The Sam68 interaction was also observed in *Saccharomyces cerevisiae* by the two-hybrid system. In situ chemical cross-linking studies in mammalian cells demonstrated that Sam68 oligomerized in vivo. These Sam68 complexes bound homopolymeric RNA and the SH3 domains of $p59^{fyn}$ and phospholipase Cγ1 in vitro, demonstrating that Sam68 associates with RNA and signaling molecules as a multimer. The formation of Sam68 complex was inhibited by $p59^{fyn}$, suggesting that tyrosine

phosphorylation regulates Sam68 oligomerization. Other Sam68 family members including *Artemia salina* GRP33, *Caenorhabditis elegans* GLD-1, and mouse Qk1 also oligomerized. In addition, Sam68, GRP33, GLD-1, and Qk1 associated with other KH domain proteins such as Bicaudal C. These observations indicate that the single KH domain found in the Sam68 family, in addition to mediating protein-RNA interactions, mediates protein-protein interactions.

INTRODUCTION

The K homology (KH) domain is a small protein module consisting of 70 to 100 amino acids that was originally identified as a repeated sequence in the heterogeneous nuclear ribonucleoprotein particle (hnRNP) K (Siomi et al., 1993). There are over 30 proteins that have been identified in prokaryotes and eukaryotes that contain single or multiple copies of this domain (Gibson et al., 1993). The KH domain is an RNA binding motif that is thought to make direct protein-RNA contacts with a three-dimensional βαααββα fold (Musco et al., 1996). The RNA binding properties of the KH domain proteins were initially demonstrated for FMR1, the gene product of the human fragile X syndrome, and hnRNP K (Ashley et al., 1993; Siomi et al., 1993). It has been demonstrated that multiple KH domains are required for optimal RNA binding (Siomi et al., 1994). In addition, there is evidence that KH domain proteins may function as multimers. For example, FMR1 associates with FXR family members through coiled-coil motifs and may function in translation, mRNA stability, or both (Zhang et al., 1995; Siomi et al., 1996).

Mutations have been isolated in a number of genes encoding KH domain proteins in various species, making the KH domain one of the protein modules with genetic data supporting its essential physiological role. In humans, gene lesions that prevent expression of the KH protein FMR1 result in the fragile X mental retardation syndrome (Pieretti et al., 1991; Verkerk et al., 1991), the most common form of heritable mental retardation (Nussbaum and Ledbetter, 1995). The significance of the KH domain was underscored by a point mutation altering a conserved isoleucine 304 to asparagine in the second KH domain of FMR1 (DeBoulle et al., 1993). The point mutation alters the structure of the KH domain (Musco et al., 1996) and severely impairs RNA binding activity (Siomi et al., 1994). In *Caenorhabditis elegans*, GLD-1 is a cytoplasmic protein required for germ cell differentiation (Francis et al., 1995a; Francis et al., 1995b; Jones et al., 1996). Thirty-two mutant alleles of GLD-1 can be divided into six phenotypic classes (Jones and Schedl, 1995). Alteration of GLD-1 glycine 227 results in a recessive tumorous germ line phenotype. The structure of the vigilin KH domain predicts that this conserved glycine

forms part of the RNA binding surface (Musco et al., 1996). Missense mutations were identified for GLD-1 glycine 248 or 250, and this class results in masculinization of the germ line (Jones and Schedl, 1995). The two glycines are located at the beginning of loop 4 and are not predicted to be involved in RNA binding (Musco et al., 1996). In mice, the *quaking* viable mutation severely impairs myelination, and as a result, mice develop a rapid tremor at postnatal day 10 (Sidman et al., 1964). A missense mutation in the *quaking* gene product, Qk1, is embryonic lethal (Ebersole et al., 1996), and alteration of Qk1 expression leads to the *quaking*viable phenotype (Hardy et al., 1996). In *Drosophila melanogaster*, Bicaudal C (BicC) contains five KH domains (Mahone et al., 1995). Gene lesions that truncate the BicC protein or a point mutation that replaces glycine 295 with an arginine in the third KH domain is a strong allele, leading to defects in oogenesis and anterior-posterior embryonic patterning (Mahone et al., 1995).

Alignment of the KH domains (Gibson et al., 1993; Musco et al., 1996) reveals a subfamily of KH-domain-containing proteins including human and murine Sam68 (Wong et al., 1992; Richard et al., 1995), *Artemia salina* GRP33 (Cruz-Alvarez and Pellicer, 1987), *C. elegans* GLD-1 (Jones and Schedl, 1995), mouse Qk1 (Ebersole et al., 1996), human ZFM1 (Toda et al., 1994), and predicted protein BO280.11b from *C. elegans* (Wilson et al., 1994). Most recently, two additional members have been identified, including the mammalian splicing factor SF1 (Arning et al., 1996) and the *Drosophila* 'held-out wings' gene product (Genbank accession no. U72331). SF1 and ZFM1 are different spliced variants derived from the same gene (Arning et al., 1996). All these proteins, except BO280.11b, contain a single KH domain embedded within a larger protein domain of ~170 residues (Jones and Schedl, 1995; Arning et al., 1996). This single extended KH domain is approximately 26 amino acids longer than most other KH domains. The extra conserved amino acids are specifically localized in loops 1 and 4; loop 1 contains an extra 6 amino acids, and loop 4 contains an additional 20 (Musco et al., 1996).

The Src-associated-in-mitosis protein of 68 kDa, Sam68, was previously called p62 (Wong et al., 1992; Lock et al., 1996) and is the only known substrate for Src-family

tyrosine kinases during mitosis (Fumagalli et al., 1994; Taylor and Shalloway, 1994). Sam68 was identified as a binding protein for various SH3- and SH2-domain-containing signaling molecules [for a review see (Pawson, 1995)]. Sam68 binds Src-family tyrosine kinases, the adapter protein Grb2, and phospholipase Cγ (PLCγ) (Fumagalli et al., 1994; Maa et al., 1994; Taylor and Shalloway, 1994; Weng et al., 1994; Richard et al., 1995; Vogel and Fujita, 1995). Most recently, Sam68 has been shown to associate with Nck (Lawe et al., 1997), the poliovirus RNA-dependent RNA polymerase 3D, (McBride et al., 1996), Itk/Tsk (Bunnell et al., 1996), Grap (Trub et al., 1997), Cbl and Jak3 in Hayai cells (Fusaki et al., 1997). These interactions support the potential role for Sam68 as an adaptor molecule for tyrosine kinases (Richard et al., 1995; Taylor et al., 1995). In addition to its SH3- and SH2-domain-binding property, Sam68 is an RNA-binding protein. It has been shown to bind single-stranded RNA and single-stranded and double-stranded DNA as well as homopolymeric RNA in vitro (Wong et al., 1992; Taylor and Shalloway, 1994). The RNA binding activity of Sam68 is regulated by tyrosine phosphorylation. Phosphorylation of Sam68 by the Src-family kinase p59^{fyn} abolishes homopolymeric RNA binding activity in vitro (Wang et al., 1995). Although the function of Sam68 is unknown, a splice variant of Sam68, devoid of the KH domain, has been identified and implicates the KH domain of Sam68 in cell cycle progression (Barlat et al., 1997).

The presence of multiple KH domains in proteins prompted us to investigate whether the single-KH-domain family member Sam68 could oligomerize. We present evidence that Sam68 can bind signaling molecules and RNA as a multimer and that Sam68 oligomerization is regulated by the tyrosine kinase p59^{fyn}.

MATERIALS AND METHODS

DNA constructions. The plasmid encoding His-Sam68 was generated as follows. The *Eco*RI fragment of myc-Sam68 [previously called myc-p62 (Richard et al., 1995); missing the N-terminal 67 amino acids] was subcloned in the *Eco*RI site of His-Bluescript KS$^+$). His-Bluescript was constructed by annealing two oligonucleotides (5'-CGC GGA TCC ACC ATG GGC AGC AGC CAT CAT CAT-3' and 5'-GCG CTC GAG GGA ATT CCC GCT GCT GTG ATG ATG ATG ATG ATG-3') and filling in the ends with DNA polymerase I (Klenow fragment). This DNA fragment was subcloned into the *Bam*HI and *Xho*I sites of Bluescript KS$^+$ (Stratagene). HA-Sam68 was constructed by subcloning the *Eco*RI fragment of myc-Sam68 in HA-Bluescript KS$^+$. HA-Bluescript KS$^+$ was constructed by annealing two oligonucleotides (5'-TA CCC ATG GCG TAC CCC TAC GAC GTG CCC GAC-3' and 5'-CTG GAATTC CAG CTG GCG TAG TCG GGC ACG TC-3') and filling in the ends with DNA polymerase I (Klenow fragment). This DNA fragment was subcloned into the *Nco*I and *Eco*RI sites of myc-Bluescript KS$^+$ (Richard et al., 1995). Underlined nucleotides denote restriction endonuclease sites.

Sam68Δ294-405 and Sam68Δ280-339 were constructed by inverse PCR with myc-Sam68 as DNA template. The sequences of the oligonucleotides used are 5'-TCT TGG CAG CTC CTC GTC CTC TCA C-3' and 5'-TCT TCC AAG ATT CTT ACG AAG CCT ACG-3' (Sam68Δ294-405) and 5'-TCT TGG GTT CAG GTA CTC CGT TCA A-3' and 5'-TCT CTG GAC GTG GTG TTG GAC CAC C-3' (Sam68Δ280-339). Sam68:294 and Sam68:330 were constructed by using the T7 promoter primer as forward primer and 5'-TAG AAT TCA GGC AGC TCC TCG TCC TCT CAC-3' and 5'-AGG AAT TCA TGG CAC CCC TCG AGT CAC A-3' as reverse primers, respectively. The DNA template was myc-Sam68, and the amplified fragments were digested with *Eco*RI and subcloned in myc-Bluescript KS$^+$ (Richard et al., 1995). The DNA fragment for Sam68ΔN was amplified by PCR with 5'-GAG AAT TCG TAC CCG CCT GAA CTC A-3' and the universal reverse primer as oligonucleotides and myc-Sam68 as a DNA template. The fragment was digested with *Eco*RI and subcloned in myc-Bluescript KS$^+$. This construct deletes the N-terminal

102 amino acids of the murine Sam68 protein. Sam68:103-269 was constructed by digesting Sam68ΔN with *Xba*I and subcloning it in the *Xba*I site of Bluescript KS$^+$. The DNA fragment for Sam68:KH was amplified by PCR with myc-Sam68 as a DNA template and oligonucleotides 5'-TAA GAATTC GAA GCT GAA AGA ACG CGT G-3' and 5'-TCC CTC GAG ATA TCA TCC ATC ATA AC-3'. The DNA fragment was digested with *Eco*RI and *Xho*I and subcloned in myc-Bluecript KS$^+$.

The construct containing the entire coding region of Sam68 was generated as follows. A λ ZAP II mouse brain library (Stratagene) was screened with a ^{32}P-labeled random-primed DNA fragment encompassing nucleotides 527 to 940 of the mouse Sam68 cDNA (Richard et al., 1995). Clone B contained nucleotides 2 to1170 of the mouse Sam68 cDNA. The coding region of this clone was amplified with a forward primer that replaces the initiator methionine with an *Eco*RI site (5'-CCC GAA TTC GAG AAG AGA CGA TCC TGC CTC GCG CC-3') and the T7 promoter primer (located at the 3' end of the cDNA). The amplified fragment was subcloned in myc-Bluescript KS$^+$ with *Eco*RI, and this plasmid was called myc-Bluescript-B. myc-Sam68 was digested with *Sac*II, removing the sequences between the *Sac*II site in the polylinker and the *Sac*II site located in the Sam68 cDNA at nucleotide 378. The DNA fragment obtained from the digestion of myc-Bluescript-B with *Sac*II was inserted in between the *Sac*II sites of myc-Sam68. The new plasmid was called myc-Sam68f for full-length.

Sam68:I→N, Sam68:G→D, Sam68:E→G, Sam68ΔL1, Sam68ΔL4, Sam68:2G→R, and Sam68ΔKH were constructed by inverse PCR with myc-Sam68f as a DNA template. The oligonucleotides were 5'-TGT ATT TCC TTG TGG TCC AAG AAT-3' and 5'-AAC AAA AGA CTC CAG GAA GAG ACT G-3' (Sam68:I→N), 5'-AAG AAT CTT CCC CAC AAA ATT GAA C-3' and 5'-GAC CCA CAA GGA AAT ACA ATC A-3' (Sam68:G→D), 5'-TAC GGA CAG CAG CTG CAT GGC GTG AGT-3' and 5'-GGA ATT GAG AAG ATT CAG AAG GGA G-3' (Sam68:E→G), 5'-GAC AGG TAT CAG CAC GCG TTC-3' and 5'-GAC CTT GTG GGG AAG ATT CTT GGA-3' (Sam68ΔL1), 5'-CCT CAA GAC AGA GAT CTT TGC ACC AGT CTC-3' and 5'-CCT ATG GAT

CTG CAT GTC TTC ATT GAA GTC-3' (Sam68ΔL4), 5'-CCT AGA TCA ATG AGA GAC AAA GCC AAG G-3' and 5'-CCT CAA GAC AGA GAT CTT TGC ACC AGT CTC-3' (Sam68:2G→R), 5'-GAG ATG GAT CTG CAT GTC TTC ATT C-3' and 5'-GAG CAG TTC ATG TTC TTA TGA GA-3' (Sam68ΔKH). Underlined nucleotides denote changes introduced.

The plasmid constructs were verified by dideoxynucleotide sequencing with Sequenase (U.S. Biochemical).

Cloning of *A. salina* GRP33, *C.elegans* GLD-1, and mouse Qk1 cDNAs. Total RNA was extracted from *A. salina, C. elegans,* and mouse brain as previously described (Richard and Zingg, 1991). The oligonucleotides, sense and antisense, were 5'-TAC CTC GAG AAA TGG CTG CCA AAC CCG AGC AAG-3' and 5'-TTG CTC GAG GGC AAC TCA GTA GGG TGC TG-3' for GRP33, 5'-CAC CTC GAG GAA TGC CGT CGT GCA CCA CTC CAA C-3' and 5'-GTG CTC GAG TTA GAA AGA GGT GTT GTT GAC TGA AG-3' for GLD-1, and 5'-CTG GAA TTC GGT CGG GGA AAT GGA AAC GAA GG-3' and 5'-GCT GAA TTC TAG TCC TTC ATC CAG CAA GTC-3' for mouse Qk1. The amplified DNA fragment for the mouse Qk1 was digested with *Eco*RI and subcloned in the *Eco*RI site of myc-Bluescript KS and HA-Bluescript KS. The amplified DNA fragments for GRP33 and GLD-1 were digested with *Xho*I and subcloned in the corresponding sites in myc-Bluescript KS and HA-Bluescript KS. The plasmid constructs were verified by dideoxynucleotide sequencing with Sequenase (U.S. Biochemical) and by the Sheldon Biotechnology Institute automated sequencing facility (McGill University). The DNA sequences of GRP33, GLD-1, and Qk1 were identical to the previously reported sequences (Cruz-Alvarez and Pellicer, 1987; Jones and Schedl, 1995; Ebersole et al., 1996). The GLD-1 sequence obtained did not contain the leucine-leucine-lysine sequence found in the alternatively spliced GLD-1 (Jones and Schedl, 1995). The coding sequences amplified from Qk1-7 started at methionine 287 (Ebersole et al., 1996).

The human FMR1-coding region was amplified using the following oligonucleotides,

5'-GAA GAA TTC GGA GCT GGT GGT GGA AGT GC-3' and 5'-TTA GAA TTC TTA GGG TAC TCC ATT CAC-3', with pBlueBacHis2B containing the human FMR1 cDNA as a DNA template [kindly provided by Stephen Warren (Verkerk et al., 1991)]. The amplified fragment was digested with *Eco*RI and subcloned in the *Eco*RI sites of myc-Bluescript KS and HA-Bluescript KS.

Protein expression and protein analysis. HeLa cells were transfected with the vaccinia virus T7 expression system and lysed as previously described (Richard et al., 1995). Histidine-tagged Sam68 was affinity purified with His-Bind Resin (Novagen) charged with nickel according to the manufacturer's instructions. The lysis buffer was adjusted with 1x His binding buffer: 5 mM imidazole-0.5 M NaCl-20 mM Tris-HCl (pH 7.9). Samples were analyzed on sodium dodecyl sulfate (SDS)-polyacrylamide gels and transferred to nitrocellulose membranes. Immunoblotting was performed with the following monoclonal antibodies, anti-myc 9E10 (Evan and Bishop, 1985) and antihemagglutinin [anti-HA (BAbCO)], and the following polyclonal rabbit antibodies, anti-Sam68 (Santa Cruz Biotechnology Inc.), anti-fyn (kindly provided by André Veillette), and anti-BicC antibodies. The designated primary antibody was followed by goat anti-rabbit or goat anti-mouse antibody conjugated to horse radish peroxidase (Organon Teknika-Cappel), and chemiluminescence was used for protein detection (Dupont).

The yeast two-hybrid system. The yeast two-hybrid system was obtained from Stephen Elledge (Durfee et al., 1993). pAS1/Sam68 and pAS1/Sam68ΔN were constructed by amplifying the Sam68 sequences from plasmids myc-Sam68ΔN and myc-Sam68 with the T7 promoter primer and the universal reverse primer containing a *Sal*I site (5'-ACC GGT CGA CGG AAA CAG CTA TGA CCA TGA TTA C-3'). The DNA fragments were digested with *Nco*I (located at the beginning of the myc tag in myc-Bluescript KS$^+$) and *Sal*I and subcloned in similar sites in the poly-linker of pAS1 (Durfee et al., 1993). pACT/Sam68 has been previously described (Richard et al., 1995).

Saccharomyces cerevisiae Y190 was transformed with pACT/Sam68, pACT/Sam68 and pAS1/Sam68ΔN, pACT/Sam68 and pAS1/Sam68, pAS1/Sam68ΔN, or pAS1/Sam68. The expression of pAS1/Sam68 and pAS1/Sam68ΔN fusion proteins was verified by immunoblotting with anti-HA antibodies. The ability of pACT/Sam68 to interact with pAS/SNF1 or pAS/lamin was determined by mating as described previously (Richard et al., 1995). The ability of the yeast to activate the *GAL1-lacZ* reporter gene was assessed by the colony lift assay and scored as white (negative) or blue (positive). The strength of the interactions was quantitated by the O-nitrophenyl-β-D-galactopyranoside spectrophotometric assay, and the β-galactosidase activity was expressed as units of enzyme (Fields and Song, 1989).

In vitro self-association assay using His-Sam68. HeLa cells transfected with His-Sam68 and the indicated myc-Sam68 construct were lysed separately. The nuclei were removed by centrifugation, and the supernatants were mixed for 1 h on ice in the presence of 2-mg/ml heparin. The mixture was divided into two and immunoprecipitated with immunoglobulin G (IgG) [control (Sigma)] or anti-myc antibodies. The bound proteins were separated on SDS-9% polyacrylamide gels, and the proteins were transferred to nitrocellulose membranes and immunoblotted with anti-Sam68 antibodies.

RNase treatments and RNA binding analysis. RNase treatment was performed at either 4°C or 37°C for 1 h with a final concentration of 1 mg RNase A (Boehringer Mannheim) per ml. Incubating the cell lysates at 37°C without RNase was considered a mock treatment. For experiments with RNase inhibitor, RNAguard (Pharmacia) was addedat a final concentration of 1,700 U/ml. Poly(U) binding was performed with poly(U)-Sepharose beads (Pharmacia) or 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 previously described (Wang et al., 1995). Poly(A), poly(G), and poly(C) were coupled to cyanogen-activated Sepharose beads (Sigma) at 5 mg/ml. ^{32}P-labeled cellular RNA was

prepared by incubating 10^7 HeLa cells with 250-μCi/ml orthophosphate overnight. RNA was isolated with the RNeasy minikit (Qiagen) and verified on agarose gels. The labeled RNA (3 x10^6 cpm) was incubated with control or anti-myc immunoprecipitates of either untransfected or transfected cells for 30 min at 4°C, washed three times, and counted. The counted samples were resuspended in sample buffer and separated by SDS-polyacrylamide gel electrophoresis (PAGE); the proteins were transferred to nitrocellulose and analyzed by immunoblotting with anti-myc antibodies.

In situ chemical cross-linking. HeLa or NIH 3T3 cells (10^7) were concentrated into 0.5 ml of phosphate-buffered saline (PBS), pH 8.0, by centrifugation and incubated at room temperature for 30 min with or without 1 mM bis(maleimido)hexane (BMH) (Sigma and Pierce) or disuccinimidyl suberate (DSS) (Pierce) as described elsewhere (Taggart and Pugh, 1996). The cells remained intact during the BMH or DSS treatments, as visualized by light microscopy and trypan blue exclusion. The cells were pelleted by centrifugation, resuspended in 75 μl of 2 x sample buffer [60 mM Tris-HCl (pH 8.0), 2% SDS, 2% dithiotheitol, 10% glycerol, 0.01 % bromophenol blue, 0.25 mM phenylmethanesulfonyl fluoride, and 2 mM sodium vanadate], and loaded on SDS-6.5% polyacrylamide gels. The proteins were transferred to nitrocellulose membranes and immunoblotted with anti-Sam68 polyclonal antibodies.

The poly(U) and SH3-domain binding ability of the Sam68 complex was examined by leaving HeLa cells untreated or treated with BMH in situ; the cells were lysed and incubated with poly(U)-Sepharose or affinity matrices coupled to glutathione S-transferase (GST)-SH3 domains as described elsewhere (Richard et al., 1995). The bound proteins were analyzed on SDS-6.5% polyacrylamide gels, transferred to nitrocellulose, and immunoblotted with anti-Sam68 antibodies.

For transfected HeLa cells, 0.5 x 10^7 adherent cells were incubated at room temperature for 30 min with or without 1 mM BMH. The cells were washed with PBS and either lysed in sample buffer as described above or lysed in lysis buffer and incubated with poly(U)-

Sepharose or uncharged or nickel-charged His-Bind resin. The beads were washed, and the bound proteins were separated by SDS-PAGE, transferred to nitrocellulose membranes, and immunoblotted with the indicated antibody.

Characterization of the Sam68 complex. HeLa cells transfected with His-Sam68, myc-Sam68, and HA-Sam68 were left untreated or treated with BMH. The cells were lysed in lysis buffer supplemented with 50 mM Tris-HCl (pH 8.0), and the lysates were incubated with charged or uncharged His-Bind Resin. The beads were washed, and the bound proteins were eluted with imidazole. The eluted proteins were dialyzed overnight with PBS to remove the excess imidazole. The buffer of the eluted protein was adjusted to 1% Triton X-100 and immunoprecipitated with anti-myc or control antibodies. The bound proteins were loaded on SDS-6.5% polyacrylamide gels, transferred to nitrocellulose membranes, and immunoblotted with anti-HA antibodies.

RESULTS

Interaction of Sam68 with itself by coimmunoprecipitation studies. To determine if Sam68 could oligomerize, we performed coimmunoprecipitation studies with HeLa cells transfected with truncated versions of epitope-tagged Sam68. Two plasmids that encoded Sam68 proteins that were shorter than the endogenous Sam68 were constructed (Fig. 2-1A). The histidine-tagged Sam68 (His-Sam68) contained a deletion of the N-terminal 67 amino acids and migrated with an apparent molecular mass of 60 kDa on an SDS-polyacrylamide gel. The myc-tagged Sam68ΔN (myc-Sam68ΔN) construct contained a deletion of the N-terminal 102 amino acids and had an apparent molecular mass of 50 kDa. HeLa cells were transfected with plasmids encoding myc-Sam68ΔN, His-Sam68, or both myc-Sam68ΔN and His-Sam68 as indicated (Fig. 2-1B). An aliquot of the cell lysate was kept to represent total cell lysate, and the remainder was divided equally and immunoprecipitated with IgG (control) or anti-myc antibodies or incubated with uncharged or charged nickel resin. The bound proteins were separated by SDS-PAGE, transferred to nitrocellulose, and immunoblotted with anti-Sam68 antibodies. Anti-myc immunoprecipitates of cells transfected with myc-Sam68ΔN contained endogenous Sam68 (Fig. 2-1B, lanes 3 and 13). His-Sam68 was also observed in anti-myc immunoprecipitates, when it was cotransfected with myc-Sam68ΔN (lane 13). The nickel-charged resin was less effective in coprecipitating Sam68. Nevertheless, endogenous Sam68 (lane 10) and myc-Sam68ΔN (lane 15) coprecipitated with His-Sam68. These observations demonstrated that Sam68 self-associated and that the N-terminal 102 amino acids were not required for self-association.

Self-association of Sam68 by the yeast two-hybrid system. The yeast two-hybrid system was used to investigate whether the interaction between Sam68 could occur in yeast cells. GAL4 DNA binding hybrids that contained Sam68 or Sam68ΔN in the plasmid pAS-1 were constructed (Durfee et al., 1993). The yeast strain Y190 was transformed with each of these plasmids alone or in combination with a plasmid expressing Sam68 as a GAL4-

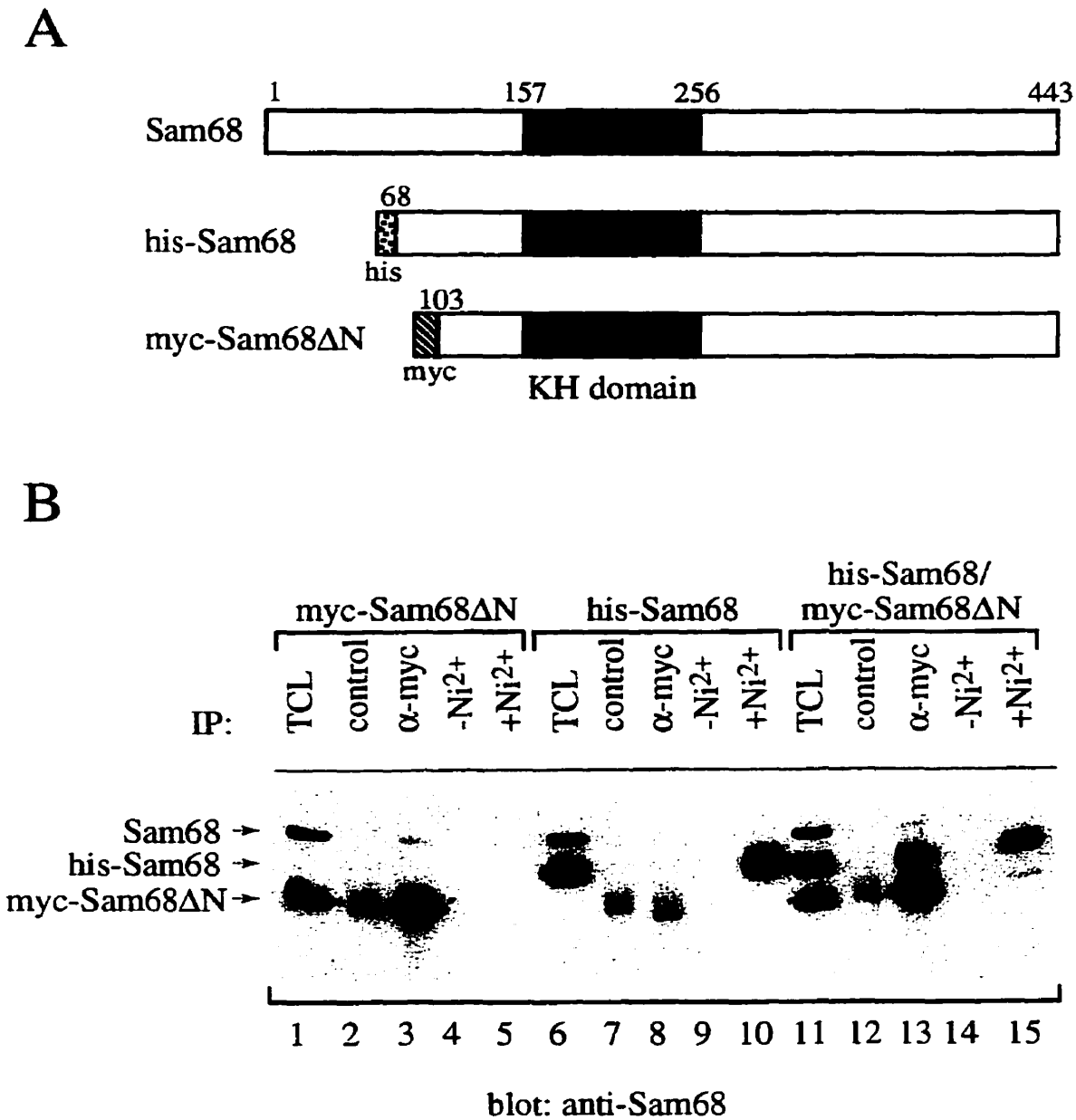


FIG. 2-1. Self-association of Sam68 by coimmunoprecipitation studies. (A) Schematic diagram of the truncated epitope-tagged Sam68 proteins used. The black box represents the Sam68 KH domain. The N-terminal 67 amino acids of Sam68 were replaced by six histidine residues (His-tag; small box with black dots) in His-Sam68. The construct myc-Sam68ΔN is missing the N-terminal 102 amino acids of Sam68 and has a myc epitope tag (hatched box). (B) HeLa cells were transfected with the plasmids encoding myc-Sam68ΔN (lanes 1 to 5), His-Sam68 (lanes 6 to10), or myc-Sam68ΔN plus His-Sam68 (lanes 11 to15). The cells were lysed and immunoprecipitated (IP) with a control or anti-myc antibody or incubated with His-Bind resin uncharged (-Ni2+) or charged with nickel (+Ni2+). The bound proteins as well as total cell lysates (TCL) were separated on SDS-9% polyacrylamide gels, transferred to nitrocellulose, and immunoblotted with anti-Sam68 antibodies. The positions of the endogenous Sam68 (Sam68), His-Sam68, and myc-Sam68ΔN are indicated by arrows, and their approximate molecular masses are 68, 60 and 50 kDa, respectively. The broad bands observed at 50 kDa in lanes 2, 7, 8, and 12 represent the antibody heavy chains.

transactivating hybrid (Durfee et al., 1993). Yeast cells were plated on appropriate selective media and assayed for β-galactosidase activity. The *lac Z* gene was activated, and the colonies turned blue only in the yeast expressing Sam68-Sam68 and Sam68ΔN-Sam68 in both hybrids (Table 2-1). The interaction was strong, with β-galactosidase activities greater than 7 Miller units compared to those of the controls, which did not exceed 1 (Table 2-1). The Sam68-Sam68 interaction was specific because no interaction between Sam68 and SNF1 or lamin was observed as described previously (Table 2-1) (Richard et al., 1995).

TABLE 2-1. Detection of self-association of Sam68 by using the yeast two-hybrid system[a]

DNA-binding hybrid	Activation hybrid	Colony color	β-Gal activity	n
Sam68	Sam68	Blue	7.28 ± 2.43	18
Sam68ΔN	Sam68	Blue	15.36 ± 8.54	6
Sam68	None	White	0.087 ± 0.076	4
Sam68ΔN	None	White	0.837 ± 0.23	4
None	Sam68	White	0.31 ± 0.087	6
SNF1	Sam68	White	ND	3
Lamin	Sam68	White	ND	3

[a]The yeast strain Y190 was cotransformed with plasmids containing Sam68 and Sam68ΔN in pACT (activation hybrid) or pAS (DNA-binding hybrid) as indicated. The strength of the interactions was quantitated by the O-nitrophenyl-β-D-galactopyranoside spectrophotometric assay, and the activity of b-galactosidase is expressed as units of enzyme (Miller units). ND, not determined.

In situ chemical cross-linking of Sam68. In situ chemical cross-linking studies in human and mouse cells were performed to determine whether Sam68 formed dimers, trimers, or larger complexes. HeLa cells were treated or not with BMH and lysed in sample buffer, and the complexes were separated by SDS-PAGE and immunoblotted with anti-Sam68 antibodies. In addition to the 68-kDa band representing monomeric Sam68, a band with an apparent molecular mass of 200 kDa was observed (Fig. 2-2A, lane 2). To confirm that this was not an artifact of the BMH cross-linker, we treated cells in situ with another chemical cross-linker. A similar Sam68 complex was also observed with DSS cross-linking (Fig. 2-2A, lane 4). The complex could be observed also in mouse cells, as BMH treatment of NIH 3T3 resulted in a similar Sam68 complex (Fig. 2-2B).

Oligomerization of Sam68 in vivo. As the apparent molecular mass of the Sam68 complex was similar to the mass predicted for a Sam68 trimer, we investigated whether two and three molecules of Sam68 were present in the complex. HeLa cells were transfected with two different epitope-tagged Sam68 molecules, His-Sam68 and myc-Sam68. The cells were mock or BMH treated, lysed, and incubated with uncharged or charged nickel beads to purify the histidine-tagged Sam68. The bound proteins were separated by SDS-PAGE, transferred to nitrocellulose, and immunoblotted with anti-myc antibodies. A band corresponding to the Sam68 complex was detected (Fig. 2-2C, lane 4). This demonstrated that at least two Sam68 molecules were present in the complex. To verify that a third Sam68 molecule was present in the complex, HeLa cells were transfected with three different epitope-tagged Sam68 molecules. The cells were mock (lanes 1 to 4) or BMH (lanes 5 to 8) treated, lysed, and incubated with uncharged or charged nickel beads to purify the histidine-tagged Sam68 ("1st IP" in Fig. 2-2D). The bound proteins were eluted from the beads and reimmunoprecipitated for myc-Sam68 with control or anti-myc antibodies ("2nd IP" in Fig. 2-2D). The bound complexes were separated by SDS-PAGE, transferred to nitrocellulose membranes, and immunoblotted with anti-HA antibodies to detect HA-Sam68. A band with an approximate molecular mass of 190 kDa was observed (Fig. 2-2D,

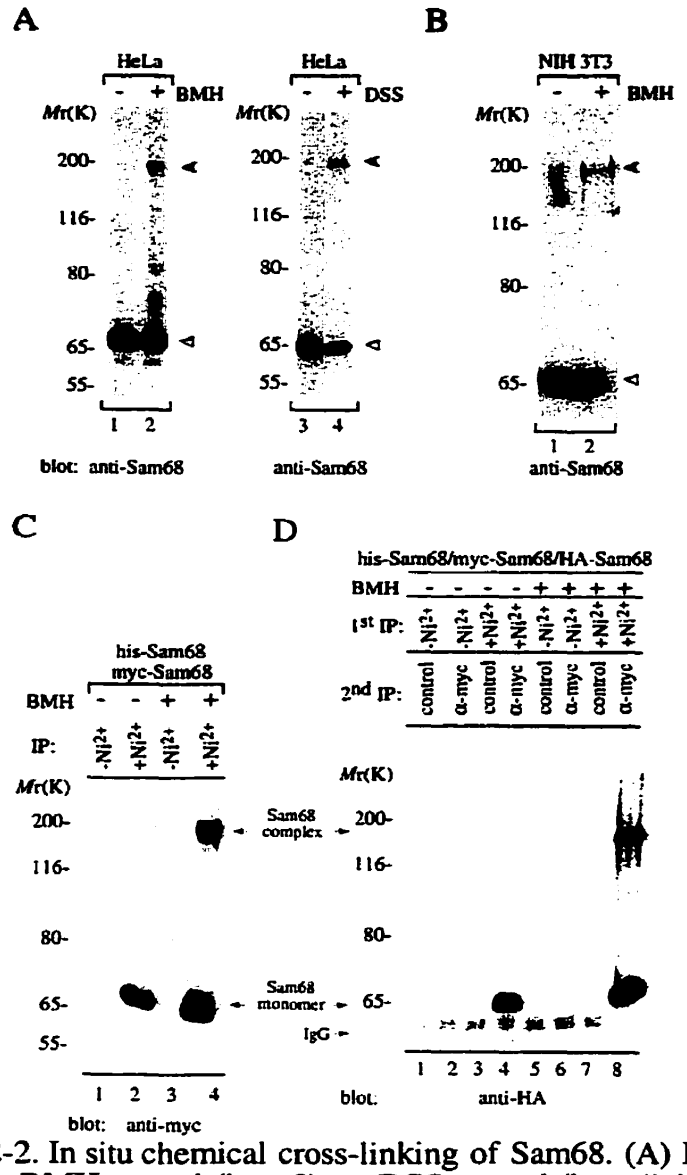


FIG. 2-2. In situ chemical cross-linking of Sam68. (A) HeLa cells were mock treated (lanes 1 and 3), BMH treated (lane 2), or DSS treated (lane 4) in situ. The cells were lysed in sample buffer and separated by SDS-PAGE. The content was transferred to a nitrocellulose membrane and immunoblotted with anti-Sam68 antibodies. (B) NIH3T3 cells were mock or BMH treated and analyzed as for panel A. The open arrowhead represents the un-cross-linked Sam68, and the solid arrowhead represents the cross-linked Sam68 complex. (C) The plasmids encoding His-tagged and myc-tagged Sam68 were transfected in HeLa cells. The cells were left untreated or treated with BMH in situ. The cells were lysed, and the lysates were incubated with His-Bind resin in the presence or absence of nickel. The bound proteins were eluted from the beads and separated on SDS-6.5% acrylamide-gels, transferred to nitrocellulose, and immunoblotted with anti-myc antibodies. (D) The plasmids encoding His-tagged, myc-tagged, and HA-tagged Sam68 were transfected in HeLa cells. The cells were left untreated or treated with BMH in situ. The cells were lysed, and the lysates were incubated with His-Bind resin in the presence or absence of nickel [first immunoprecipitation (1st IP)]. The bound proteins were eluted from the beads and reimmunoprecipitated with control or anti-myc antibodies [second immunoprecipitation (2nd IP)]. The proteins were separated on SDS-6.5% acrylamide gels, transferred to nitrocellulose, and immunoblotted with anti-HA antibodies. The relative molecular weight markers are shown on the left of each panel in kDa. IP, immunoprecipitation.

lane 8). These findings demonstrated that at least three different Sam68 molecules are associated with each other, suggesting that Sam68 forms homotrimers. However, it is possible that three Sam68 molecules associated as a result of an interaction between two separate complexes; therefore, the cross-linked Sam68 complex could also consist of a Sam68 dimer that migrates with an aberrant molecular mass of 190 kDa, or the complex could consist of a Sam68 dimer with an unknown protein with an apparent molecular mass of 60 to 70 kDa that is not p59*fyn* (see below). Although these experiments do not address which of the possibilities is occurring, they clearly show that Sam68 oligomerizes in vivo. The complex in Fig. 2-2C and D had a faster mobility compared to that in Fig. 2-2A and B because the epitope-tagged Sam68 molecules migrate between 60 and 65 kDa due to the absence of the N-terminal 67 amino acids. The 65 kDa band observed in lane 4 (Fig. 2-2D) indicated that the Sam68 complex was purified, but since the complex was not treated with BMH, the complex was dissociated in the sample buffer, and monomeric HA-Sam68 was detected.

The cross-linked Sam68 complex associates with SH3 domains and homopolymeric RNA. We examined whether the Sam68 complex formed after in situ chemical cross-linking could associate with homopolymeric RNA in vitro. HeLa cells were left untreated or treated in situ with BMH, lysed, and examined for their ability to associate with poly(U)-Sepharose. The 200 kDa Sam68 complex observed with BMH treatment associated with poly(U)-Sepharose (Fig. 2-3A, lane 2). These findings demonstrated that Sam68 can associate with RNA as a multimer. As Sam68 is known to associate with the SH3 domains of p59*fyn* and PLCγ1 (Richard et al., 1995), we examined whether Sam68 could associate with these domains as a multimer. Indeed, this was the case, as the cross-linked Sam68 complex associated with the SH3 domain of p59*fyn* and PLCγ1 but not p120rasGAP (Fig. 2-3B). This binding specificity was similar to what has been described previously (Richard et al., 1995). Sam68, therefore, has the capability to associate with both SH3-domain-containing proteins and RNA as a multimer.

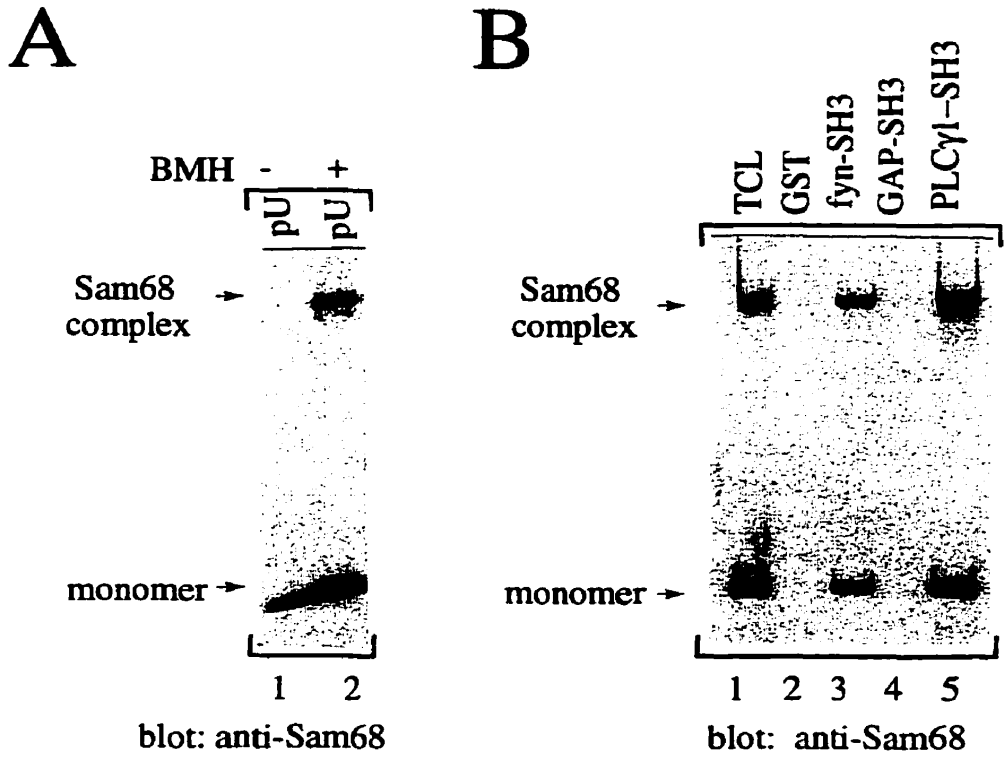


FIG. 2-3. (A) The Sam68 complex binds RNA. HeLa cells were left untreated (lane 1) or BMH treated in situ (lane 2), lysed, and incubated with poly(U)-Sepharose. The bound proteins were analyzed on SDS-6.5% polyacrylamide gels, transferred to nitrocellulose membranes, and immunoblotted with anti-Sam68 antibodies. (B) The Sam68 complex associates with the SH3 domains of p59fyn and PLCγ1. HeLa cells were BMH treated in situ, lysed, and incubated with affinity matrices containing GST, GST-fynSH3, GST-GAP-SH3, and GST-PLCγ1-SH3. The bound proteins were analyzed as for panel A. The Sam68 complex and the Sam68 monomer are indicated. TCL, total cell lysates.

Mapping the Sam68 homopolymeric RNA binding domain. We next wanted to map the homopolymeric RNA binding domain of Sam68 and determine whether it overlapped with the region required for self-association. Sam68 contains two putative RNA binding domains: the N-terminal RGG box, which consists of two RGG repeats, and the KH domain (Burd and Dreyfuss, 1994). We transfected a series of myc-tagged Sam68 constructs containing deletions or amino acid substitutions in HeLa cells and examined their ability to associate with homopolymeric RNA. Poly(U) covalently coupled beads were used because no known physiological RNA targets have been identified for Sam68 and poly(U) has been previously shown to associate with Sam68 (Taylor and Shalloway, 1994). The full-length murine Sam68 protein bound poly(U)-Sepharose beads as well as Sam68 proteins containing N-terminal deletions of 67 and 102 residues [compare Sam68 with Sam68Δ1-67 and Sam68ΔN (Fig. 2-4A)], suggesting that the RGG box is not required. In contrast, a large deletion within the KH domain abolished the poly(U) binding activity of Sam68 (Fig. 2-4A, Sam68ΔKH). However, the KH domain alone was not sufficient to associate with poly(U)-Sepharose (Fig. 2-4A, Sam68:KH). These data demonstrate that the KH domain of Sam68 is necessary but not sufficient for the homopolymeric RNA binding activity.

Deletion of the C-terminal 113 residues had no effect on poly(U) binding activity of Sam68 (Fig. 2-4A, Sam68:330). However, deletion of the C-terminal 149 or 174 amino acid abolished poly(U) binding (Fig. 2-4A, Sam68:294 and Sam68:103-269). This region between residues 294 and 330 is rich in arginine-glycine (RG) repeats (Wong et al., 1992). The Sam68:294 protein contains 4 RG repeats, whereas Sam68:330 contains 11. Internal deletions removing amino acids 295 to 404 had no effect on poly(U) binding (Sam68:Δ294-405). The deletion of amino acids 281 to 338 maintained some poly(U) binding activity (Fig. 2-4A, Sam68:Δ280-339). Sam68:Δ280-339 contains one RG repeat, and Sam68:Δ294-405 contains four. We interpret these data as suggesting that approximately 100 amino acids C-terminal to the KH domain are required to stabilize poly(U) binding in the presence of heparin and that the number of RG repeats might be

irrelevant.

Point mutations previously shown to have physiological effects for FMR1, GLD-1, and Qk1 were introduced in Sam68 and tested for their ability to affect homopolymeric RNA binding activity. The FMR1 mutation of isoleucine 304 to asparagine has been shown to severely reduce poly(U) binding activity (Siomi et al., 1994). To investigate the role of this isoleucine in Sam68 poly(U) binding, we substituted asparagine for the equivalent isoleucine 184 in Sam68 and confirmed that it had severely impaired poly(U) binding activity (Fig. 2-4A, Sam68: I→N). The role of GLD-1 glycine 227 is unknown, but it has been predicted to contact RNA directly (Musco et al., 1996). The equivalent mutation in Sam68 was introduced and severely impaired poly(U) binding activity, confirming its role in RNA binding (Sam68: G→D). Three other mutations were observed in the KH domain of GLD-1, a proline (217)-to-leucine change in loop 1 and a glycine (248 or 250)-to-arginine change in loop 4 (Jones and Schedl, 1995). To investigate the roles of loops 1 and 4 in poly(U) binding, we shortened each loop separately to the size of the corresponding loop present in first KH domain of FMR1 (Verkerk et al., 1991). The deletion of Sam68 loop 1 had little or no effect on poly(U) binding activity (Fig. 2-4A, Sam68:ΔL1). In contrast, deletion of loop 4 had reduced poly(U) binding activity (Sam68:ΔL4). The replacement of glycines 199 and 201 with arginines also resulted in reduced poly(U) binding activity, similar to that observed with Sam68:ΔL4 (Fig. 2-4A, Sam68:2G→R). The point mutation identified in the mouse *quaking* gene (Ebersole et al., 1996) was also introduced in Sam68, replacing glutamic acid 128 with a glycine. This mutation had no effect on poly(U) binding activity (Fig. 2-4A, Sam68:E→G).

Localization of the Sam68 self-association region. The region responsible for Sam68 self-association was delineated by examining the ability of various myc-Sam68 proteins to associate with His-Sam68. Since the molecular mass of the several myc-Sam68 constructs was approximately 68 kDa, we examined the ability of the myc-Sam68 proteins to associate with a shorter Sam68 protein truncated in the N-terminal 67 amino acids (His-

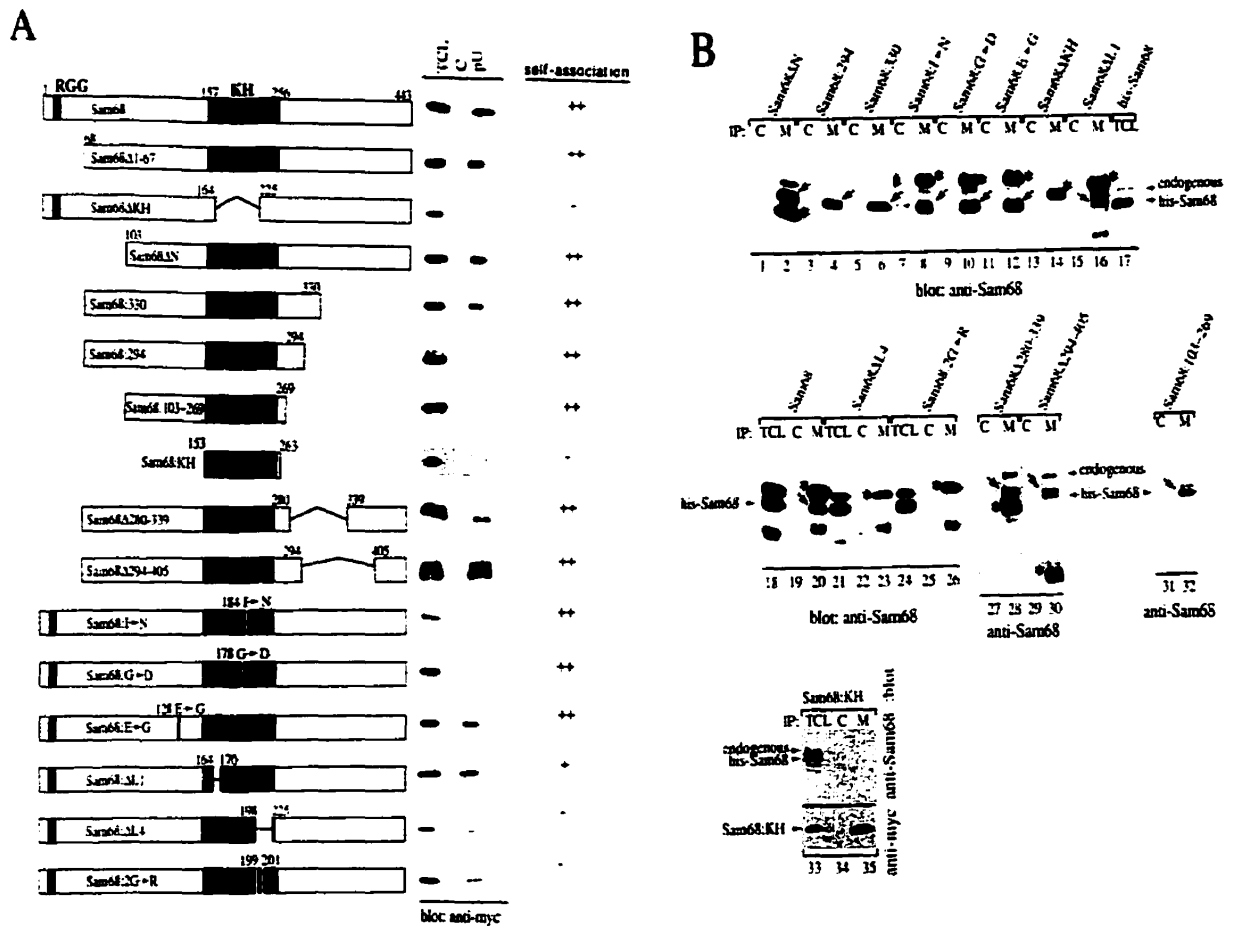


FIG. 2-4. Mapping the poly(U) binding and self-association domains. (A) A schematic diagram representing the various Sam68-myc constructs is shown on the left. The large black boxes represent the KH domain, and the small black boxes represent the arginine-glycine-glycine sequence (RGG). Δ indicates a deletion, and ➔ represents an amino acid substitution. The thin line denotes the absence of amino acid sequences or amino acid substitutions. Each plasmid construct was transfected in HeLa cells, the cells were lysed, and the cell lysates containing 2-mg/ml heparin 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 self-association data shown in panel B are summarized: ++, +, - denote strong, moderate, and no self-association, respectively. (B) HeLa cell lysates containing His-Sam68 were mixed with HeLa cell lysates containing the indicated myc-tagged protein. The mixed lysates were divided equally and immunoprecipitated (IP) with control (C) or anti-myc (M) antibodies. The presence of His-Sam68 in myc immunoprecipitations was analyzed by SDS-PAGE, transferred to nitrocellulose, and immunoblotted with anti-Sam68 antibodies. The endogenous Sam68 and His-Sam68 are indicated. Lane 17 represents the input His-Sam68 for lanes 1 to16. Lanes 18, 21, 24, and 33 represent the input after mixing. The arrows represent His-Sam68, and the asterisks represent myc-Sam68 constructs. In lanes 4, 6, and 32, myc-Sam68 constructs were expressed to levels seen in panel A (data not shown). The slower-migrating band in lane 2 is endogenous Sam68. Longer exposure of lanes 4, 6, and 32 reveals endogenous Sam68 (data not shown). The myc expression of Sam68:KH is shown in the lower panel for lanes 33 to 35. The upper panel for lanes 33 to 35 was immunoblotted with anti-Sam68 antibodies. TCL, total cell lysates.

Sam68). The ability to associate with His-Sam68 correlated with the ability to associate with endogenous Sam68 and to form cross-linked Sam68 complexes in vitro (Fig. 2-1 and data not shown). HeLa cells were transfected with various myc-Sam68 constructs and mixed with HeLa lysates containing His-Sam68. The lysates were immunoprecipitated with control or anti-myc antibodies. The bound proteins were separated by SDS-PAGE, transferred to nitrocellulose membranes, and immunoblotted with anti-Sam68 antibodies. myc immunoprecipitates of Sam68ΔN, Sam68:294, Sam68:330, Sam68:I→N, Sam68:G→D, Sam68:E→G, Sam68, Sam68Δ280-339, Sam68Δ294-405, and Sam68:103-269 contained abundant levels of His-Sam68 (Fig. 2-4B, lanes 2, 4, 6, 8, 10, 12, 20, 28, 30, and 32, respectively). In contrast, myc immunoprecipitates of Sam68ΔKH, Sam68ΔL4, Sam68:2G→R, and Sam68:KH did not coprecipitate any detectable levels of His-Sam68 (lane 14, 23, 26, and 35, respectively). Sam68ΔL1 constantly coprecipitated reduced levels of His-Sam68 (lane 16). The equivalent GLD-1 glycine 227 (Sam68:G→D) and the FMR1 isoleucine 304 (Sam68:I→N) in Sam68, essential for poly(U) binding, were not required for self-association. The *quaking* missense mutation when introduced in Sam68 had no effect on the ability of Sam68 to self-associate or to bind homopolymeric RNA. We determined that the minimal region required for Sam68 self-association was from residues 103 to 269. Within this region, the KH domain loops 1 and 4 were essential. The fact that the KH domain alone (Sam68:KH) did not associated with His-Sam68 suggests that the 50 residues N-terminal to the KH domain are required and might be important to maintain the structure of the KH domain for self-association.

Investigation of the role of RNA in self-association of Sam68. To further examine the role of RNA in Sam68 self-association, cell lysates were treated with RNase before and after mixing. HeLa cell lysates expressing His-Sam68 or myc-Sam68ΔN were lysed separately, and the cell lysates were treated with or without RNase separately and then mixed (Fig. 2-5A, lanes 1 to 6) or mixed and then treated with or without RNase (pre-mix, lanes 7 to12). An aliquot of the mixed proteins was kept to represent total cell lysate and to

demonstrate that Sam68 did not degrade during the RNase treatment. The His-Sam68-myc-Sam68ΔN mixture was immunoprecipitated with IgG (control) or anti-myc antibodies. myc immunoprecipitates contained abundant levels of His-Sam68 and endogenous Sam68 in the mock-treated samples (Fig. 2-5A, lanes 3 and 9). RNase treatment of the His-Sam68- and myc-Sam68ΔN-containing HeLa cell lysates prior to mixing resulted in little if any His-Sam68 in myc immunoprecipitates (lane 6). If the myc-Sam68ΔN and His-Sam68 lysates were mixed before RNase treatment, His-Sam68 was detected in myc immunoprecipitates, albeit at a lower level than without RNase treatment (compare lanes 9 and 12). These data strongly imply that cellular RNA is required for the initial steps of the complex formation. Once the complex is formed, RNase treatment has no effect because either the complex is protecting a fragment of RNA required to stabilize the complex or RNA is no longer required once the complex is formed. The weak association observed after mixing and after RNase treatment (Fig. 2-5A, lane 12) suggests that RNA is required to stabilize the complex.

Although the RNase treatment experiments demonstrated that cellular RNA was required for Sam68 self-association, several Sam68 proteins that self-associated did not bind homopolymeric RNA (Fig. 2-4, Sam68:294, Sam68:103-269, Sam68:I→N, and Sam68:G→D). To address this issue, similar mixing-RNase treatment experiments with the full-length Sam68 as well as two homopolymeric RNA-defective proteins, Sam68:I→N and Sam68:G→D, were performed. The lysates for myc-Sam68, myc-Sam68:I→N, Sam68G→D, and His-Sam68 were incubated without RNase, with an RNase inhibitor, or with RNase at 4°C or 37°C for 1h; mixed; and immunoprecipitated with control or anti-myc antibodies. The bound proteins were separated by SDS-PAGE and analyzed by immunoblotting with anti-Sam68 antibodies. myc-Sam68 associated with both endogenous Sam68 and His-Sam68 in the absence of RNase treatment (Fig. 2-5B, top panel, lanes 3 and 12). The addition of an RNase inhibitor consistently increased the amount of His-Sam68 associating with myc-Sam68 (lanes 6 and 15). RNase treatment at 37°C completely abolished association between myc-Sam68 and His-Sam68 (lane 18), as was observed for

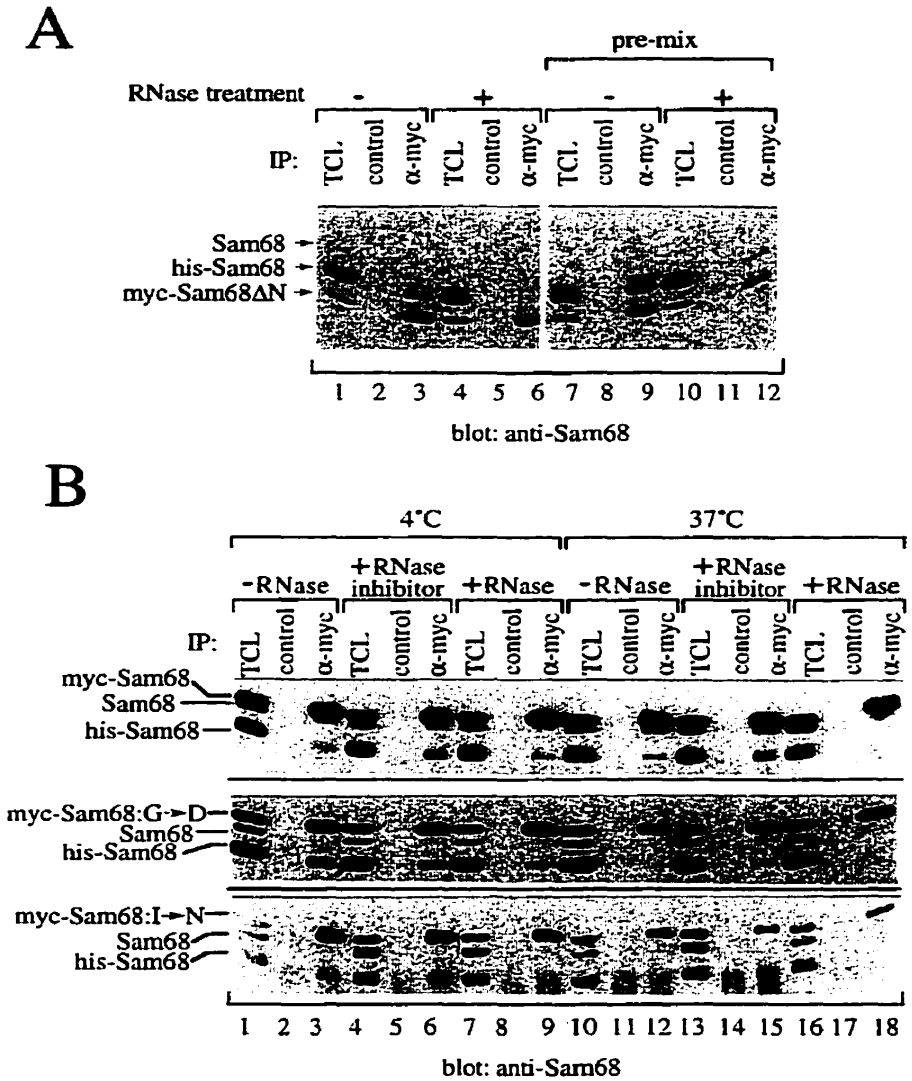


FIG. 2-5. The role of RNA in mediating Sam68 self-association. (A) HeLa cells were transfected with His-Sam68 or myc-Sam68ΔN and lysed separately. Cell lysates were mock treated (lanes 1 to 3) or treated with RNase at 37°C before mixing (lanes 4 to 6). In lanes 7 to12, cell lysates were mock treated and RNase treated at 37°C after mixing. After the RNase treatment and the mixing, cell lysates were immunoprecipitated (IP) with control (lanes 2, 5, 8, and 11) or anti-myc (lanes 3, 6, 9, and 12) antibodies. The bound proteins as well as total cell lysates (TCL) were separated on SDS-9% polyacrylamide gels, transferred, and immunoblotted with anti-Sam68 antibodies. The positions of endogenous Sam68 (Sam68), His-Sam68, and myc-Sam68DN are indicated on the left. (B) HeLa cells were transfected with myc-Sam68, myc-Sam68:G→D, myc-Sam68:I→N, and His-Sam68 separately. Each lysate was incubated at 4° or 37°C for 1 h in the absence of RNase (-RNase), in the presence of RNA inhibitor (+RNase inhibitor), or in the presence of RNase (+RNase). The myc-expressing lysates were mixed with His-Sam68 at 4°C for 1 h and immunoprecipitated (IP) with IgG (control) or anti-myc antibodies. The bound proteins were separated on SDS-9% polyacrylamide gels and immunoblotted with anti-Sam68 antibodies. The migration of the Sam68 constructs is shown on the left. Each experiment shown is a representation of at least three separate experiments. In the lowest panel (myc-Sam68:I→N), the diffuse bands seen in lanes 5, 8, 9, 11, 12, 14, 15, 17, and 18 represent the heavy chain of the antibodies. TCL, total cell lysates.

myc-Sam68ΔN (Fig. 2-5A, lane 6). RNase treatment at 4°C had no effect on self-association of myc-Sam68 (Fig. 2-5A, lane 9).

His-Sam68 associated with myc-Sam68:G→D at 4°C (Fig. 2-5B, middle panel, lane 3) as observed previously in Figure 2-4B. Treatment with the RNase inhibitor at 4°C had no effect (Fig. 2-5B, lane 6), and RNase treatment at 4°C slightly reduced self-association (lane 9). However, myc immunoprecipitates of myc-Sam68:G→D contained very little His-Sam68 at 37°C compared to 4°C (Fig. 2-5B, middle panel, compare lanes 3 and 12). To investigate whether the RNases from the cell lysates were preventing the association, we incubated the lysates with an RNase inhibitor. Interestingly, this treatment consistently increased the association between Sam68:G→D and His-Sam68 (lane 15). The addition of exogenous RNase completely abrogated the association between Sam68:G→D and His-Sam68 (lane 18). We next tested whether Sam68:I→N also required RNA for self-association. myc immunoprecipitates of Sam68:I→N contained His-Sam68 at 4°C (Fig. 2-5B, lower panel, lane 3) as described for Fig. 2-4B. This association was increased with an RNase inhibitor (lane 6) and was completely abolished with RNase treatment at 4°C or any incubation at 37°C (lanes 7 to18). Our findings suggest that Sam68:I→N and Sam68:G→D require RNA to self-associate with Sam68. The RNA bound by these proteins appears to be loosely associated or bound with low affinity because RNase treatment at 4°C or incubation at 37°C diminished or abolished the association.

Although Sam68:I→N and Sam68:G→D self-associated, they were inactive in homopolymeric RNA in vitro (Fig. 2-4A). We also wanted to confirm that the cellular RNA binding capabilities of these proteins was impaired. Sam68, Sam68:I→N, and Sam68:G→D were expressed in HeLa cells and immunoprecipitated with control or anti-myc antibodies. The immunoprecipitations were washed and incubated with labeled cellular RNA. As shown in Fig. 2-6A, Sam68 associated with cellular RNA as described previously (Wong et al., 1992), whereas both Sam68:I→N and Sam68:G→D proteins were unable to associate with cellular RNA. These findings demonstrate that although Sam68:I→N and Sam68:G→D self-associate with Sam68, the complexes are unable to bind RNA.

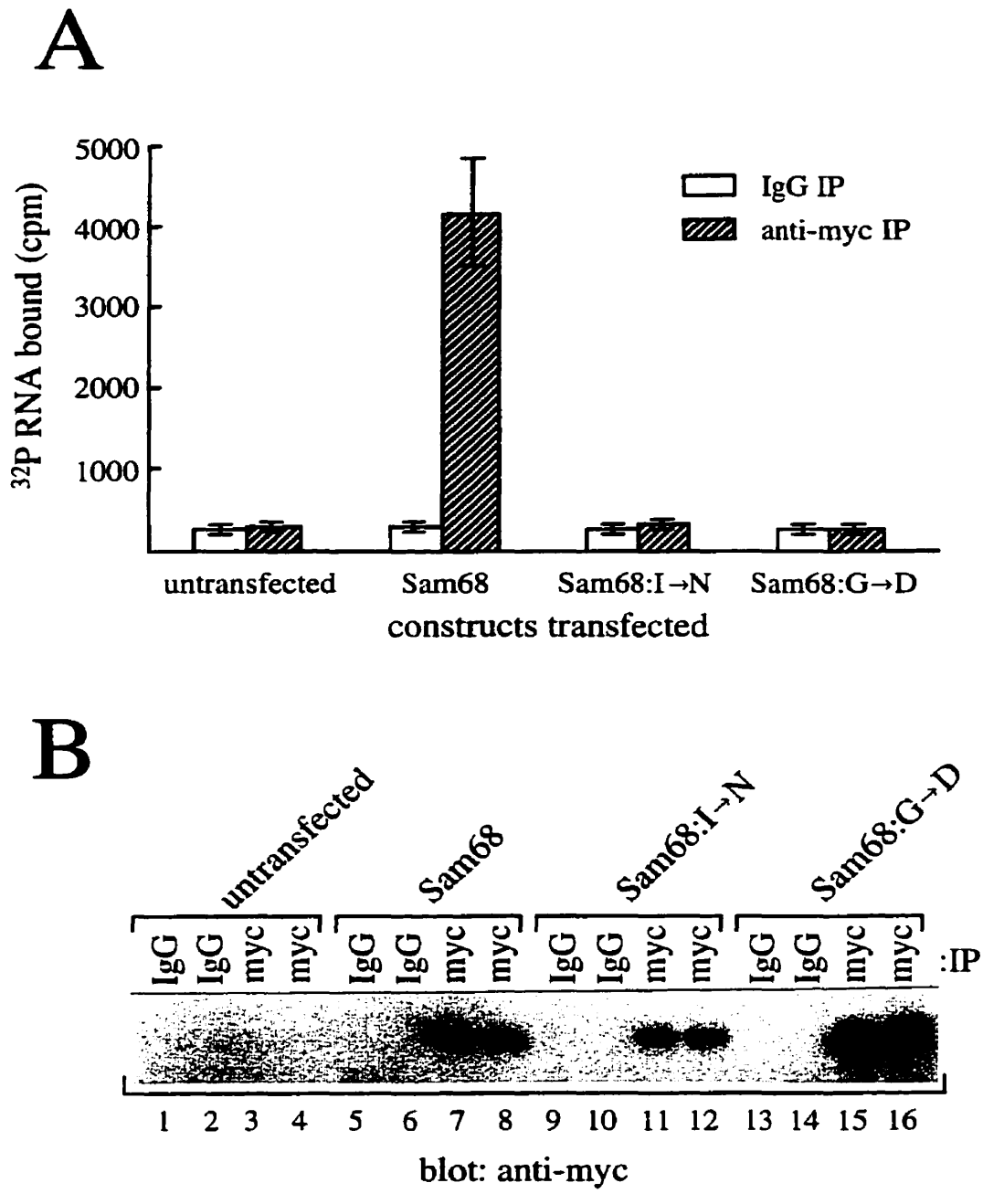


FIG. 2-6. The Sam68:I→N and Sam68:G→D proteins do not associate with cellular RNA. (A) HeLa cells or HeLa cells transfected with Sam68, Sam68:I→N, or Sam68:G→D were immunoprecipitated with a control IgG antibody (white bars) or anti-myc antibody (striped bars). The immunoprecipitated proteins were incubated with ^{32}P-labeled total RNA. The beads were washed, and the radioactivity associated with them was quantitated. Each bar represents the mean ± standard error of at least three separate experiments. (B) A typical representation of the expression of the myc-tagged constructs is shown. IP, immunoprecipitation.

The kinase activity of fyn abrogates self-association. The Src-family tyrosine kinase, p59^{fyn}, phosphorylates Sam68 in vivo when coexpressed in HeLa cells (Richard et al., 1995). To determine whether p59^{fyn} affected Sam68 self-association, HeLa cells were transfected with constructs encoding myc-Sam68ΔN and His-Sam68 alone, in the presence of p59^{fyn}, or in the presence of kinase-inactive p59^{fyn} [D-fyn (Wang et al., 1995)]. The cells were lysed and immunoprecipitated with control or anti-myc antibodies. The bound proteins were separated by SDS-PAGE and immunoblotted with anti-Sam68 antibodies. His-Sam68 coprecipitated with myc-Sam68ΔN in the absence of fyn (Fig. 2-7A, lane 2). The cotransfection of p59^{fyn} significantly reduced the amount of His-Sam68 coprecipitating with myc-Sam68ΔN (lane 4), which was partially restored by D-fyn (lane 6). The expression levels of fyn and D-fyn (Fig. 2-7B, lanes 1 to 3) as well as myc-Sam68ΔN and His-Sam68 (Fig. 2-7B, lanes 4 to 6) were equivalent. Since we have shown elsewhere that the effect of fyn compared to D-fyn was the tyrosine phosphorylation of Sam68 (Wang et al., 1995), our observations suggest that tyrosine phosphorylation regulates self-association.

The effect of fyn on in situ Sam68 complex formation was also tested. In situ cross-linking of cells expressing myc-Sam68 resulted in Sam68 complex formation (Fig. 2-7C, lane 4), which was severely reduced or abolished when fyn was coexpressed (Fig. 2-7C, lane 2). The same membrane was immunoblotted with anti-fyn antibodies (Fig. 2-7D). Besides the monomeric form of p59^{fyn} expressed in lanes 1 and 2, no higher-molecular-mass-containing complex was observed with BMH treatment (lane 2). These data demonstrate that p59^{fyn} is not a component of the Sam68 complex and that p59^{fyn} regulates oligomerization of Sam68 in vivo.

Characterization of GLD-1, GRP33, and Qk1 RNA and protein binding activities. The cDNAs encoding the *A. salina* GRP33, the *C.elegans* GLD-1, and the mouse Qk1 were expressed in HeLa cells as myc epitope-tagged proteins. The ability of GRP33, GLD-1, and Qk1 to bind homopolymeric RNA was compared with that of Sam68.

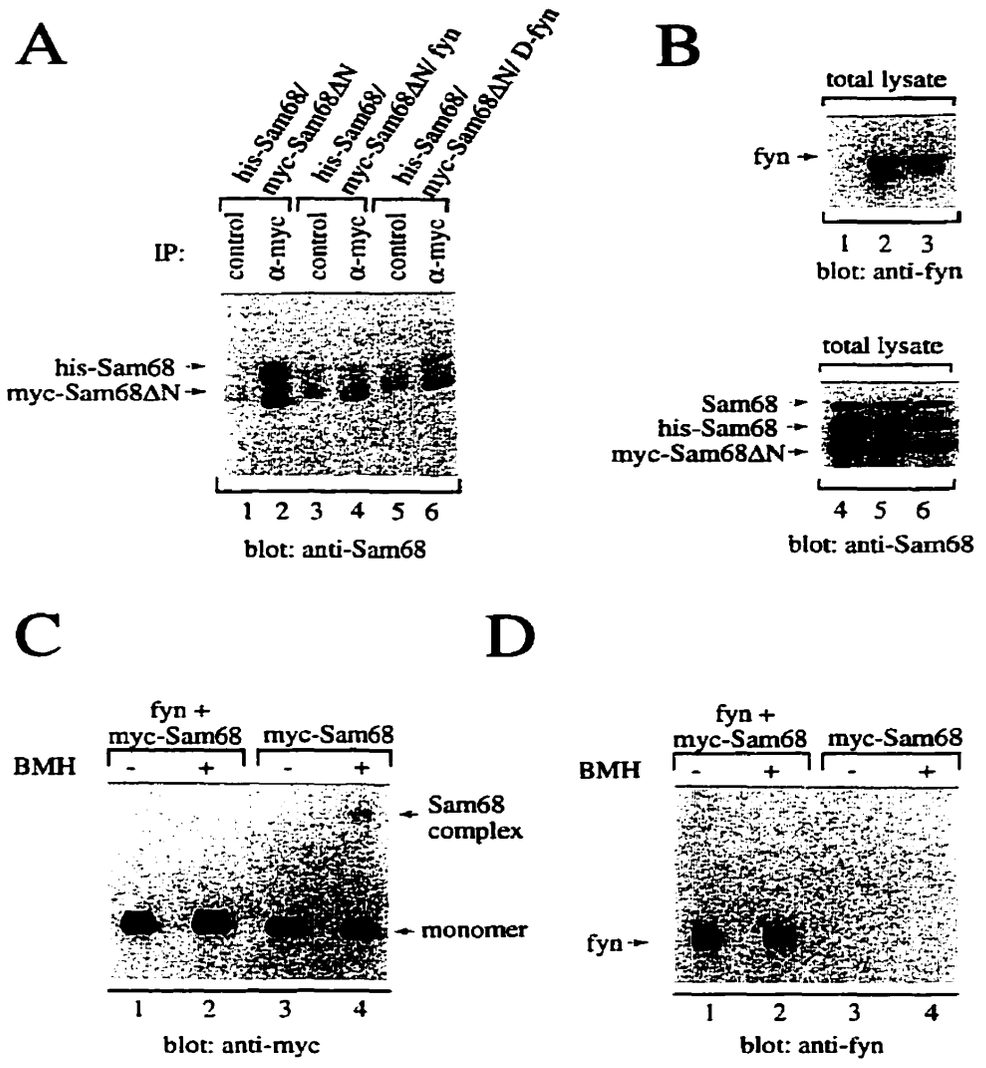


FIG. 2-7. Self-association of Sam68 is severely impaired by $p59^{fyn}$. (A) His-Sam68 was cotransfected with myc-Sam68ΔN in HeLa cells alone (lanes 1 and 2), in the presence of $p59^{fyn}$ (lanes 3 and 4), or in the presence of kinase-inactive $p59^{fyn}$ [dead fyn (D-fyn), lanes 5 and 6]. The cells were lysed and immunoprecipitated (IP) with control or anti-myc antibodies. The bound proteins were separated by SDS-PAGE and immunoblotted with anti-Sam68 antibodies. The positions of His-Sam68 and myc-Sam68ΔN are indicated on the left. The bands with the size of myc-SamΔN in lanes 1, 3, and 5 represent the antibody heavy chain. (B) Equivalent expression levels of $p59^{fyn}$ were obtained for cells transfected with His-Sam68/myc-Sam68ΔN/fyn (lane 2) or His-Sam68/myc-Sam68ΔN/D-fyn (lane 3). Lane 1, HeLa cells transfected with His-Sam68/myc-Sam68ΔN. Equivalent expression levels of His-Sam68 and myc-SamΔN proteins were also obtained for cells transfected with His-Sam68/myc-Sam68ΔN (lane 4), His-Sam68/myc-Sam68ΔN/fyn (lane 5), or His-Sam68/myc-Sam68ΔN/D-fyn (lane 6). (C and D) Cells expressing myc-Sam68 with (lanes 1 and 2) or without (lanes 3 and 4) $p59^{fyn}$ were left untreated (lanes 1 and 3) or treated in situ with BMH (lanes 2 and 4). The cells were lysed in sample buffer, and the proteins were separated on SDS-6.5% polyacrylamide gels, transferred to nitrocellulose membranes, and immunoblotted with anti-myc antibodies (C) or anti-$p59^{fyn}$ antibodies (D). The position of the un-cross-linked (monomer) or cross-linked (Sam68 complex) myc-Sam68 and $p59^{fyn}$ are indicated by arrows.

Sam68 bound both poly(A)- and poly(U)-Sepharose (Fig. 2-8A, lanes 1 to 6). The poly(A) binding activity was weaker than the poly(U) binding activity, consistent with the competition data previously obtained by Taylor and Shalloway (1994). GRP33 bound both poly(A)- and poly(U)-Sepharose, similar to Sam68 (Fig. 2-8A, lanes 13 to18). GLD-1 bound poly(U)-Sepharose (Fig. 2-8A, lanes 7 to12), and surprisingly, Qk1 did not bind any homopolymeric RNA (Fig. 2-8A, lanes 19 to 24). The inability of Qk1 to bind homopolymeric RNA suggests that for Qk1 homopolymeric RNA may not mimic its cellular RNA targets. To verify this hypothesis and to examine whether GRP33 and GLD-1 associated with cellular RNA, ^{32}P-labeled cellular RNA from HeLa cells was incubated with anti-myc immunoprecipitates of HeLa cells expressing Qk1, GRP33, and GLD-1 (Table 2-2). GLD-1 bound about 0.04% of the input RNA whereas GRP33 and Qk1 bound 0.55 and 0.17%, respectively. The weaker RNA binding activity of GLD-1 may suggest that *C.elegans* GLD-1 has few RNA targets in HeLa cells. These findings show that GLD-1, GRP33 and Qk1 are RNA-binding proteins.

We performed in situ chemical cross-linking studies to determine whether GRP33, GLD-1, Qk1, FMR1, and BicC formed a single distinct complex as observed for Sam68. BMH addition of myc-GLD-1-transfected cells resulted in two distinct bands. The approximate molecular masses were 120 and 200 kDa, representing potential dimers and trimers (Fig. 2-8B, lanes 1 and 2). The three distinct bands that were observed after BMH treatment of myc-FMR1-transfected cells had approximate molecular masses ranging from 160 to 210 kDa (lanes 3 and 4). These complexes most likely represent FMR1 homodimers and FMR-FXR heterodimers (Zhang et al., 1995). BMH treatment of GRP33- or Qk1-transfected cells resulted in a single complex that potentially represents homodimers (lanes 5 to 8). The BMH pattern obtained with BicC was different and contained many distinct bands after chemical cross-linking, suggesting that BicC may mediate many protein-protein interactions (lanes 9 and 10).

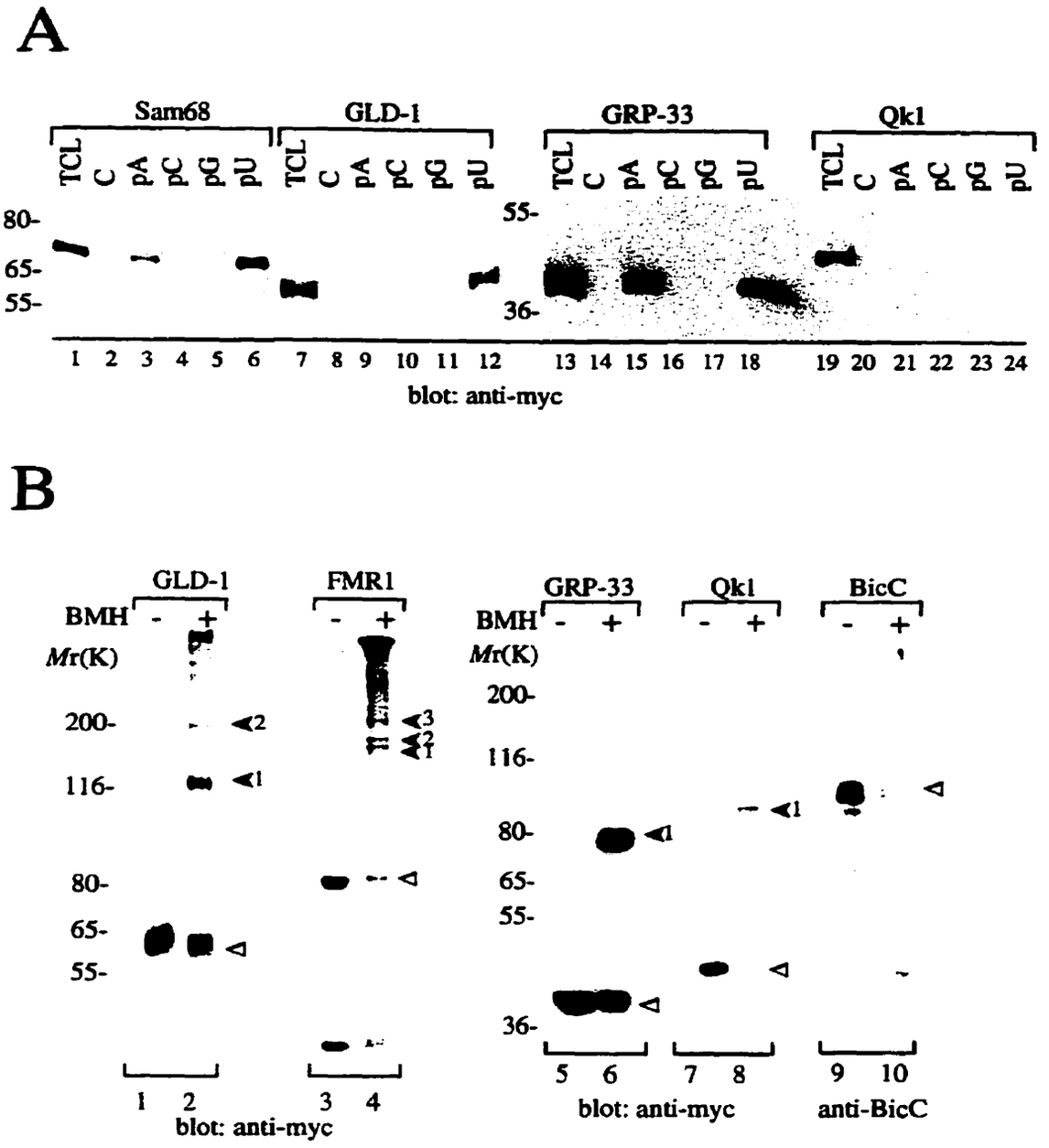


FIG. 2-8. RNA binding characteristics and chemical cross-linking of GRP33, GLD-1, and Qk1. (A) myc-tagged Sam68, GLD-1, GRP33, and Qk1 were expressed in HeLa cells. The cells were lysed, and an aliquot was kept for total cell lysate (TCL) or incubated with control Sepharose (C) or poly(A)-, poly(C)-, poly(G)-, or poly(U)-Sepharose beads in the presence of 2-mg/ml heparin. The bound proteins were separated by SDS-PAGE, transferred to nitrocellulose, and immunoblotted with anti-myc antibodies. The molecular mass markers are shown on the left in kDa. (B) HeLa cells transfected with the myc-GLD-1, myc-FMR1, myc-GRP33, myc-Qk1, and BicC were left untreated or treated in situ with BMH. The cells were lysed in sample buffer, and the proteins separated on SDS-6.5% polyacrylamide gels, transferred to nitrocellulose membranes, and immunoblotted with anti-myc or BicC antibodies. The open arrowheads indicate the monomeric proteins, and the solid arrowheads represent the cross-linked complexes. The relative molecular weight markers are shown on the left in kDa.

TABLE 2-2. Association of GRP33, Qk1, and GLD-1 with cellular RNA[a]

Construct transfected	IP Ab	32P-RNA bound	n	% Bound[b]
None	IgG	259 ± 122	6	<0.01
	Anti-myc	263 ± 108	6	<0.01
myc-GRP33	IgG	235 ± 120	6	<0.01
	Anti-myc	16,362 ± 2,192	6	0.55
myc-Qk1	IgG	251 ± 114	6	<0.01
	Anti-myc	5,243 ± 847	6	0.17
myc-GLD-1	IgG	282 ± 104	6	<0.01
	Anti-myc	1,342 ± 36	6	0.04

[a]HeLa cells transfected with myc-GRP33, -GLD-1, or -Qk1 were lysed and immunoprecipitated (IP) with IgG or anti-myc antibodies (Ab). The beads were washed and incubated with ^{32}P-labeled total cellular RNA. The bound RNA was quantitated and expressed as counts per minute ± the standard errors of the mean.

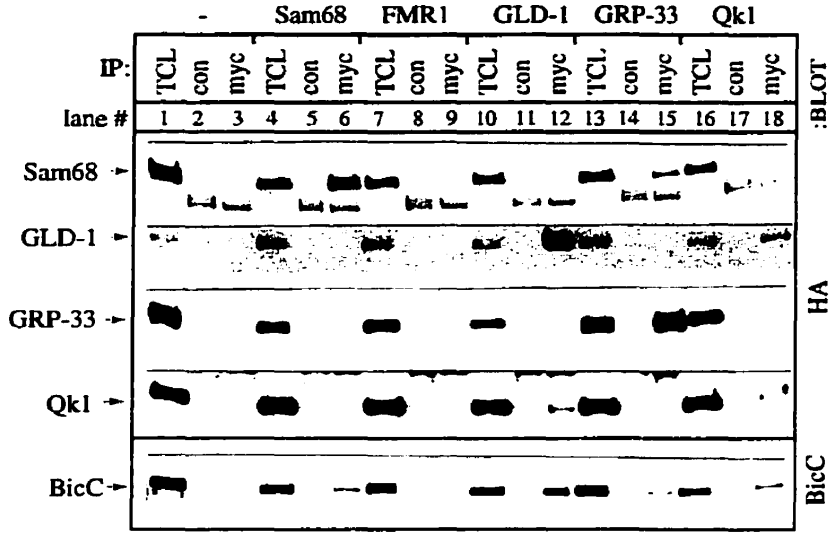
[b]% bound RNA as a percentage of input RNA.

The ability of Sam68, GRP33, GLD-1, and Qk1 to self-associate and associate with each other or unrelated KH domain proteins FMR1 and BicC was examined by cotransfection in HeLa cells. Cells were transfected with the plasmid expressing HA-Sam68 alone (Fig. 2-9A, lanes 1 to 3) or cotransfected with plasmids expressing HA-Sam68 and myc-Sam68 (lanes 4 to 6), HA-Sam68 and myc-FMR1 (lanes 7 to 9), HA-Sam68 and myc-GLD-1 (lanes 10 to12), HA-Sam68 and myc-GRP33 (lanes 13 to15), and HA-Sam68 and myc-Qk1 (lanes 16 to18). The cells were lysed and immunoprecipitated with control or anti-myc antibodies, and immunoblotted with anti-HA antibodies. HA-Sam68 was observed only in myc-Sam68 and myc-GRP33 immunoprecipitates (lanes 6 and 15, Sam68 panel). Similar experiments were performed with all the myc constructs described above with HA-GLD-1, HA-GRP33, and HA-Qk1. HA-GLD-1 was observed in myc-GLD-1 and myc-Qk1 immunoprecipitates (lane 12 and 18, GLD-1 panel). This demonstrated that GLD-1 self-associated and associated with Qk1. HA-Qk1 was observed in myc-GLD-1 and myc-Qk1 immunoprecipitates (lanes 12 and 18, Qk1 panel). These findings demonstrated that Qk1 self-associated. HA-GRP33 was observed in myc-GRP33 immunoprecipitates, indicating that GRP33 self-associated (lane 15, GRP33 panel). We suspect that the inability of myc-Sam68 to immunoprecipitate HA-GRP33 is due to the interference of the epitope tags.

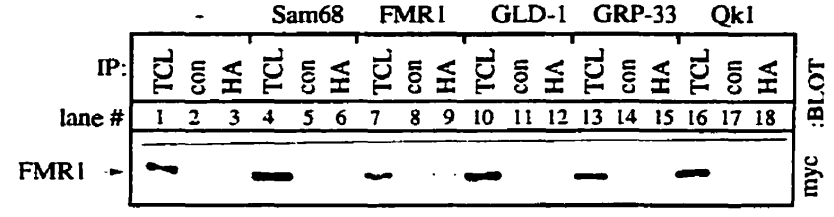
The ability of FMR1 to associate with Sam68 family members was tested. myc-FMR1 was cotransfected with HA-Sam68, HA-FMR1, HA-GLD-1, HA-GRP33 and HA-Qk1. myc-FMR1 was observed only in HA-FMR1 immunoprecipitates, demonstrating that the FMR1 interacted only with itself (lane 2-9B, lane 9). We next tested whether Sam68 family members could associate with BicC. BicC was transfected alone or cotransfected with myc-Sam68, myc-FMR1, myc-GLD-1, myc-GRP33, and myc-Qk1. BicC was observed in myc immunoprecipitates of Sam68, GLD-1, GRP33 and Qk1 but not FMR1 (Fig. 2-9A, BicC panel). This interaction demonstrates that KH domain proteins from different classes can interact. In addition, Sam68:I→N and Sam68:G→D also associated with BicC (data not shown). In summary, GRP33, GLD-1, and Qk1 self-associated (Fig. 2-9C). The closest

family members, GRP33 and Sam68, interacted with each other, and the same was true for GLD-1 and Qk1. Sam68, GLD-1, GRP33, and Qk1 interacted with BicC but not with FMR1 when transfected in mammalian cells.

A



B



C

	Sam68	FMR1	GLD-1	GRP-33	Qk1
Sam68	++	-	-	+	-
GLD-1	-	-	++	-	+
GRP-33	-	-	-	++	-
Qk1	-	-	+	-	+
BicC	+	-	+	+	+
FMR1	-	+	-	-	-

FIG. 2-9. Self-association of GRP33, GLD-1, and Qk1: association with BicC but not FMR1. (A) The HA epitope-tagged Sam68, GLD-1, GRP33, or Qk1 or untagged BicC was transfected alone or cotransfected with myc-tagged Sam68, FMR1, GLD-1, GRP33, or Qk1. The cells were lysed and an aliquot was keep for the total cell lysate (TCL). The remaining cell lysate was divided equally into two and immunoprecipitated (IP) with control (con) or anti-myc antibodies (myc). For GLD-1, the antibodies were covalently coupled to protein A-Sepharose. The bound proteins were separated by SDS-PAGE, transferred, and immunoblotted with anti-HA or anti-BicC antibodies. The position of the HA proteins or BicC is shown on the left. The other bands in the control and anti-myc lanes represent the antibody heavy chains. (B) myc-FMR1 was transfected alone or with HA-tagged Sam68, FMR1, GLD-1, GRP33, or Qk1. Cell lysates were divided equally into two and immunoprecipitated (IP) with control or anti-HA antibodies. The bound proteins were separated by SDS-PAGE and immunoblotted with anti-myc antibodies. The position of myc-FMR1 is shown on the left. (C) The data from panel A and B are summarized. ++, +,- denote strong, moderate, and no association, respectively.

DISCUSSION

We have shown that Sam68 coimmunoprecipitated with itself in mammalian fibroblast cells by using truncated versions of Sam68. Sam68 also self-associated in yeast when expressed with the two-hybrid system. A single Sam68 complex of ~200 kDa was observed after in situ chemical cross-linking experiments in mammalian cells, demonstrating a degree of interaction specificity between the cross-linked Sam68 molecules. By using three different epitope-tagged Sam68 constructs, we demonstrated that the ~200-kDa complex was most likely a Sam68 homotrimer. As Sam68 associates with several SH3- and SH2-domain-containing signaling molecules (Fumagalli et al., 1994; Maa et al., 1994; Taylor and Shalloway, 1994; Weng et al., 1994; Richard et al., 1995; Vogel and Fujita, 1995), it was surprising to observe a single complex after cross-linking. We speculate that Sam68 molecules are in closer molecular proximity with each other than in other complexes, and as a result, cross-linking studies reveal only one Sam68 complex.

The KH domain of Sam68 is embedded in a larger conserved protein motif of ~170 amino acids initially observed by Jones and Schedl (1995). This protein motif spans amino acids 95 to 279 in Sam68, and the KH domain spans amino acids 157 to 256. The minimal region sufficient for self-association was from residues 103 to 269, located within the larger protein motif. Although a role for this conserved motif is unknown, in Sam68 it is necessary and sufficient for RNA-dependent self-association. It will be necessary to determine whether this region fulfills a similar function in the other Sam68 family members.

The KH domain was essential for Sam68 self-association. More specifically, the KH domain loops 1 and 4, unique to this family of single-KH-domain proteins, were required for self-association. The physiological importance of the conserved residues in these loops is underscored by three missense mutations identified in GLD-1 loops 1 and 4 that impair germ line differentiation (Jones and Schedl, 1995). Replacement of the GLD-1 glycine 248 or 250 with arginine in loop 4 results in both loss-of-function (defective oocyte differentiation) and gain-of-function (inappropriate male sex determination in the hermaphrodite germ line) phenotypes (Jones and Schedl, 1995). The GLD-1 proline 217-

to-leucine mutation in loop 1 reverts the gain-of-function sex determination phenotype of the glycine 248-to-arginine mutation and displays the loss-of-function defective oogenesis phenotype. Based on the biochemical studies reported here, we propose that an inability of GLD-1 to self-associate may explain the loss of function caused by these mutations.

The presence of multiple KH domains in most proteins suggests that more than one KH domain is required to form specific and stable RNA interactions (Gibson et al., 1993). Indeed, it has been demonstrated for hnRNP K and FMR1 that their multiple RNA binding domains are all required to optimally associate with RNA (Siomi et al., 1994). The ability of the Sam68 complex to bind homopolymeric RNA demonstrates that single-KH-domain proteins can bind RNA as multimers. However, this does not exclude the possibility that KH domains interact with RNA as monomers. The multimers might have a higher affinity for RNA than the monomer and might have more than one RNA binding site. On the other hand, multimers might have different RNA binding specificities than the monomers. Multimers might interact with each other on neighboring sites on RNA, forming interactions with RNA that resemble beads on a string as seen for hnRNP particles (Burd and Dreyfuss, 1994)

The formation of Sam68 complexes requires cellular RNA because RNase treatment prior to mixing abolished Sam68 self-association. However, once Sam68 complexes were formed, RNase treatment had only a mild effect on Sam68 complex formation. We interpret these data as suggesting that RNA might be required for the initial steps of complex formation and that once the complex is formed, RNA is not required to maintain the complex. Alternatively, RNA could be required at all times to form and stabilize the complex. The Sam68 complex would be resistant to RNase treatment because it protects a fragment of RNA required for its association and stability. The KH domain and approximately 100 residues C-terminal were required for homopolymeric RNA binding in the presence of heparin. Sam68 glycine 178 (GLD-1 glycine 227) and isoleucine 184 (FMR1 isoleucine 304) were essential for RNA binding activity. These results are consistent with the essential role of FMR1 isoleucine 304 in RNA binding (Siomi et al.,

1994) and with the data predicting that GLD-1 glycine 227 forms part of the RNA binding surface (Musco et al., 1996). Our data demonstrate that cellular RNA is required for the self-association of Sam68:G→D and Sam68:I→N with Sam68, but unlike Sam68, the RNA associated with these constructs appeared to be hypersensitive to RNases. As it is known that replacement of the isoleucine with asparagine alters the structure of the vigilin KH domain (Musco et al., 1996), we speculate that the Sam68:G→D and Sam68:I→N proteins weakly associate with RNA and are unable to form stable protein-RNA complexes. These complexes would be destabilized by temperature and RNases and stablilized with RNase inhibitors. In contrast, Sam68:E→G, the conserved glutamic acid residue identified in the *quaking* mouse (Ebersole et al., 1996), had no effect on homopolymeric RNA binding activity or the ability of Sam68 to self-associate. Therefore, this glutamic acid residue must be involved in a separate unidentified function. It is interesting that the mouse Qk1 was able to associate with cellular RNA but not homopolymeric RNA. In agreement with these results, it has been observed that the *Xenopus* Qk1 can associate with *Xenopus*cellular RNA (Zorn et al., 1997).

The tyrosine phosphorylation of Sam68 can be observed when overexpressed with a Src-family tyrosine kinase and/or in cells arrested in mitosis with nocodazole (Wong et al., 1992; Fumagalli et al., 1994; Taylor and Shalloway, 1994; Weng et al., 1994; Richard et al., 1995). To date, no cell surface receptor has been shown to physiologically induce the tyrosine phosphorylation of Sam68. We demonstrated previously that the coexpression of p59^{fyn} with Sam68 results in Sam68 tyrosine phosphorylation (Richard et al., 1995) and that the phosphorylation of Sam68 abrogated its homopolymeric RNA binding activity (Wang et al., 1995). This was a direct effect of phosphorylation because phosphatase treatment restored RNA binding activity (Wang et al., 1995). We hypothesized two models by which RNA binding activity was inhibited by phosphorylation. The first model proposed was that Sam68 bound RNA as an oligomer and phosphorylation interfered with the oligomerization. The second model proposed was that the C-terminal tyrosine-rich region of Sam68 sterically inhibits the KH domain from accessing RNA. Since oligomerization of

Sam68 was abrogated with $p59^{fyn}$, this suggested that tyrosine phosphorylation negatively regulates complex formation. Because Sam68 does not bind RNA when phosphorylated and RNA is required for self-association, oligomerization may be inhibited because the phosphorylated Sam68 is unable to bind RNA, or alternatively, RNA binding might be affected because the phosphorylated Sam68 is unable to oligomerize. Our current working model for Sam68 function is that in the nonphosphorylated state Sam68 associates with RNA and with SH3 domain proteins such as Src kinases and PLCγ as a multimer. The activation of Src-family kinases phosphorylates Sam68 and disassembles the Sam68 complex. As a result, the RNA binding activity of Sam68 would be inhibited. The tyrosine-phosphorylated monomeric form of Sam68 would now be available to associate with Grb2 via its SH2 domain, and the Src kinases and PLCγ would be available to associate with their SH3 and SH2 domains. This model suggests that the monomeric form of Sam68 is mainly involved in signaling and that the multimer is involved in RNA binding and some signaling functions. These distinct functions of Sam68 would then be regulated by tyrosine phosphorylation. Consistent with this model of Sam68 function is the discovery of a splice variant of Sam68 that contains all the signaling motifs but not the KH domain (Barlat et al., 1997). This splice variant would only be able to fulfill the Sam68 signaling functions. It has been shown that this variant does not bind RNA (Barlat et al., 1997), and based on the data presented here, we expect that this splice variant would be unable to self-associate.

The ability of GRP33, GLD-1, and Qk1 to oligomerize and form specific complexes after chemical cross-linking studies demonstrates that oligomerization is a general feature of this subfamily of KH domain proteins. Our studies demonstrate that these family members are also RNA-binding molecules. However, the regulation by tyrosine phosphorylation may not be a common property of all members, and it remains to be determined whether GRP33, GLD-1, and Qk1 are phosphoproteins. It is intriguing that GRP33 and Sam68 as well as GLD-1 and Qk1 associated with one another, implying that heteromultimers might be forming as seen for FMR and FXR proteins (Zhang et al., 1995). However, we suspect that the formation of these complexes with other family members in their respective species is

unlikely because of the chemical cross-linking experiments. These experiments demonstrated that a single complex for each member was formed with the exception of GLD-1 and most likely represented homomultimers. The association of Sam68 family members with BicC, a protein with five KH domains (Mahone et al., 1995), demonstrates that the Sam68 family members can associate with other KH domain proteins.

In conclusion, we demonstrate that Sam68 self-associated in the presence of cellular RNA and that the KH domain loops 1 and 4 were required for self-association. Sam68 formed multimers in vivo that bound homopolymeric RNA and SH3 domains in vitro. Other family members, including the *A. salina* GRP33, *C. elegans* GLD-1, and the mouse Qk1, also oligomerized. These observations indicate that the single KH domain found in the Sam68 family can mediate both protein-RNA and protein-protein interactions.

ACKNOWLEDGMENTS

We thank John Hiscott, Antonis Koromilas, Louise Larose, Tim Schedl, Andrey Shaw, and André Veillette for valuable comments and for critically reading the manuscript. The homopolymeric RNA binding analysis was initiated in Andrey Shaw's laboratory. We warmly acknowledge this and thank him for his constant encouragement. We are grateful to Stephen Elledge for the use of the yeast two-hybrid components. We thank Stephen Warren for the human FMR1 cDNA, André Veillette for anti-fyn antibodies, and Mary Sutherland for screening the mouse cDNA library. We thank the members of the Terry Fox Group in Molecular Oncology for stimulating discussions.

B.B.D is supported with funds from a Canderel fellowship. C.H. is supported by a Challenge 1996 summer award. P.L is supported by grants from the National Cancer Institute of Canada (NCIC) with funds from the Canadian Cancer Society and is a Research Scientist of the NCIC. This work was supported by grants from the Medical Research Council of Canada and from an establishment grant from the Fonds de la Recherche en Santé du Québec to S.R. S.R. is a Scholar of the Medical Research Council of Canada.

REFERENCES

Arning, S., Gruter, P., Bilbe, G. and Kramer, A. 1996. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. RNA, 2:794-810.

Ashley, C. T., Wilkinson, K. D., Reines, D. and Warren, S. T. 1993. FMR1 protein: conserved RNP family domains and selective RNA binding. Science, 262:563-566.

Barlat, I., Maurier, F., Duchesne, M., Guitard, E., Tocque, B. and Schweighoffer, F. 1997. A role for Sam68 in cell cycle progression antagonized by a spliced variant within the KH domain. J. Biol. Chem., 272:3129-3132.

Bunnell, S. C., Henry, P. A., Kolluri, R., Kirchhausen, T., Rickles, R. J. and Berg, L. J. 1996. Identification of Itk/Tsk Src homology 3 domain ligands. J. Biol. Chem., 271:25646-25656.

Burd, C. G. and Dreyfuss, G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. Science, 265:615-621.

Cruz-Alvarez, M. and Pellicer, A. 1987. Cloning of a full-length complementary cDNA for as *Artemia salina* glycine-rich protein. J. Biol. Chem., 262:13377-13380.

DeBoulle, K., Verkerk, A. J. M. H., Reyniers, E., Vits, L., Hendrickx, J., Roy, B. V., Bos, F. V. D., DeGraaff, E., Oostra, B. A. and Willems, P. J. 1993. A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nat. Genet., 3:31-35.

Durfee, T., Becherer, K., Chen, P. L., Yeh, S. H., Yang, Y., Kilburn, A. E., Lee, W. H. and Elledge, S. J. 1993. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Genes Dev., 7:555-569.

Ebersole, T. A., Chen, Q., Justice, M. J. and Artzt, K. 1996. The *quaking* gene unites signal transduction and RNA binding in the developing nervous system. Nat. Genet., 12:260-265.

Evan, G. I. and Bishop, J. M. 1985. Isolation of monoclonal antibodies specific for the human c-myc protooncogene product. Mol. Cell. Biol., 4:2843-2850.

Fields, S. and Song, O. 1989. A novel genetic system to detect protein-protein interactions.

Nature, 340:245-246.

Francis, R., Barton, M. K., Kimbel, J. and Schedl, T. 1995a. Control of oogenesis, germline proliferation and sex determination by the *C. elegans*gene *gld-1*. Genetics, 139:579-606.

Francis, R., Maine, E. and Schedl, T. 1995b. *gld-1*, a cell-type specific tumor suppressor gene required in *C. elegans*. Genetics, 139:607-630.

Fumagalli, S., Totty, N. F., Hsuan, J. J. and Courtneidge, S. A. 1994. A target for Src in mitosis. Nature, 368:871-874.

Fusaki, N., Iwamatsu, A., Iwahima, M. and Fujisawa, J. 1997. Interaction between Sam68 and Src family tyrosine kinases, fyn and lck, in T cell receptor signaling. J. Biol. Chem., 272:6214-6219.

Gibson, T. J., Thompson, J. D. and Heringa, J. 1993. The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. FEBS Lett., 324:361-366.

Hardy, R. J., Loushin, C. L., Friedrich Jr., V. L., Chen, Q., Ebersole, T. A., Lazzarini, R. A. and Artzt, K. 1996. Neural cell type-specific expression of QKI proteins is altered in the *quaking* viable mutant mice. J. Neurosci., 16:7941-7949.

Jones, A. R., Francis, R. and Schedl, T. 1996. GLD-1, a cytoplasmic protein essential for oocyte differentiation, shows stage- and sex-specific expression during *Caenorhabditis elegans* germline development. Dev. Biol., 180:165-183.

Jones, A. R. and Schedl, T. 1995. Mutations in GLD-1, a female germ cell-specific tumor suppressor gene in *C.elegans*, affect a conserved domain also found in Sam68. Genes Dev., 9:1491-1504.

Lawe, D. C., Hahn, C. and Wong, A. J. 1997. The Nck SH2/SH3 adaptor protein is present in the nucleus with the nuclear protein Sam68. Oncogene, 14:223-231.

Lock, P., Fumagalli, S., Polakis, P., McCormick, F. and Courtneidge, S. A. 1996. The human p62 cDNA encodes Sam68 and not the ras-GAP-associated p62 protein. Cell, 84:23-24.

Maa, M.-C., Leu, T.-H., Trandel, B. J., Chang, J.-H. and Parsons, S. J. 1994. A protein that is related to GTPase activating protein-associated p62 complexes with phospholipase Cγ. Mol. Cell. Biol., 14:5466-5473.

Mahone, M., Saffman, E. E. and Lasko, P. F. 1995. Localised Bicaudal-C RNA encodes a protein containing a KH domain, the RNA-binding motif for FMR1. EMBO J., 14:2043-2055.

McBride, A. E., Schlegel, A. and Kirkegaard, K. 1996. Human protein Sam68 relocalization and interaction with poliovirus RNA polymerase in infected cells. Proc. Natl, Acad. Sci. USA, 93:2296-2301.

Musco, G., Stier, G., Joseph, C., Morelli, M. A. and Pastore, A. 1996. Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. Cell, 85:237-245.

Nussbaum, R. L. and Ledbetter, D. H. 1995. The fragile X syndrome, p. 795-810. In C. R. Scriver, A. Beaudet, W. S. Sly and D. Valles (ed.), Metabolic basis of inherited disease. MacGraw-Hill, New York.

Pawson, T. 1995. Protein modules and signalling networks. Nature, 373:573-580.

Pieretti, M., Zhang, F., Fu, Y.-H., Warren, S. T., Oostra, B. A., Caskey, C. T. and Nelson, D. L. 1991. Absence of expression of the FMR-1 gene in fragile X syndrome. Cell, 66:817-822.

Richard, S., Yu, D., Blumer, K. J., Hausladen, D., Olszowy, M. W., Connelly, P. A. and Shaw, A. S. 1995. Association of p62, a multi-functional SH2- and SH3-binding protein, with src-family tyrosine kinases, Grb2, and phospholipase Cγ-1. Mol. Cell. Biol., 15:186-197.

Richard, S. and Zingg, H. H. 1991. Identification of a retinoic acid response element in the human oxytocin promoter. J. Biol. Chem., 266:21428-21433.

Sidman, R. L., Dickie, M. M. and Appel, S. H. 1964. Mutant mice (quaking and jimpy) with deficient myelination in the central nervous system. Science, 144:309-311.

Siomi, H., Choi, M., Siomi, M. C., Nussbaum, R. L. and Dreyfuss, G. 1994. Essential role for KH domain in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. Cell, 77:33-39.

Siomi, H., Matunis, M. J., Michael, W. M. and Dreyfuss, G. 1993. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. Nucleic Acids Res., 21:1193-1198.

Siomi, M. C., Zhang, Y., Siomi, H. and Dreyfuss, G. 1996. Specific sequences in the Fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interaction among them. Mol. Cell. Biol., 16:3825-3832.

Taggart, A. K. P. and Pugh, B. F. 1996. Dimerization of TFIID when not bound to DNA. Science, 272:1331-1333.

Taylor, S. J., Anafi, M., Pawson, T. and Shalloway, D. 1995. Functional interaction between c-src and its mitotic target, Sam68. J. Biol. Chem., 270:10120-10124.

Taylor, S. J. and Shalloway, D. 1994. An RNA-binding protein associated with src through its SH2 and SH3 domains in mitosis. Nature, 368:867-871.

Toda, T., Iida, A., Miwa, T., Nakamura, Y. and Imai, T. 1994. Isolation and characterization of a novel gene encoding nuclear protein at a locus (D11S636) tightly linked to multiple endocrine neoplasia type 1 (MEN1). Hum. Mol. Genet., 3:465-470.

Trub, T., Frantz, J. D., Miyazaki, M., Band, H. and Shoelson, S. E. 1997. The role of a lymphoid-restricted, Grb2-like SH3-SH2-SH3 protein in T cell receptor signaling. J. Biol. Chem., 272:894-902.

Verkerk, A. J. M. H., Pieretti, M., Sutcliffe, J. S., Fu, Y.-H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F., Eussen, B. E., van Ommen, G. J. B., Blonden, L. A. J., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A. and Warren, S. T. 1991. Identification of a gene (FMR1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell, 65:905-914.

Vogel, L. B. and Fujita, D. J. 1995. p70 phosphorylation and binding to p56lck is an early event in interleukin-2 induced onset of cell cycle progression in T-lymphocytes. J. Biol.

Chem., 270:2506-2511.

Wang, L. L., Richard, S. and Shaw, A. S. 1995. p62 association with RNA is regulated by tyrosine phosphorylation. J. Biol. Chem, 270:2010-2013.

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. Identification of Src, Fyn, and Lyn SH3-binding proteins: Implications for a function of SH3 domains. Mol. Cell. Biol., 14:4509-4521.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J. et al. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature, 368:32-38.

Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M. F., Polakis, P. and McCormick, F. 1992. Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. Cell, 69:551-558.

Zhang, Y., O'Connor, J. P., Siomi, M. C., Srinivasan, S., Dutra, A., Nussbaum, R. L. and Dreyfuss, G. 1995. The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. EMBO J., 14:5358-5366.

Zorn, A. M., Grow, M., Patterson, K. D., Ebersole, T. A., Chen, Q., Artzt, K. and Krieg, P. A. 1997. Remarkable sequence conservation of transcripts encoding amphibian and mammalian homologues of *quaking* a KH domain RNA-binding protein. Gene, 188:199-206.

Chapter 3

STRUCTURE-FUNCTION ANALYSIS OF QK1: A LETHAL POINT MUTATION IN MOUSE *quaking* PREVENTS HOMODIMERIZATION

PREFACE

Genetic mutations in the GSG domain that result in growth or developmental defects have been isolated in a number of genes. In the previous study (chapter 2), we characterized some of these mutations using Sam68. An interesting finding was that introduction of the Qk1 lethal point mutation (glutamic acid 48 to glycine) in Sam68 has no effect on the ability of Sam68 to oligomerize and to bind RNA. In this study, we intend to characterize this mutation using Qk1, the protein from which the mutation was originally isolated.

SUMMARY

Qk1 is a member of the KH domain family of proteins that includes Sam68, GRP33, GLD-1, SF1, and Who/How. These family members are RNA binding proteins that contain an extended KH domain embedded in a larger domain called the GSG (for GRP33-Sam68-GLD-1) domain. An ethylnitrosourea-induced point mutation in the Qk1 GSG domain alters glutamic acid 48 to a glycine and is known to be embryonically lethal in mice. The function of Qk1 and the GSG domain as well as the reason for the lethality are unknown. Here we demonstrate that the Qk1 GSG domain mediates RNA binding and Qk1 self-association. By using in situ chemical cross-linking studies, we showed that the Qk1 proteins exist as homodimers in vivo. The Qk1 self-association region was mapped to amino acids 18 to 57, a region predicted to form coiled coils. Alteration of glutamic acid 48 to glycine (E→G) in the Qk1 GSG domain (producing protein Qk1:E→G) abolishes self-association but has no effect on the RNA binding activity. The expression of Qk1 or Qk1:E→G in NIH 3T3 induces cell death by apoptosis. Approximately 90% of the remaining transfected cells are apoptotic 48 h after transfection. Qk1:E→G was consistently

more potent at inducing apoptosis than was wild-type Qk1. These results suggest that the mouse *quaking* lethality (E→G) occurs due to the absence of Qk1 self-association mediated by the GSG domain.

INTRODUCTION

The mouse *quaking* gene encodes the Qk1 RNA binding proteins (Ebersole et al., 1996). The type of RNA binding domain found in Qk1, known as a KH domain, was originally identified in the heterogeneous nuclear ribonucleoprotein K [hnRNP K (Gibson et al., 1993; Siomi et al., 1993)]. KH domains are evolutionarily conserved domains that are thought to make direct protein-RNA contacts with a three-dimensional βααββα-fold (Musco et al., 1996). The Qk1 KH domain is embedded in a larger conserved domain of ~200 amino acids called the GSG domain. The GSG domain was initially identified by aligning the first three family members [GRP33, Sam68, GLD-1 (Jones and Schedl, 1995)]. The boundaries of this new protein module have become clearer with the identification of new family members (Arning et al., 1996; Ebersole et al., 1996). This domain is also called STAR [for signal transduction and activation of RNA (Vernet and Artzt, 1997)] and the SGQ [Sam68, GLD-1, Qk1 (Lin et al., 1997)] domain. GSG domain family members include the *Artemia salina* GRP33 (Cruz-Alvarez and Pellicer, 1987), human Sam68 (Wong et al., 1992), *Caenorhabditis elegans* GLD-1(Jones and Schedl, 1995), human SF1(Arning et al., 1996), *Drosophila* Who/How (Baehrecke, 1997; Fyrberg et al., 1997; Zaffran et al., 1997), *Xenopus* Xqua (Zorn et al., 1997), and mouse Qk1 (Ebersole et al., 1996). The features of the GSG domain include a single KH domain that is longer than most other KH domains (Musco et al., 1996). In addition to the KH domain, the GSG domain is composed of ~75 amino acids N-terminal and ~25 amino acids C-terminal of the KH domain [for a review, see (Vernet and Artzt, 1997)].

GSG proteins share several properties, including RNA binding (Wong et al., 1992; Arning et al., 1996; Chen et al., 1997; Lin, et al., 1997; Zorn et al., 1997) and self-association (Chen et al., 1997; Zorn and Krieg, 1997). With the exception of the human SF1 protein, which functions as a splicing factor (Arning et al., 1996), the roles of the GSG proteins in cellular processes are not known. Genetic studies with GSG domain proteins have demonstrated the roles of these proteins in development, differentiation, myelination, and tumorigenesis. In *C. elegans*, GLD-1 is required for germ cell differentiation (Francis

et al., 1995a; Francis et al., 1995b; Jones and Schedl, 1995). One class of alleles results in a recessive tumorous germ line phenotype, suggesting that GLD-1 functions as a tumor suppressor (Jones and Schedl, 1995). The Qk1 proteins of *Drosophila melanogaster*, *Xenopus laevis*, and mice have been characterized. The *Drosophila* Who/How protein, a Qk1 homolog, has been shown to be critical for skeletal muscle development since weak alleles result in flies with "held-out" wings (Baehrecke, 1997; Zaffran et al., 1997). One such allele contains a point mutation in loop 4 of the Who/How KH domain (Baehrecke, 1997). The *Xenopus* Xqua protein, another Qk1 homolog, has been shown to be necessary for notochord development (Zorn and Krieg, 1997). Mice that are homozygous for the *quaking* viable allele have a severe deficiency of myelin throughout their nervous system and, as a consequence, develop a characteristic tremor (Sidman et al., 1964). The genetic lesion in the *quaking* viable mouse has been mapped to the *qk1* promoter-enhancer region (Ebersole et al., 1996). The defect in these mice is the absence of Qk1-6 and Qk1-7 protein expression from the myelin-forming oligodendrocytic cells (Hardy et al., 1996). Another class of mouse *quaking* mutations is embryonic lethal (Bode, 1984; Justice and Bode, 1988; Shedlovsky et al., 1988). One such allele, qk^{kt4}, was found to alter glutamic acid 48 to glycine in the QUA1 region of the Qk1 GSG domain (Ebersole et al., 1996); the cause for the lethality is unknown.

We have characterized the genetic point mutations identified in GLD-1 and Qk1 by using Sam68 (Chen et al., 1997). Substitution of the corresponding GLD-1 glycine 227 to aspartic acid in Sam68 abolishes RNA binding, suggesting that the mutation alters GLD-1 RNA binding in *C. elegans*. Substitution of the corresponding GLD-1 glycines 248 and 250 to arginine in Sam68 abolishes self-association, suggesting that some of the GLD-1 loss-of-function phenotypes observed in *C. elegans* may be due to the absence of protein-protein interactions. However, the replacement of Qk1 glutamic acid 48 by glycine in Sam68 had no effect on Sam68 RNA binding and oligomerization (Chen et al., 1997). Therefore, to better understand Qk1 and its lethal point mutation, we characterized the properties of these proteins in vitro and in vivo. Here we report that Qk1 self-associates into

dimers via a GSG domain region predicted to form coiled coils. The introduction of the Qk1 lethal point mutation altering glutamic acid 48, located in the coiled-coil region, to a glycine (E48G; resulting protein, Qk1:E→G) abolished self-association. We also demonstrated that the expression of Qk1 and Qk1:E→G in NIH 3T3 cells induces apoptosis. These data implicate GSG domain-mediated self-association in the normal function of Qk1.

MATERIALS AND METHODS

DNA constructions. The deletion constructs encoding Qk1:1-205, Qk1:1-180, and Qk1:81-325 were generated by PCR with myc-Qk1 (Chen et al., 1997) as a DNA template. The sequences of the oligonucleotide pairs used are 5'-CTG GAA TTC GGT CGG GGA AAT GGA AAC GAA GG-3' and 5'-ATG GAA TTC TAT CTG TAG GTG CCA TTC AG-3' (for Qk1:1-205), 5'-CTG GAA TTC GGT CGG GGA AAT GGA AAC GAA GG-3' and 5'-TCA GAA TTC TAT ACC AGT AAC TTC TTC AC-3' (for Qk1:1-180), and 5'-ACC GAA TTC TCA GTT ACA AGA GAA ACT T-3' and 5'-GCT GAA TTC TAG TCC TTC ATC CAG CAA GTC-3' (for Qk1:81-325). The amplified DNA fragments were digested with *Eco*RI (the restriction sites are underlined) and subcloned into the *Eco*RI site of myc-Bluescript KS$^+$(Richard et al., 1995) and hemagglutinin (HA)-Bluescript KS$^+$(Chen et al., 1997). myc-Qk1:E→G was constructed by inverse PCR with myc-Qk1 as a DNA template and the following oligonucleotides as primers: 5'-GGA ATT AGC AGA GTA CGG AAA GAC-3' and 5'-TTC GTC CAG CAG CCG CTC GAG-3'. HA-Qk1:E→G was generated by subcloning the *Eco*RI fragment of myc-Qk1:E→G into HA-Bluescript KS$^+$. The constructs encoding HA-Qk1, myc-GLD-1, myc-GRP33, and HA-GRP33 were previously described (Chen et al., 1997). The plasmids encoding glutathione S-transferase (GST)-Qk1 and GST-Qk1:E→G fusion proteins were constructed by a two-step subcloning strategy. The *Bam*HI-*Xho*I fragment of myc-Qk1 was first inserted in the corresponding sites of pGEX-KG, generating pGST-Qk1(*Bam*HI-*Xho*I). The *Xho*I fragments of myc-Qk1 and myc-Qk1:E→G were subcloned into the *Xho*I site of pGST-Qk1(*Bam*HI-*Xho*I), resulting in pGST-Qk1 and pGST-Qk1:E→G, respectively. The GST-Qk1 deletion constructs were generated by PCR amplification with myc-Qk1 as the DNA template. For Qk1:1-80, Qk1:1-57, and Qk1:1-37, the T7 primer was used as the forward primer and the following oligonucleotides were used as the reverse primers: 5'-CTC TCT AGA CTA AAC AAT GGG TCC CAC CGC-3' (for Qk1:1-80), 5'-TAA TCT AGA CTA GTA CAT GTC TTT CCG TAC-3' (for Qk1:1-57), and 5'-GAA TCT AGA TCA GAA GTT GGG CAG GCT GCT-3' (for Qk1:1-37). The DNA

fragments encoding Qk1:18-57 and 28-57 were generated by using the reverse primer employed for Qk1:1-57 and the following forward primers: 5'-CCA GGA TCC TTG ATG CAG CTG ATG AAC-3' (for Qk1:18-57) and 5'-AAG GGA TCC ATG AGC AGC CTG CCC AAC-3' (for Qk1:28-57). The oligonucleotides used to generate Qk1:38-80 were 5'-TGC GGA TCC TTC AAC CAC CTC GAG CGG-3' and 5'-CTC TCT AGA CTA AAC AAT GGG TCC CAC CGC-3'. All of the amplified fragments were digested with *Bam*HI and *Xba*I (the underlined nucleotide sequences denote the restriction sites) and subcloned into the corresponding sites of pGEX-KG. The GST proteins were purified by affinity chromatography with glutathione beads. The purified GST proteins were covalently coupled to Affi-Gel 10 (Bio-Rad) as described previously (Richard et al., 1995). The green fluorescence protein (GFP) fusion constructs GFP-Qk1 and GFP-Qk1:E→G were generated by subcloning the *Eco*RI DNA fragments of myc-Qk1 and myc-Qk1:E→G, respectively, into vector pEGFP-C1 (Clontech). pcDNA-Qk1, pcDNA-Qk1:E→G, and pcDNA-GLD-1 were generated by subcloning the *Eco*RI fragments of myc-Qk1 and myc-Qk1:E→G and the *Xho*I fragment of myc-GLD-1, respectively, into the corresponding sites of myc-pcDNA. myc-pcDNA was constructed by digesting myc-Bluescript KS^+ (Richard et al., 1995) with *Bam*HI and *Xho*I and subcloning the DNA fragment in the corresponding sites of pcDNA1 (Invitrogen). The identities of the plasmid constructs were verified by dideoxynucleotide sequencing with Sequenase (U.S. Biochemical).

Protein expression and analysis. Proteins were expressed in HeLa cells, using the vaccinia virus T7 expression system as described previously (Richard et al., 1995). The HeLa cells were lysed in lysis buffer [1% Triton X-100, 150 mM NaCl, 20 mM Tris-HCl (pH 8.0), 50 mM NaF, 100 μM sodium vanadate, 0.01% phenylmethanesulfonyl fluoride, 1 μg of aprotinin per ml, 1 μg of leupeptin per ml], and the cellular debris and nuclei were removed by centrifugation. For immunoprecipitations, cell lysates were incubated on ice with the appropriate antibody for 1 h; then 20 μl of a 50% protein A-Sepharose slurry was added, and the mixture was incubated at 4°C for 30 min with constant end-over-end

mixing.The beads were washed twice with lysis buffer and once with phosphate-buffered saline (PBS). The samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to nitrocellulose membranes. For GST pull-down assays, 20 µl of 50% slurry containing 2 mg of GST fusion protein, covalently coupled to beads, per ml was incubated with cell lysate expressing HA-Qk1 for 1 h with constant mixing. The samples were washed and analyzed as described previously (Richard et al., 1995). Immunoblotting was performed with anti-myc 9E10 (Evan and Bishop, 1985), anti-HA, or anti-Qk1 rabbit polyclonal antibody. The anti-Qk1 antibody was generated by using GST-Qk1(81-180) as an antigen. The designated primary antibody was followed by goat anti-mouse or goat anti-rabbit antibodies conjugated to horseradish peroxidase (Organon Teknika-Cappel), and chemiluminescence (Dupont) was used for protein detection.

In situ chemical cross-linking and analysis of Qk1 dimers. HeLa cells transfected with Qk1 plasmids and rat C6 glioma cells (American Type Culture Collection) were treated in situ with 1 mM bis(maleimido)hexane (BMH) and analyzed as described previously (Chen et al., 1997). Rat astrocytes and oligodendrocytes were prepared by Guillermina Almazan (McGill University) as described elsewhere (Radhakrichna and Almazan, 1994). Dimer formation was analyzed in HeLa cells transfected with myc-Qk1 and HA-Qk1. The transfected cells were treated with BMH in situ, and the cell lysates were immunoprecipitated with anti-myc antibody or mouse immunoglobulin G (IgG). The bound proteins as well as the cell lysates were separated by SDS-PAGE, transferred to nitrocellulose membranes, and immunoblotted with anti-HA antibody.

RNase treatment and RNA binding analysis. RNase treatment was carried out by incubating cell lysates at 37°C for 1 h with RNase A (Boehringer Mannheim) at 1 mg/ml. Incubating the cell lysates at 37°C without RNase A was considered a mock treatment. RNA binding studies were carried out by incubating Qk1 immunoprecipitates with radiolabeled

total cellular RNA. To obtain radiolabeled RNA, HeLa cells were incubated overnight with 50 μCi of [^{32}P]orthophosphate (Dupont)/ml in phosphate-free Dulbecco's modified Eagle's medium (ICN). The cells were harvested, and RNA was extracted by using an RNeasy Mini Kit (Qiagen). myc-Qk1 expressed in HeLa cells was immunoprecipitated with anti-myc antibody or mouse IgG (control). The immunoprecipitates were incubated at 4°C for 30 min with ^{32}P-labeled RNA (3 x 10^6 cpm) in the presence of 2 mg of heparin/ml. The beads were washed extensively, and the bound radioactivity was counted with a liquid scintillation counter. The bound proteins were then dissociated in 1 x Laemmli sample buffer, separated by SDS-PAGE, transferred to nitrocellulose membranes, and analyzed by immunoblotting with anti-myc antibodies.

Apoptosis Assays. NIH 3T3 cells were plated 12 h before transfection, typically at a density of 10^5 cells/22-mm^2 coverslip (Fisher Scientific Co.). Cells were transfected with DNA constructs encoding GFP alone, GFP-Qk1, GFP-Qk1:E→G, pcDNA-Qk1, pcDNA-Qk1:E→G, or pcDNA-GLD-1, using LipofectAMINE PLUS reagent (Gibco BRL). At 12, 24, 36, or 48 h after transfection, the cells were fixed with 4% paraformaldehyde in PBS for 15 min and permeabilized with 1% Triton X-100 in PBS for 5 min. For immunostaining, the fixed cells were incubated with the anti-myc 9E10 antibody (1:1000) at room temperature for 1 h and subsequently with a rhodamine-conjugated goat anti-mouse secondary antibody (Jackson Laboratories; 1:300) for 30 min. The nuclei were stained with 3 μg of 4, 6-diamidino-2-phenylindole (DAPI)/ml. The morphology of transfected cells was examined by fluorescence microscopy, and cells with morphological features such as nuclear condensation and fragmentation were considered apoptotic. Apoptosis was also examined by TUNEL (terminal deoxynucleotidyl transferase-mediated fluorescein-dUTP nick end labeling). The TUNEL reagents were obtained from Boehringer Mannheim, and the assay was performed as suggested by the manufacturer.

RESULTS

The mouse *quaking* gene products exist as homodimers in vivo. The *quaking* gene encodes three different alternatively spliced transcripts that generate the Qk1-5, Qk1-6, and Qk1-7 proteins (Ebersole et al., 1996). These Qk1 proteins differ in their C-terminal 30 amino acids and are predicted to migrate with apparent molecular masses of 45 to 38 kDa on SDS-polyacrylamide gels (Ebersole et al., 1996). To characterize the endogenous Qk1 proteins, a rabbit polyclonal antibody against Qk1 amino acids 81 to 180, encompassing the KH domain, was generated. This region is identical in all Qk1 splice variants, and therefore the antibody should recognize all three Qk1 isoforms. The specificity of the anti-Qk1 antibody was examined by using HeLa cells transfected with vector alone, myc epitope-tagged Qk1-7, or myc-Sam68. The cell lysates from the transfected cells were resolved by SDS-PAGE, transferred to nitrocellulose membranes, and immunoblotted with anti-myc (Fig. 3-1A, lanes 1 to 3), normal rabbit serum (lanes 4 to 6), anti-Qk1 (lanes 7 to 9), or anti-Qk1 antibodies preabsorbed with the GST-Qk1 antigen (lanes 10 to 12). The anti-Qk1 antibody recognized the transfected Qk1 protein but not Sam68 (lanes 8 and 9). Qk1 was not observed when using normal rabbit serum or anti-Qk1 antibodies preabsorbed with the antigen (lanes 4 to 6 and 10 to 12, respectively).

To identify a cell line that contained all three Qk1 splice variants, a panel of neuronal cell lines was analyzed by immunoblotting with the anti-Qk1 antibody. Three major immunoreactive proteins in the 45- to 38-kDa range were detected in the rat C6 glioma cell line (Fig. 3-1B, lane 1) (Benda et al., 1968). Similar patterns of expression were observed for purified rat astrocytes (Fig. 3-1B, lane 2), purified rat oligodendrocytes, and whole-brain extracts from BALB/c, heterozygous, and homozygous *quaking* viable mice (data not shown). These data are consistent with the recent finding that Qk1-5, Qk1-6, and Qk1-7 are expressed in different brain cell types of normal and *quaking* viable mice (Hardy et al., 1996), and therefore extracts from whole brains should contain all three Qk1 splice variants. Our results identify the C6 glioma cell line as a suitable cell system with which to investigate the properties of the endogenous Qk1 proteins.

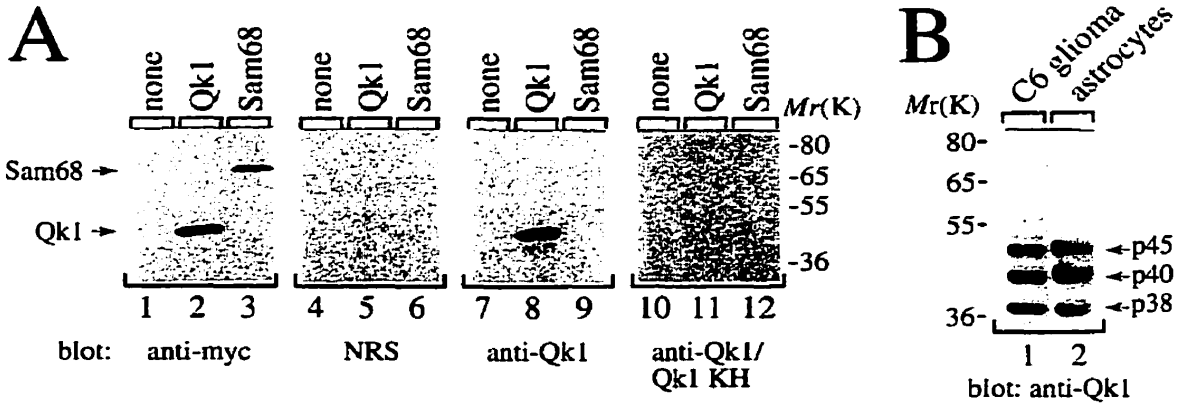


FIG. 3-1. Characterization of the anti-Qk1 antibody. (A) Total cell lysates from HeLa cells transfected with vector (none), myc-Qk1 (Qk1), or myc-Sam68 (Sam68) were separated by SDS-PAGE. The proteins were transferred to nitrocellulose and immunoblotted with anti-myc, normal rabbit serum (NRS), anti-Qk1, or anti-Qk1 antibodies preabsorbed with 1 mg of GST-Qk1KH antigen/ml (anti-Qk1/Qk1 KH). The positions of Qk1 and Sam68 are shown on the left, and those of molecular mass markers (in kDa) are on the right. (B) Total cell extracts from the rat C6 glioma cells and rat astrocytes were immunoblotted with anti-Qk1. The presence of three Qk1-immunoreactive proteins with approximate molecular masses of 45, 40, and 38 kDa is shown.

Since we have previously shown that the transfected Qk1-7 protein forms complexes in HeLa cells (Chen et al., 1997), we performed chemical cross-linking studies with C6 glioma cells to determine whether endogenous Qk1 proteins self-associated into similar complexes. C6 glioma cells were either left untreated or treated in situ with BMH, an irreversible chemical cross-linker, and the cell lysates were resolved by SDS-PAGE, transferred to nitrocellulose membranes, and immunoblotted with the anti-Qk1 antibody (Fig. 3-2A). In addition to the three monomeric Qk1 proteins (lanes 1 and 2), three distinct cross-linked Qk1 complexes with apparent molecular masses of 90 to 80 kDa were observed after cross-linking (lane 2). These data suggest that Qk1 proteins exist as dimers.

To verify that Qk1 cross-linked complexes represented Qk1 dimers, HeLa cells coexpressing myc- and HA-Qk1 were treated in situ with BMH, and the cell lysates were immunoprecipitated with IgG (Fig. 3-2B, lane 2) or anti-myc antibody (lane 3). The immunoprecipitated proteins as well as the total cell lysate (lane 1) were analyzed by immunoblotting with anti-HA antibodies. BMH treatment of HeLa cells expressing Qk1 resulted in a 90-kDa complex, in addition to the 45-kDa Qk1 monomer (lane 1). This HA-Qk1-containing complex was present in anti-myc immunoprecipitates (lane 3), demonstrating that HA- and myc-Qk1 proteins dimerized. The presence of HA-Qk1 monomers in anti-myc immunoprecipitates indicated that not all complexes were chemically cross-linked (lane 3).

Since the mouse Qk1-7 protein is an RNA-binding protein (Chen et al., 1997), we sought to investigate whether cellular RNA was required for Qk1 self-association. myc- and HA-Qk1 were expressed in HeLa cells separately, the cell lysates were or were not treated with RNase A for 1 h at 37°C and then mixed and immunoprecipitated with anti-myc antibodies. The bound proteins were separated by SDS-PAGE, transferred to nitrocellulose, and immunoblotted with anti-HA antibodies. HA-Qk1 coprecipitated with myc-Qk1 regardless of whether the lysates were treated with RNase A (Fig. 3-2C, lanes 1 and 2), indicating that RNase treatment had no effect on the ability of Qk1 to self-associate. Under similar conditions, we also tested the ability of GRP33 to self-associate in the presence or

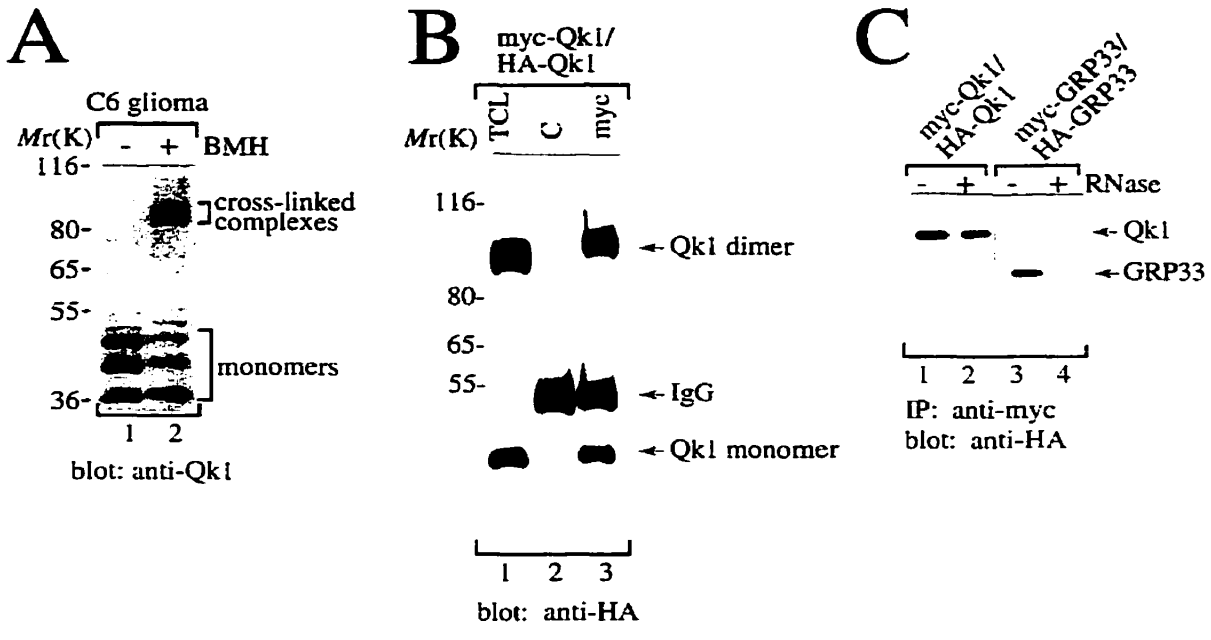


FIG. 3-2. Qk1 self-associates into homodimers in the absence of cellular RNA. (A) C6 glioma cells were treated in situ with (+) or without (-) BMH. The cells were lysed in sample buffer, and the proteins were separated by SDS-PAGE. The proteins were transferred to nitrocellulose and immunoblotted with rabbit anti-Qk1 antibody. The bands representing the three Qk1 isoforms are shown as monomers, and the cross-linked complexes are indicated. The positions of the molecular mass markers (in kDa) are indicated on the left. (B) HeLa cells cotransfected with myc- and HA-Qk1 were treated in situ with BMH. The cells were lysed, and an aliquot of the total cell lysate (TCL) as well as anti-myc (myc) and IgG (C) immunoprecipitates were separated by SDS-PAGE. The proteins were transferred to nitrocellulose and immunoblotted with anti-HA antibodies. (C) Qk1, unlike GRP33, does not require cellular RNA for self-association. HeLa cells expressing myc-Qk1, HA-Qk1, myc-GRP33, or HA-GRP33 were lysed separately, and each cell lysate was divided into two portions, either treated with RNase A(+) or not treated (-), mixed, and incubated with anti-myc antibodies. The anti-myc immunoprecipitates (IP) of the mixed lysates were separated by SDS-PAGE and immunoblotted with anti-HA antibodies.

absence of RNase. HA-GRP33 was observed in anti-myc immunoprecipitates when no RNase treatment was performed (lane 3), but it was not seen when RNase treatment was performed (lane 4). These findings show that unlike GRP33 and Sam68 (Chen et al., 1997), Qk1 does not require RNA for self-association.

The self-association and RNA binding properties of Qk1 map to the GSG domain. The Qk1 GSG domain spans amino acids 9 to 205, and the embedded KH domain spans amino acids 81 to 180 (Fig. 3-3A). A deletion analysis was performed to identify whether the GSG domain and the Qk1 C-terminal region were necessary and sufficient for Qk1 self-association and RNA binding. The truncated Qk1 proteins were tested for their ability to associate with HA-Qk1. HeLa cells were cotransfected with DNAs encoding HA-Qk1 and with wild-type myc-Qk1 or the truncated myc-Qk1 constructs (Fig. 3-3A). The cells were lysed and immunoprecipitated with control or anti-myc antibodies. The bound proteins were immunoblotted with anti-HA antibodies for detection of the presence of HA-Qk1. myc immunoprecipitates of wild-type myc-Qk1 (Fig. 3-3B, lane 3), myc-Qk1:1-205 (lane 6), and myc-Qk1:1-180 (lane 9) contained abundant levels of HA-Qk1. However, Qk1 lacking its N-terminal 80 amino acids, or the QUA1 region (myc-Qk1:81-325), did not coprecipitate HA-Qk1 (Fig. 3-3B, lane 12). These data demonstrate that the C-terminal 145 amino acids of Qk1 are dispensable and that the QUA1 region of the GSG domain is required for self-association. The levels of expression of the myc-Qk1 constructs were equivalent (Fig. 3-3B, lanes 13 to 16).

The RNA binding abilities of wild-type Qk1 and its truncated forms were compared. Wild-type and the various truncated myc-Qk1 proteins were expressed in HeLa cells, and the anti-myc and control IgG immunoprecipitates were incubated with ^{32}P-labeled RNA in lysis buffer supplemented with 2 mg of heparin/ml. The radioactivity in the bound RNA was counted and expressed in counts per minute (Fig. 3-3C). Anti-myc immunoprecipitates of wild-type Qk1 bound 20 times more labeled RNA than did control immunoprecipitates (Fig. 3-3C, Qk1). The deletion of the C-terminal 120 amino acids had no effect on Qk1

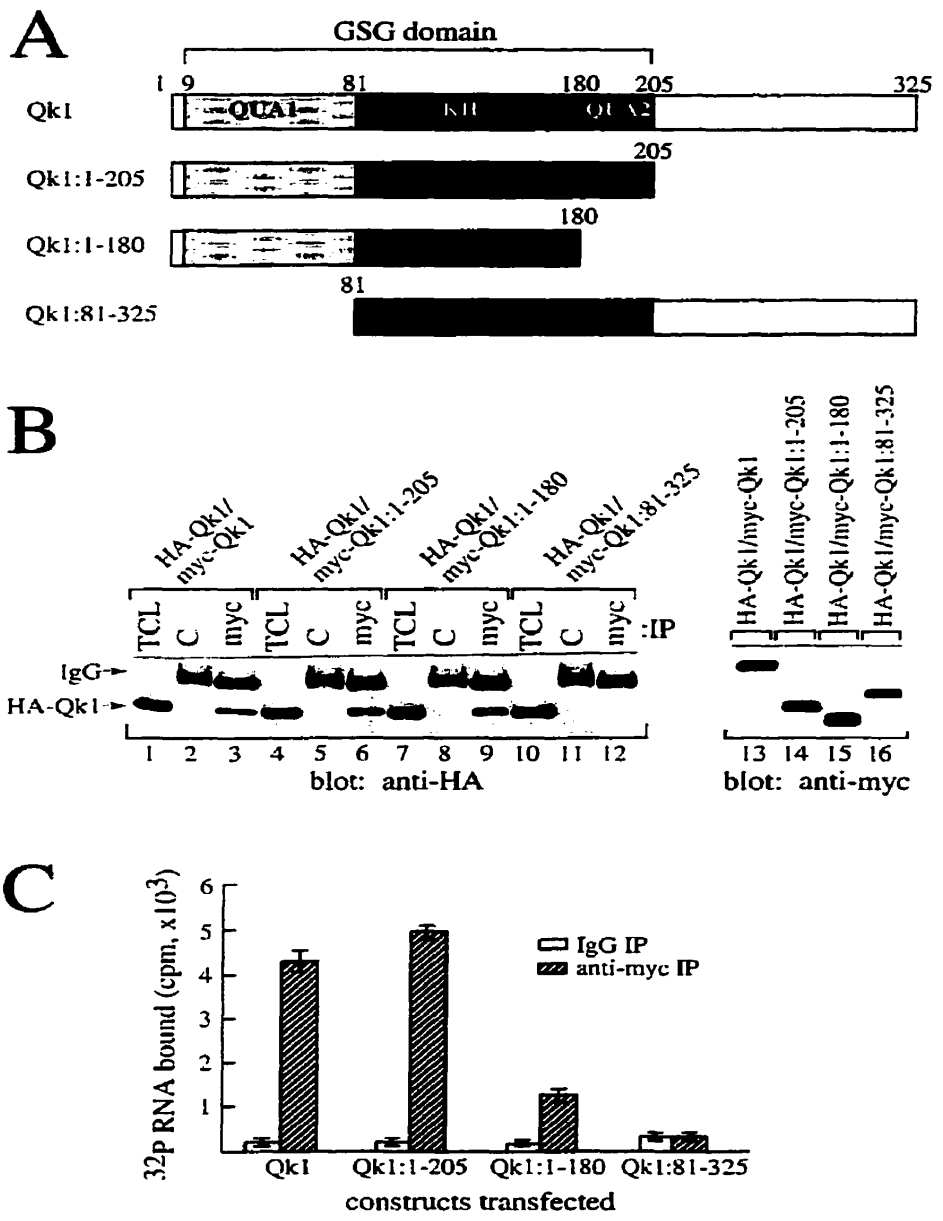


FIG. 3-3. Mapping of the Qk1 self-association and the RNA binding regions within the GSG domain. (A) Schematic diagrams of the Qk1 constructs utilized are shown. The black box denotes the KH domain, whereas the gray boxes represent the QUA1 and QUA2 regions as indicated. The entire region encompassing QUA1, KH, and QUA2 is the GSG domain. (B) Truncation of the N-terminal 80 amino acids of Qk1 or QUA1 abolishes self-association. HA-Qk1 was cotransfected in HeLa cells with various myc-Qk1 deletion constructs as indicated. Total cell lysates (TCL) as well as anti-myc (myc) and control IgG (C) immunoprecipitates were analyzed by immunoblotting with anti-HA antibodies. Total cell lysates of the myc-Qk1 proteins were immunoblotted with anti-myc antibodies (lanes 13 to 16). (C) The Qk1 GSG domain is required for RNA binding. HeLa cell lysates containing myc-tagged Qk1 or various truncated forms of Qk1 were immunoprecipitated with anti-myc antibody (hatched bars) or control IgG (white bars) and then incubated with ^{32}P-labeled total cellular RNA. Each bar represents the mean $\pm$ standard deviation of data from more than six independent immunoprecipitations carried out during more than three separate experiments.

RNA binding activity (Fig. 3-3C, Qk1:1-205). However, the deletion of the QUA2 region in the Qk1 GSG domain reduced the bound RNA by more than half (Fig. 3-3C, Qk1:1-180). Furthermore, the deletion of the QUA1 region of the Qk1 GSG domain completely abolished RNA binding (Fig. 3-3C, Qk1:81-325). These data suggest that the entire Qk1 GSG domain is required for optimal RNA binding.

The genetic mutation (E48G) identified in *quaking* lethal mice abolishes self-association but not RNA binding. Our deletion studies indicated that the QUA1 region of the Qk1 GSG domain is required for Qk1 self-association and RNA binding. Interestingly, one ethylnitrosourea-induced point mutation induced that causes a mouse *quaking* lethal phenotype (Justice and Bode, 1988) has been identified in this region (Ebersole et al., 1996). This amino acid substitution, altering glutamic acid 48 to a glycine, was introduced in the mouse *quaking* protein (Qk1:E→G) and tested for its effect on self-association and RNA binding. The abilities of Qk1:E→G to associate with Qk1 and to self-associate were examined (Fig. 3-4A). HeLa cells were transfected with combinations of myc- and HA-Qk1 or Qk1:E→G as indicated in Fig. 3-4A. The cells were lysed and immunoprecipitated with control or anti-myc antibodies. The proteins were separated by SDS-PAGE and analyzed by immunoblotting with anti-HA antibodies. HA-Qk1 coimmunoprecipitated with myc-Qk1 (Fig. 3-4A, lane 3) but not with myc-Qk1:E→G (lane 6). Moreover, HA-Qk1:E→G did not coprecipitate with myc-Qk1 (lane 9) or myc-Qk1:E→G (lane 12), suggesting that the E48G mutation prevents Qk1 self-association. The levels of expression of the myc epitope-tagged constructs used were equivalent (Fig. 3-4B).

The RNA binding properties of Qk1:E→G and Qk1 were investigated. The Qk1:E→G protein bound RNA to the same extent as the wild-type Qk1 protein (Fig. 3-4C), demonstrating that this point mutation has no effect on RNA binding. However, we cannot exclude the possibility that a difference in RNA binding will be observed once a high-affinity RNA target for Qk1 is identified.

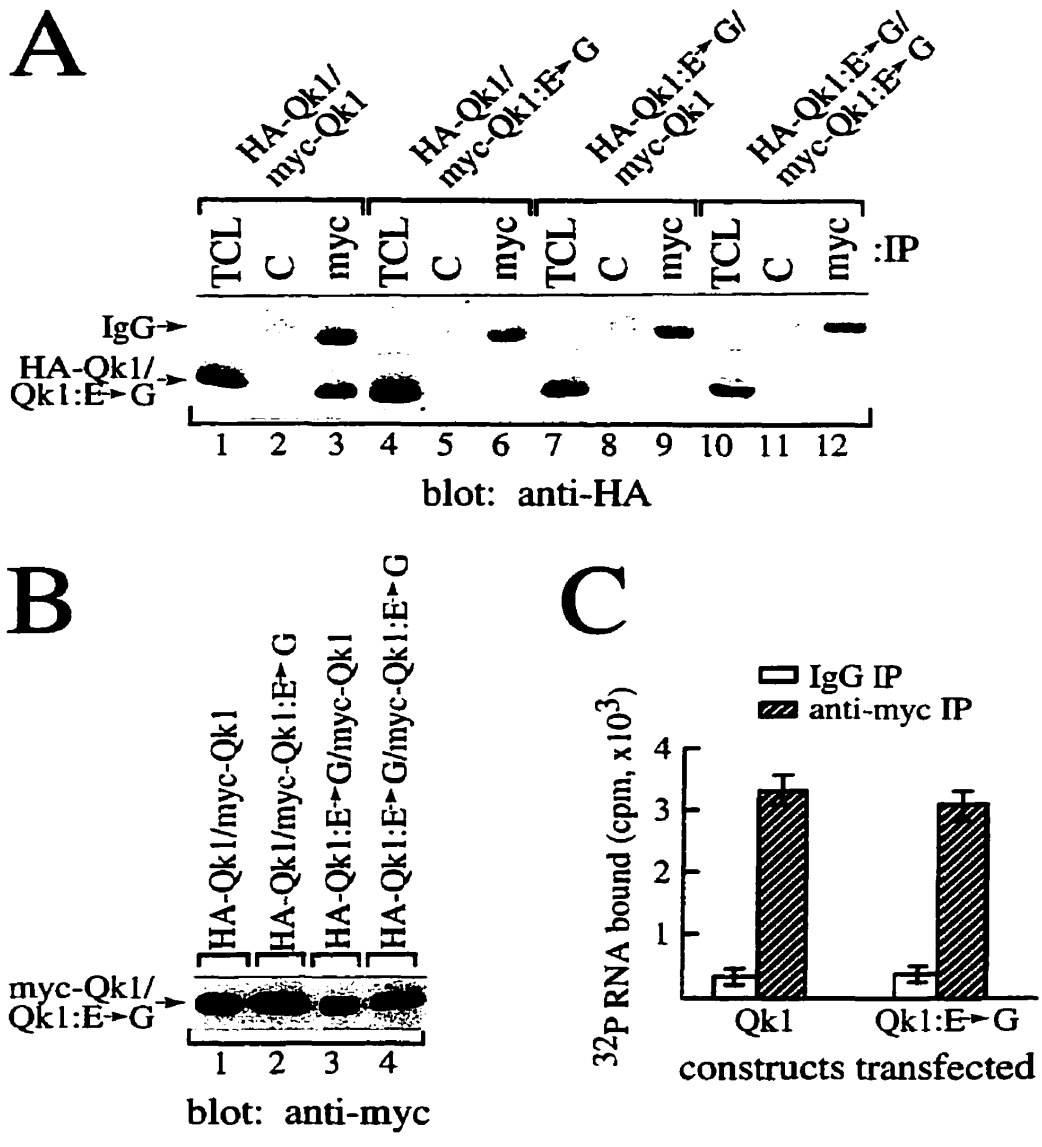


FIG. 3-4. E48G substitution in Qk1 abolishes homodimerization but not RNA binding. (A) HeLa cells were transfected with HA- or myc-tagged Qk1 and/or Qk1:E→G as indicated. The total cell lysate (TCL) as well as anti-myc (myc) and IgG (C) immunoprecipitates were immunoblotted with anti-HA antibody. The bands representing HA-Qk1 or HA-Qk1:E→G are indicated on the left. The migration of the heavy chain of IgG is also indicated. (B) Total cell lysates corresponding to those of panel A were immunoblotted with anti-myc antibodies. (C) Immunoprecipitated Qk1 or Qk1:E→G was incubated with labeled RNA as described in Materials and Methods.

We next examined the ability of wild-type and mutated Qk1 to self-associate in vitro. GST pull-down assays using bacterial fusion proteins were performed (Fig. 3-5A). HeLa cell lysates containing HA-Qk1 or HA-Qk1:E→G were incubated with affinity matrices coupled with GST alone (lanes 2 and 6), GST-Qk1 (lanes 3 and 7), or GST-Qk1:E→G (lanes 4 and 8). The proteins that bound the GST proteins were separated by SDS-PAGE and analyzed with anti-HA antibodies. HA-Qk1 bound the GST-Qk1 fusion protein (lane 3), indicating that Qk1 was able to self-associate in vitro. However, HA-Qk1 did not associate with GST-Qk1:E→G (lane 4), and no interaction between HA-Qk1:E→G and either GST-Qk1 (lane 7) or GST-Qk1:E→G (lane 8) was observed. These results are consistent with our coimmunoprecipitation data, confirming that the *quaking* lethal point mutation abolishes Qk1 self-association. These observations demonstrate that Qk1 containing the *quaking* lethal point mutation is defective in GSG-mediated protein-protein interactions, such as the ability to self-associate.

The QUA1 region of the GSG domain mediates Qk1 self-association. The severe effect of the *quaking* point mutation on Qk1 self-association suggested that the QUA1 region of the Qk1 GSG domain is directly involved in Qk1 self-association. To further delineate the region responsible for self-association, a series of GST fusion proteins were generated. These fusion proteins were utilized in GST pull down assays and tested for their ability to interact with HA-Qk1 produced in HeLa cells (Fig. 3-5B). The GST-Qk1:1-80 and GST-Qk1:1-57 fusion proteins bound HA-Qk1 (lane 3 and 4), but GST-Qk1:1-37 did not (lane 8). Amino terminal deletions revealed that Qk1:18-57 was the minimum region capable of binding HA-Qk1 (lanes 5 to 7).

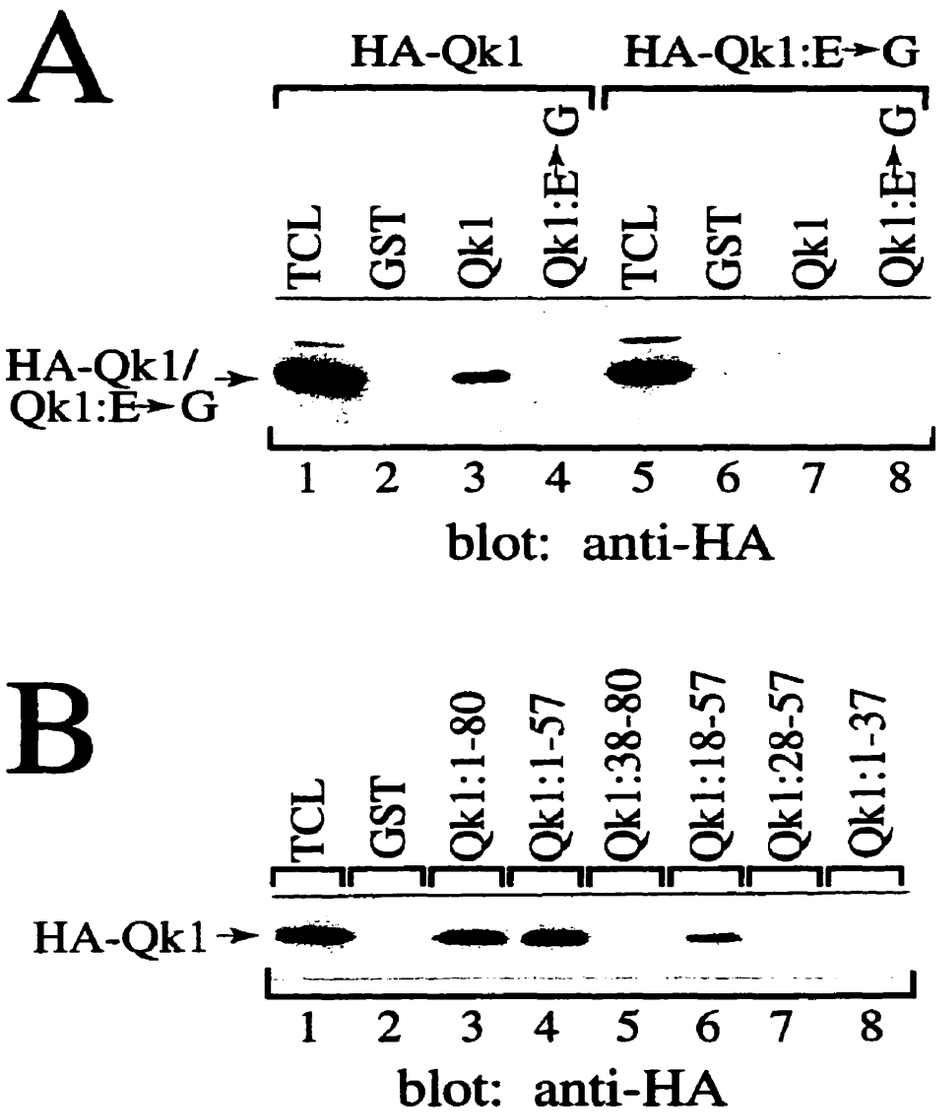


FIG. 3-5. Association of Qk1 in vitro and localization of the minimal region to amino acids 18 to 57. (A) HA-Qk1 or HA-Qk1:E→G was transfected into HeLa cells. The cells were lysed, and GST pull-down assays were performed with full-length Qk1 or Qk1:E→G expressed as a GST fusion protein. The bound proteins were separated by SDS-PAGE, transferred to nitrocellulose, and immunoblotted with anti-HA antibodies. (B) HA-Qk1 was expressed in HeLa cells and incubated with various Qk1 GST fusion proteins as indicated. The bound HA-Qk1 was analyzed as described for panel A. The migration of HA-Qk1 or HA-Qk1:E→G is shown on the left. TCL, total cell lysate.

Qk1 and Qk1:E48→G induce apoptosis in fibroblasts. The phenotype of the *quaking* lethal mutation is arrested growth of an embryo with generalized abnormalities (Justice and Bode, 1988), suggesting that Qk1 may play a role in cell proliferation or differentiation. To determine the role of Qk1 and its lethal mutation, we expressed wild-type Qk1 and Qk1 with the E48G mutation in mouse fibroblast cells and examined their effects on cell growth. NIH 3T3 cells were transfected with expression vectors encoding GFP alone, GFP-Qk1, or GFP-Qk1:E→G. Twelve hours after transfection, approximately 25% of the cells expressed GFP, GFP-Qk1 or GFP-Qk1:E→G, as visualized by fluorescence microscopy. Interestingly, only 6 to 8 % of the cells expressed GFP-Qk1 and GFP-Qk1:E→G at 36 h , suggesting that cells transfected with wild-type or mutant Qk1 were not surviving (data not shown). The cells expressing GFP alone appeared normal and healthy (Fig. 3-6A, left panels). Cells expressing GFP-Qk1 or Qk1:E→G exhibited morphological changes characteristic of apoptosis, including cell shrinkage, cytoplasm condensation, and membrane blebbing (Fig. 3-6A, middle and right panels). GFP-transfected cells displayed normal nuclear morphology as visualized by DAPI staining (GFP panels, lower halves). GFP-Qk1- or GFP-Qk1:E→G-transfected cells had irregular (GFP-Qk1 and GFP-Qk1:E→G panels, 12 h, lower halves), condensed, or fragmented nuclei (36 h), consistent with apoptotic cell death. To confirm that the morphological changes induced by GFP-Qk1 and GFP-Qk1:E→G were indeed associated with apoptosis, we performed the TUNEL assay with fluoresceinated nucleotides, NIH 3T3 cells were transfected with myc-Qk1, myc-Qk1:E→G, or myc-GLD-1 for 36 h. The cells were fixed, and the myc epitope-tagged proteins were detected by indirect immunofluorescence with a rhodamine-conjugated secondary antibody.(Fig. 3-6B, top panels). As observed by TUNEL assay, most of the myc-Qk1- and myc-Qk1:E→G-transfected cells fluoresced green, consistent with apoptotic cell death. The transfection of another cytoplasmic GSG protein, myc-GLD-1 (Jones et al., 1996), did not induce apoptosis and served as a negative control (Fig. 3-6B, left panels). All of the cells that stained in the TUNEL assay contained condensed or fragmented nuclei as visualized by DAPI staining (Fig. 3-6B). Therefore, the presence of nuclear condensation

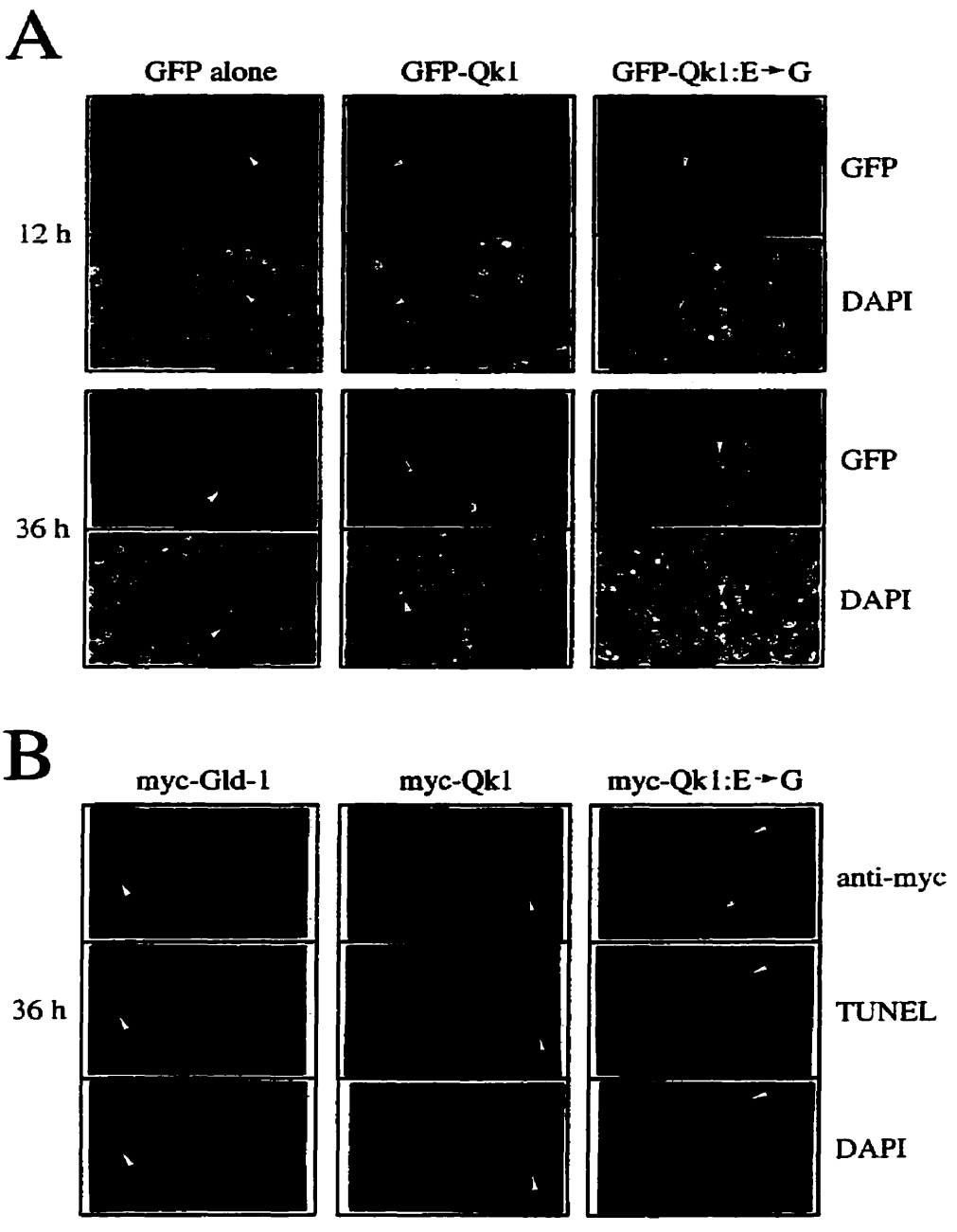


Fig. 3-6. Qk1 and Qk1:E→G induce apoptosis in NIH 3T3 cells. (A) NIH 3T3 cells were transfected with an expression vector encoding GFP alone, GFP-Qk1, or GFP-Qk1:E→G. After 12, 24, 36, and 48 h, the cells were fixed and stained with DAPI to visualize the nuclei. The top photograph in each pair shows the fluoresceinated cells containing GFP, and the lower photograph shows the DAPI-stained nuclei. The white arrowheads were used to align the top and bottom photographs. (B) NIH 3T3 cells were transfected with expression vector encoding myc-GLD-1, myc-Qk1, or myc-Qk1:E→G. The myc epitope-tagged proteins were visualized by indirect immunofluorescence with a rhodamine-conjugated secondary antibody (anti-myc). The apoptotic cells were visualized by TUNEL with fluorescein-containing nucleotides (TUNEL), and the nuclei were stained with DAPI (DAPI). The three photographs each for myc-GLD-1, myc-Qk1, and myc-Qk1:E→G represent the same field of cells as visualized with different filters. The white arrowheads were used to orient the cells in the photographs.

and fragmentation, as detected by DAPI staining, is a good indication of apoptotic cell death.

The levels of apoptosis induced by Qk1 and the Qk1:E→G proteins were 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-Qk1, or GFP-Qk1:E→G. A small fraction of GFP expressing cells were apoptotic, and this fraction (~15%) remained steady up to 48 h (Fig. 3-7). The transfection of Qk1 or Qk1:E→G resulted in a significant increase in the number of apoptotic cells with time. At 48 h posttransfection, ~90% of the remaining cells were apoptotic (Fig. 3-7). Qk1:E→G consistently resulted in a larger fraction of apoptotic cells upon transfection than did Qk1. This difference was more prominent at the early time points. At 12 and 24 h, 36.7% and 68.6% of the Qk1:E→G-transfected cells were apoptotic. These values are in contrast to 24.4% and 49.0%, respectively, for Qk1-transfected cells, and these differences were statistically significant as calculated by the χ^2 test ($P<0.01$). Since the transfection efficiencies and levels of expression of GFP-Qk1 and GFP-Qk1:E→G were similar (data not shown), these results suggested that Qk1:E→G is more potent than Qk1. The majority, ~70%, of the Qk1:E→G-transfected cells were apoptotic at 24 h, whereas it took GFP-Qk1 36 h to reach a similar level of apoptosis. These data suggest that Qk1 induces apoptosis and that the E48G mutation in Qk1 contributes to aggravated apoptotic cell death.

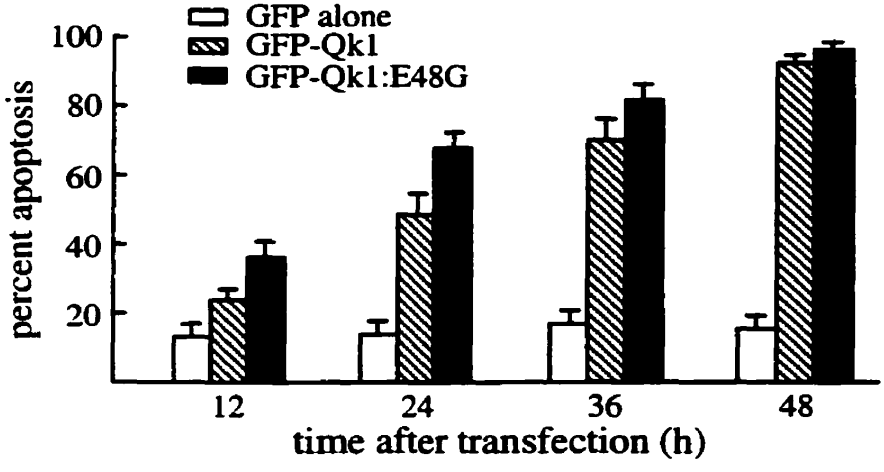


FIG. 3-7. Quantitation of the apoptosis induced by Qk1 and Qk1:E48G. Eight hundred transfected NIH 3T3 cells, in three separate experiments, were counted and assessed for the presence of apoptotic nuclei. The presence of apoptotic nuclei was scored as cells undergoing apoptosis. The white, hatched, and black bars represent GFP, GFP-Qk1, and GFP-Qk1:E48G, respectively.

DISCUSSION

We have shown by coimmunoprecipitation and in situ chemical cross-linking studies that the Qk1 proteins exist as homodimers. The minimum region required for self-association consists of amino acids 18 to 57, which are located in the QUA1 region of the Qk1 GSG domain. This region contains several conserved residues, including glutamic acid at position 48. Alteration of glutamic acid 48 to glycine is thought to be the cause of the lethality in the qk^{kt4} mice (Ebersole et al., 1996). We demonstrated that the Qk1 E48G substitution abolishes self-association. Analysis of the Qk1 protein sequence with the computer program COILS (Lupas et al., 1991) predicted that amino acids 38 to 57 have a high propensity to form coiled coils, which would be disrupted with the introduction of a glycine at position 48 (Fig. 3-8, coiled coil no. 1). Thus, it is likely that Qk1 dimerizes through coiled-coil interactions mediated by the GSG domain. These data suggest that the failure of Qk1 to dimerize causes embryonic lethality in the qk^{kt4} mice and implicate dimer formation in the normal function of Qk1 proteins. It is possible that Qk1 associates with other proteins via this region, and therefore we cannot exclude the possibility that the lethality results because Qk1 fails to mediate interactions with other proteins.

Mouse Qk1 amino acids 38 to 57, predicted to form coiled coil no. 1, are conserved in Who/How and Xqua, with 13 of 20 and 19 of 20 identical residues, respectively (Ebersole et al., 1996; Baehrecke, 1997; Fyrberg et al., 1997; Zaffran et al., 1997; Zorn et al., 1997). This region in Xqua has a coiled coil prediction similar to that of mouse Qk1 (data not shown), and since Xqua has been shown to self-associate in vitro (Zorn and Krieg, 1997), we predict that this region mediates the self-association. This region in Who/How is also predicted to form coiled coils, but with a lower propensity (0.15). Both Xqua and Who/How coiled coils in this region would be disrupted by the introduction of the corresponding mouse Qk1 E48G mutation. Based on our Qk1 data and the computer analyses, we predict that the point mutation identified in Who/How, altering arginine 185 to cysteine (Baehrecke, 1997), should not alter self-association or RNA binding. Indeed, the introduction of R185C in *Drosophila*Who/How did not alter self-association or RNA binding (Di Fruscio and

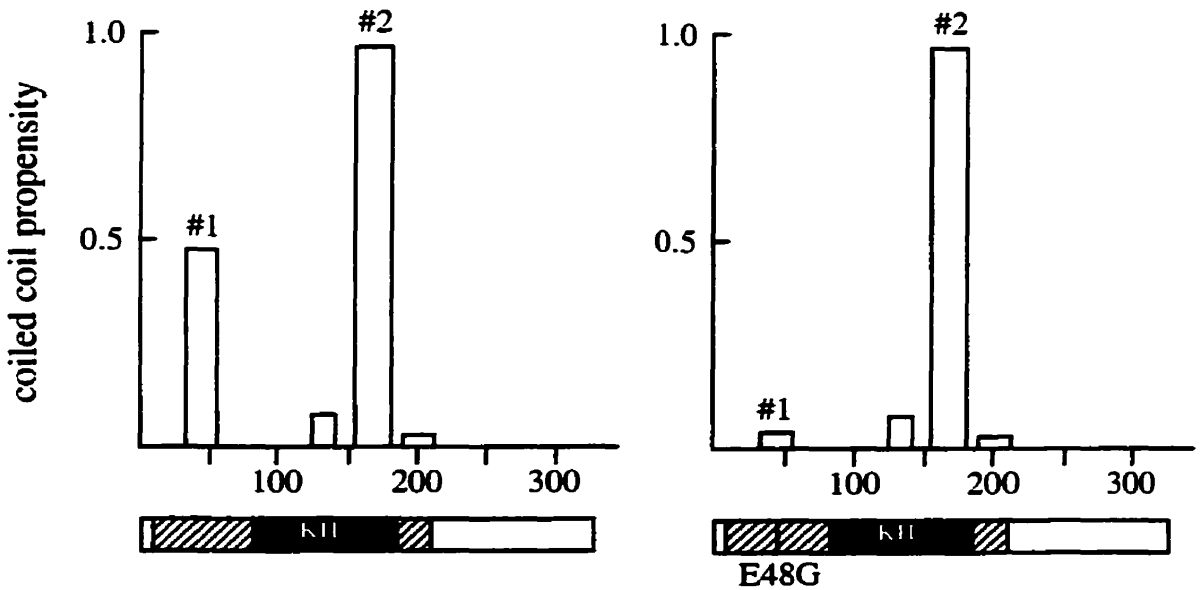


FIG. 3-8. Coiled-coil predictions for Qk1 (left) and Qk1:E48G (right). The Qk1 and Qk1:E48G protein sequences were analyzed with the computer program COILS, and the putative coiled-coil motifs are shown (#1 and #2). The abscissa and ordinate represent amino acid numbers and the propensity to form coiled coils, respectively. The structures of the Qk1 proteins are shown below, with hatched and black boxes representing the GSG and KH domains, respectively. The vertical line in the GSG domain denotes the location of the *quaking* lethal point mutation E48G.

Richard, unpublished data).

The mouse Qk1 sequence is also predicted to form a second coiled coil, at the C terminus of the KH domain (Fig. 3-8, coiled coil no. 2). Although our data suggest that coiled coil no. 2 is not sufficient for Qk1 self-association (Qk1:81-325), it is most likely involved in protein-protein interactions other than self-association. Coiled-coil interactions have been predicted for KH domain proteins FMR1 and FXRs (Siomi et al., 1996). These proteins have been observed to self-associate as well as associate with ribosomes (Khandjian et al., 1996; Siomi et al., 1996). These interactions were demonstrated to occur outside the KH domains, by regions predicted to form coiled coils (Siomi et al., 1996). Interestingly, the KH domains of the FMR1 protein are not predicted to form coiled coils like the KH domain of Qk1. Analysis of the sequences of GSG proteins SF1, Who/How, GLD-1, and Xqua demonstrated that they are all predicted to form coiled coil no. 2 at the C terminus of the KH domain (data not shown). This may represent one more difference between the KH domains of GSG proteins and other KH domains. Sam68 and GRP33 are not predicted to form such coiled coils and may represent a different subclass of GSG proteins. To this end, both GRP33 and Sam68 require RNA for self-association (Fig. 3-2C) (Chen et al., 1997). For Sam68, we have been unable to reconstitute the self-association in vitro with recombinant proteins (data not shown) as we have done with Qk1. Another difference is the minimum region required for Sam68 self-association, which is amino acids 103 to 269, the entire GSG domain, and within this region the KH domain loops 1 and 4 are essential (Chen et al., 1997).

We have mapped the Qk1 RNA binding domain to the N-terminal 205 amino acids encompassing the entire GSG domain. This demonstrates that the entire Qk1 GSG domain is sufficient for RNA binding, as has been previously demonstrating for Sam68 (Chen et al., 1997; Lin et al., 1997). However, this is in contrast to what has been reported for Xqua by Zorn and Krieg; they demonstrated that the entire Xqua protein was required for optimal RNA binding (Zorn and Krieg, 1997). Deletion of the N-terminal GSG amino acids had only a slight reduction effect on RNA binding, whereas deletion of amino acids C terminal

to the GSG domain (231 to 357) abolished RNA binding. Since the amino acid sequence identity between mouse, human and *Xenopus* Qk1 proteins is 94% (Zorn et al., 1997), Xqua and Qk1 are predicted to have similar RNA binding properties. Either the difference is intrinsic to the proteins or is due to the different stringencies of the RNA binding assays.

Although the exact function of the GSG domain is unknown, our observations with Qk1 and Sam68 indicate that the GSG domain is involved in oligomerization and RNA binding. It is presently unclear whether the GSG domain mediates protein-protein interactions with non-GSG family members. The presence of signaling protein motifs in GSG proteins suggests a role for these proteins in signal transduction [for a review, see (Vernet and Artzt, 1997)]. Interestingly, these potential SH2, SH3, and WW domain binding motifs lie outside the GSG domain, ruling out a direct role of the GSG domain in signal transduction. Nevertheless, the presence of signaling motifs and the presence of phosphorylation sites C terminal to the Sam68 GSG domain (Wong et al., 1992; Richard et al., 1995; Taylor et al., 1995) suggests that signal transduction pathways may regulate GSG-mediated interactions. Indeed, we have demonstrated in previous studies that Sam68 RNA binding and oligomerization are abolished by the p59^{fyn} tyrosine kinase (Wang et al., 1995; Chen et al., 1997). Moreover, the binding of Sam68 to SH3 domains inhibits RNA binding in vitro (Taylor et al., 1995). These data demonstrate that Sam68 has the potential to link signal transduction pathways with RNA metabolism. Similar data for the other GSG proteins has yet to be obtained. The absence of known signaling motifs in some GSG family members, such as GLD-1 (Jones and Schedl, 1995), demonstrates that not all family members have the potential to act as signal transduction activators of RNA metabolism (Vernet and Artzt, 1997), but may have other functions. It is possible that the only common properties shared by all family members are GSG domain-mediated RNA binding and self-association.

The *quaking* genes from *X. laevis*, *D. melanogaster*, and mice are involved in a variety of processes, such as myelination, embryogenesis, muscle development, and notochord development (Sidman et al., 1964; Bode, 1984; Hogan and Greenfield, 1984;

Justice and Bode, 1988; Shedlovsky et al., 1988; Baehrecke, 1997; Fyrberg et al., 1997; Zaffran et al., 1997; Zorn and Krieg, 1997). The pleiotropic roles and the high level of conservation of this gene suggest a general function for Qk1 in cellular processes. Our results suggest that a function of Qk1 might be to act as a regulator or effector of apoptosis. Since fibroblasts do not express Qk1 (Fig. 3-1, lanes 7 to 9), it is likely that the expression of Qk1 leads to perturbation of normal cellular processes. Qk1-7, the isoform used in our experiments, is predominantly cytoplasmic (Fig. 3-6 and data not shown) (Hardy et al., 1996), and it may induce apoptosis by regulating or interfering with the translation and/or mRNA stability of apoptotic or survival proteins. There is a precedent for these mechanisms, since an increase in RNA degradation before the onset of apoptosis has been observed in T cells (Mondino and Jenkins, 1995; Taupin et al., 1995) and hnRNP K has been shown to regulate translation (Ostareck et al., 1997).

The three mouse Qk1 splice variants have identical GSG domains and differ in their C termini (Ebersole et al., 1996). Qk1-5 is mainly expressed in the nucleus, whereas Qk1-6 and Qk1-7 are mainly expressed in the cytoplasm, suggesting that the last 30 amino acids determine the localization (Hardy et al., 1996). The identicalness of the GSG domains suggest that all three Qk1 splice variants are able to associate with RNA, homodimerize, and heterodimerize. By performing cross-linking studies in C6 glioma cells, we have demonstrated the presence of at least three cross-linked Qk1 species that may represent homodimers and/or heterodimers. The presence of multiple Qk1 splice variants in glial cells and oligodendrocytes (Hardy et al., 1996) suggests an interesting mechanism for the regulation of Qk1 cellular localization. Heterodimers of Qk1-5:Qk1-6 or Qk1-5:Qk1-7 may cause the retention of Qk1-5 in the cytoplasm. Alternatively, Qk1-6 and Qk1-7 might be dragged into the nucleus as Qk1-5 heterodimers. We speculate that the formation and balance of such dimers are crucial for Qk1 function and are responsible for the phenotypes observed in the *quaking* viable and lethal mice.

The genetic lesion in the *quaking* viable mouse has been mapped to the *qk1* promoter-enhancer region (Ebersole et al., 1996), and as a result, Qk1-6 and Qk1-7 are not

expressed in oligodendrocytes (Hardy et al., 1996). Oligodendrocytes still express nuclear Qk1-5. According to our data, Qk1-5 should be unable to form heterodimers with Qk1-6 or Qk1-7 in the oligodendrocytes. This might interfere with Qk1 function and lead to the myelin dysregulation observed in the central nervous system of these animals. The *quaking* viable mice have been extensively studied (Hogan and Greenfield, 1984), and several defects in RNA metabolism have been observed. Alterations in the levels of alternatively spliced RNAs and in the processing and/or turnover of the mRNA transcripts encoding myelin-associated glycoprotein, myelin basic protein, and proteolipid protein have been demonstrated (Frail and Braun, 1985; Bartoszewicz et al., 1995; Bo et al., 1995; Hardy et al., 1996). A defect in myelin basic protein mRNA transport has also been observed in the *quaking* viable mice (Barbarese, 1991). The challenge will be to determine whether Qk1 regulates splicing, RNA transport, mRNA stability, and/or translation. Since hnRNP K acts as a transcription factor (Michelotti et al., 1996) and Sam68 associates with double-stranded DNA (Wong et al., 1992), the possibility that nuclear Qk1 also functions as a transcription factor cannot be excluded.

The ethylnitrosourea-induced *quaking* alleles are known to be lethal at around day 9 or 10 of gestation (Bode, 1984; Justice and Bode, 1988; Shedlovsky et al., 1988). The only Qk1 isoform expressed in significant amounts at this early time is Qk1-5 (Ebersole et al., 1996). Our data suggest that this point mutation would be unable to homodimerize, thus possibly altering its function during embryogenesis. Interestingly, the Qk1:E→G protein was significantly more potent than wild-type Qk1 at inducing apoptosis in NIH 3T3 cells. Since the Qk1 E48G point mutation is lethal in mice, it is tempting to speculate that unregulated apoptotic cell death occurs due to the absence of GSG-mediated dimerization. The apoptosis we observe with Qk1 in NIH 3T3 cells may be similar to the poisoning effects observed with certain GLD-1 point mutations in *C. elegans* (Jones and Schedl, 1995).

By using a pan-Qk1 antibody, the C6 glioma cell line was identified to contain all three Qk1 isoforms. This cell line, which is of rat origin and is derived from glial cells (Benda et

al., 1968), expresses several oligodendrocytic markers, such as myelin-associated glycoprotein, proteolipid protein, and 2',3'-cyclic nucleotide 3'-phosphohydrolase (Zhu et al., 1994; Goya et al., 1996). Since the C6 glioma cell line expresses all three Qk1 isoforms, it should provide a cell system in which to study the properties of the Qk1 dimers and some of their biochemical functions.The expression of Qk1 and Qk1:E→G in these cells did not readily induce apoptosis (data not shown), suggesting that either the endogenous Qk1 proteins provide a protective effect in these cells or C6 glioma cells are not a suitable system in which to study apoptosis.

In conclusion, we have defined the Qk1 GSG domain as the region required for dimerization and RNA binding. Replacement of glutamic acid 48 with a glycine, a mutation known to be lethal in mice, abolished Qk1 self-association but not RNA binding. The expression of Qk1 and Qk1:E→G in fibroblast cells induced apoptotic cell death. Since Qk1 has signaling motifs (Ebersole et al., 1996), it will be essential to examine the potential role of signaling molecules in the regulation of Qk1 RNA binding, self-association, and apoptosis.

ACKNOWLEDGMENTS

We thank Guillermina Almazan for purified extracts of rat astrocytes and oligodendrocytes. We thank Janet Henderson and Antonis Koromilas for critically reading the manuscript and helpful comments. We are grateful to Rongtuan Lin, John Th'ng, and Hans Zingg for providing reagents and to Bassam Damaj for technical assistance with the rabbit polyclonal antibody.

T.C. is supported by a studentship from the Cancer Research Society of Canada and funds from Canderel. This work was supported by grants from the Medical Research Council of Canada, the Cancer Research Society of Canada, Fonds de la Recherche en Santé du Québec, and the Multiple Sclerosis Society of Canada. S.R. is a Scholar of the Medical Research Council of Canada.

REFERENCES

Arning, S., Gruter, P., Bilbe, G. and Kramer, A. 1996. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. RNA, 2:794-810.

Baehrecke, E. H. 1997. who encodes a KH RNA binding protein that functions in muscle development. Development, 124:1323-1332.

Barbarese, E. 1991. Spatial distribution of myelin basic protein mRNA and polypeptide in Quaking oligodendrocytes in culture. J. Neursci. Res., 29:271-281.

Bartoszewicz, Z. P., Noronha, A. B., Fujita, N., Sato, S., Bo, L., Trapp, B. D. and Quarles, R. H. 1995. Abnormal expression and glycosylation of the large and small isoforms of myelin-associated glycoprotein in dysmelinating quaking mutants. J. Neurchem. Res., 41:27-38.

Benda, P., Lightbody, J., Sato, G., Levine, L. and Sweet, W. 1968. Differentiated rat glial cell strain in tissue culture. Science, 161:370-371.

Bo, L., Quarles, R. H., Fujita, N., Bartoszewiez, Z., Sato, S. and Trapp, B. D. 1995. Endocytic depletion of L-MAG from CNS myelin in quaking mice. J. Cell Biol., 131:1811-1820.

Bode, V. C. 1984. Ethylsnitrosourea mutagenesis and the isolation of mutant alleles for specific genes located in the t region of the mouse chromosome 17. Genetics, 108:457-470.

Chen, T., Damaj, B., Herrerra, C., Lasko, P. and Richard, S. 1997. Self-association of the single-KH domain family members Sam68, GRP33, GLD-1 and Qk1: Role of the KH domain. Mol. Cell. Biol., 17:5707-5718.

Cruz-Alvarez, M. and Pellicer, A. 1987. Cloning of a full-length complementary cDNA for as *Artemia salina* glycine-rich protein. J. Biol. Chem., 262:13377-13380.

Ebersole, T. A., Chen, Q., Justice, M. J. and Artzt, K. 1996. The *quaking* gene unites signal transduction and RNA binding in the developing nervous system. Nat. Genet., 12:260-265.

Evan, G. I. and Bishop, J. M. 1985. Isolation of monoclonal antibodies specific for the

human c-myc protooncogene product. Mol. Cell. Biol., 4:2843-2850.

Frail, D. E. and Braun, P. E. 1985. Abnormal expression of the myelin-associated glycoprotein in the central nervous system of dysmyelinating mutant mice. J. Neurochem., 45:1071-1075.

Francis, R., Barton, M. K., Kimbel, J. and Schedl, T. 1995a. Control of oogenesis, germline proliferation and sex determination by the *C. elegans*gene *gld-1*. Genetics, 139:579-606.

Francis, R., Maine, E. and Schedl, T. 1995b. *gld-1*, a cell-type specific tumor suppressor gene required in *C. elegans*. Genetics, 139:607-630.

Fyrberg, C., Becker, J., Barthmaier, P., Mahaffey, J. and Fyrberg, E. 1997. A Drosophia muscle-specific gene related to the mouse quaking locus. Gene, 197:315-323.

Gibson, T. J., Thompson, J. D. and Heringa, J. 1993. The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. FEBS Lett., 324:361-366.

Goya, L., Feng, P. T., Aliabadi, S. and Timiras, P. S. 1996. Effect of growth factors on the in vitro growth and differentiation of early and late passage of C6 glioma cells. Int. J. Dev. Neurosci., 14:409-17.

Hardy, R. J., Loushin, C. L., Friedrich Jr., V. L., Chen, Q., Ebersole, T. A., Lazzarini, R. A. and Artzt, K. 1996. Neural cell type-specific expression of QKI proteins is altered in the *quaking* viable mutant mice. J. Neuroscience, 16:7941-7949.

Hogan, E. L. and Greenfield, S. 1984. Animal models of genetic disorders of myelin. p. 489-534. In P. Morells (ed.), Myelin, 2nd ed. Plenum Press, New York, N. Y.

Jones, A. R., Francis, R. and Schedl, T. 1996. GLD-1, a cytoplasmic protein essential for oocyte differentiation, shows stage- and sex-specific expression during *Caenorhabditis elegans* germline development. Dev. Bol. 180:165-183.

Jones, A. R. and Schedl, T. 1995. Mutations in GLD-1, a female germ cell-specific tumor suppressor gene in *C.elegans*, affect a conserved domain also found in Sam68. Genes and Development, 9:1491-1504.

Justice, M. J. and Bode, V. C. 1988. Three ENU-induced alleles of the murine quaking locus are recessive embryonic lethal mutations. Genet. Res. Camb., 51:95-102.

Khandjian, E. W., Corbin, F., Woerly, S. and Rousseau, F. 1996. The fragile X mental retardation protein is associated with ribosomes. Nat. Genet., 12:91-93.

Lin, Q., Taylor, S. J. and Shalloway, D. 1997. Specificity and determinants of Sam68 RNA binding. J. Biol. Chem., 272:27274-27280.

Lupas, A., van Dyke, M. and Stock, J. 1991. Predicting Coiled Coils from protein sequences. Science, 252:1162-1164.

Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I. and Levens, D. 1996. Heterogeneous nuclear ribonucleoprotein K is a transcription factor. Mol. Cell. Biol., 16:2350-2360.

Mondino, A. and Jenkins, M. K. 1995. Accumulation of sequence-specific RNA-binding proteins in the cytosol of activated T cells undergoing RNA degradation and apoptosis. J. Biol. Chem., 270:26593-26601.

Musco, G., Stier, G., Joseph, c., Morelli, M. A. and Pastore, A. 1996. Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. Cell, 85:237-245.

Ostareck, D. H., Ostareck-Lederer, A., Wilm, M., Thiele, B. J., Mann, M. and Hentze, M. W. 1997. mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. Cell, 89:597-606.

Radhakrichna, M. and Almazan, G. 1994. Protein kinases mediate basic fibroblast growth factor's stimulation of proliferation and c-fos induction in oligodendrocyte progenitors. Mol. Brain. Res., 24:118-128.

Richard, S., Yu, D., Blumer, K. J., Hausladen, D., Olszowy, M. W., Connelly, P. A. and Shaw, A. S. 1995. Association of p62, a multi-functional SH2- and SH3-binding protein, with src-family tyrosine kinases, Grb2, and phospholipase Cγ-1. Mol. Cell. Biol., 15:186-197.

Shedlovsky, A., King, T. R. and Dove, W. F. 1988. Saturation germ line mutagenesis of the

murine t region including a lethal allele of the quaking locus. Proc. Natl. Acad. Sci., 85:180-184.

Sidman, R. L., Dickie, M. M. and Appel, S. H. 1964. Mutant mice (quaking and jimpy) with deficient myelination in the central nervous system. Science, 144:309-311.

Siomi, H., Matunis, M. J., Michael, W. M. and Dreyfuss, G. 1993. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. Nucl. Acids. Res., 21:1193-1198.

Siomi, M. C., Zhang, Y., Siomi, H. and Dreyfuss, G. 1996. Specific sequences in the Fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interaction among them. Mol. Cell. Biol., 16:3825-3832.

Taupin, J.-L., Tian, Q., Kedersha, N., Robertson, M. and Anderson, P. 1995. The RNA-binding protein TIAR is translocated from the nucleus to the cytoplasm during Fas-mediated apoptotic cell death. Proc. Natl. Acad. Sci. USA, 92:1629-1633.

Taylor, S. J., Anafi, M., Pawson, T. and Shalloway, D. 1995. Functional interaction between c-src and its mitotic target, Sam68. J. Biol. Chem., 270:10120-10124.

Vernet, C. and Artzt, K. 1997. STAR, a gene family involved in signal transduction and activation of RNA. Trends in Genet., 13:479-484.

Wang, L. L., Richard, S. and Shaw, A. S. 1995. p62 association with RNA is regulated by tyrosine phosphorylation. J. Biol. Chem, 270:2010-2013.

Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M. F., Polakis, P. and McCormick, F. 1992. Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. Cell, 69:551-558.

Zaffran, S., Astier, M., Gratecos, D. and Semeriva, M. 1997. The held out wings (how) Drosophila gene encodes a putative RNA binding protein involved in the control of muscular and cardiac activity. Development, 124:2087-2098.

Zhu, W., Wiggins, R. C. and Konat, G. W. 1994. Glucocorticoid-induced upregulation of proteolipid protein and myelin-associated glycoprotein genes in C6 cells. J. Neurosci. Res., 37:208-12.

Zorn, A. M., Grow, M., Patterson, K. D., Ebersole, T. A., Chen, Q., Artzt, K. and Krieg, P. A. 1997. Remarkable sequence conservation of transcripts encoding amphibian and mammalian homologues of *quaking* a KH domain RNA-binding protein. Gene, 188:199-206.

Zorn, A. M. and Krieg, P. A. 1997. The KH domain protein encoded by quaking functions as a dimer and is essential for notochord development in Xenopus embryos. Genes & Development, 11:2176-2190.

Chapter 4

A ROLE FOR THE GSG DOMAIN IN LOCALIZING SAM68 TO NOVEL NUCLEAR STRUCTURES IN CANCER CELL LINES

PREFACE

Sam68 localizes predominantly in the nucleus at steady state. Although previous studies indicated the presence of a putative NLS at the C terminus of Sam68, it is not clear whether the NLS alone is responsible for localizing Sam68 to the nucleus. In this study, we investigated the role of the GSG domain in Sam68 localization.

SUMMARY

The GSG (GRP33, Sam68, GLD-1) domain is a protein module found in an expanding family of RNA binding proteins. The numerous missense mutations identified genetically in the GSG domain support its physiological role. Although the exact function of the GSG domain is not known, it has been shown to be required for RNA binding and oligomerization. Here it is shown that the Sam68 GSG domain plays a role in protein localization. We show that Sam68 concentrates into novel nuclear structures that are predominantly found in transformed cells. These Sam68 nuclear bodies (SNBs) are distinct from coiled bodies, gems, and promyelocytic nuclear bodies. Electron microscopic studies show that SNBs are distinct structures that are enriched in phosphorus and nitrogen, indicating the presence of nucleic acids. A GFP-Sam68 fusion protein had a similar localization as endogenous Sam68 in HeLa cells, diffusely nuclear with two to five SNBs. Two other GSG proteins, the Sam68-like mammalian proteins SLM-1 and SLM-2, colocalized with endogenous Sam68 in SNBs. Different GSG domain missense mutations were investigated for Sam68 protein localization. Six separate classes of cellular patterns were obtained, including exclusive SNB localization and association with microtubules. These findings demonstrate that the GSG domain is involved in protein localization and define a new compartment for Sam68, SLM-1, and SLM-2 in cancer cell lines.

INTRODUCTION

The GSG (GRP33, Sam68, GLD-1) domain is an ~200-amino acid protein module found in proteins closely associated with RNA (Jones and Schedl, 1995). GSG domain-containing proteins include the *Artemia salina* GRP33 (Cruz-Alvarez and Pellicer, 1987), mammalian Sam68 (Wong et al., 1992), *Caenorhabditis elegans* GLD-1 (Jones and Schedl, 1995), SF1(Arning et al., 1996), *Drosophila* Who/How (Baehrecke, 1997; Fyrberg et al., 1997; Zaffran et al., 1997), *Xenopus* Xqua (Zorn et al., 1997), mouse Qk1 (Ebersole et al., 1996), zebrafish Qk1 (Tanaka et al., 1997), mouse SLM-1 and SLM-2 (Di Fruscio et al., 1999), *Drosophila* KEP1 and Sam50 (Di Fruscio et al., 1998), and *Drosophila* Qk1-related proteins (Fyrberg et al., 1998). The GSG domain is also called STAR, for signal transduction and activation of RNA, domain (Vernet and Artzt, 1997). In the GSG domain is embedded an extended heterogeneous nuclear ribonucleoprotein (hnRNP) K homology (KH) domain that contains 6 and 20 extra amino acids in KH domain loops 1 and 4, respectively (Gibson et al., 1993). Although the roles of these extended KH domain loops are unknown, they have been postulated to mediate protein-protein interactions (Musco et al., 1996). Indeed, we have shown that the KH domain loops 1 and 4 are required for Sam68 multimerization (Chen et al., 1997). Although the exact function of the GSG domain is unknown, two properties have been ascribed to this protein module: self-association (Chen et al., 1997; Zorn and Krieg, 1997; Di Fruscio et al., 1998) and RNA binding (Berglund et al., 1997; Chen et al., 1997; Lin et al., 1997; Zorn and Krieg, 1997; Chen and Richard, 1998; Di Fruscio et al., 1998, 1999; Rain et al., 1998).

Genetic mutations in the GSG domain that result in growth or developmental defects have been isolated in a number of genes, supporting a physiological role for this protein module. In *C. elegans*, the GSG protein GLD-1 is required for normal oocyte development and has been shown to function as a tumor suppressor (Francis et al., 1995a,b). Thirty-two *gld-1* mutations, including several missense mutations within the GSG domain, have been identified and classified into six phenotypic classes (Jones and Schedl, 1995). In mice, the GSG protein Qk1 is involved in myelination and early embryogenesis (Hogan and

Greenfield, 1984). A missense mutation (E48G) identified in the N-terminal portion of the Qk1 GSG domain is known to be embryonic lethal in mice (Justice and Bode, 1988; Ebersole et al., 1996). The molecular defect is most likely due to the failure of the Qk1 proteins to dimerize (Chen and Richard, 1998). The *Drosophila melanogaster* GSG protein Who/How plays a critical role in skeletal muscle development as weak alleles result in the "wings-held-out" phenotype (Baehrecke, 1997; Fyrberg et al., 1997; Zaffran et al., 1997).

The GSG protein Sam68 is a substrate for Src kinases during mitosis (Fumagalli et al., 1994; Taylor and Shalloway, 1994), and it has been proposed to serve as an adapter protein for Src kinases (Richard et al., 1995; Taylor et al., 1995). Sam68 is an RNA-binding protein (Wong et al., 1992), and its RNA binding activity is negatively regulated by tyrosine phosphorylation (Wang et al., 1995). The nuclear localization of Sam68 (Wong et al., 1992; Ishidate et al., 1997) suggests that it may interact with cytoplasmic proteins, including Src kinases, during mitosis, when the nuclear envelope breaks down. Two Sam68-like proteins, SLM-1 and SLM-2, were recently identified (Di Fruscio et al., 1999). They are both nuclear proteins that heterodimerize with Sam68 (Di Fruscio et al., 1999). SLM-1 is tyrosine phosphorylated by Src kinases during mitosis, like Sam68, whereas SLM-2 is not a substrate for Src kinases (Di Fruscio et al., 1999). The functions of Sam68, SLM-1, and SLM-2 are currently unknown.

The nucleus contains a number of specialized subnuclear structures, including the nucleolus, interchromatin granule speckles, coiled bodies, gems, and promyelocytic (PML) nuclear bodies (Lamond and Earnshaw, 1998). Speckles are observed when immunostained with antibodies to splicing factors such as SC-35 (Fu and Maniatis, 1990) and represent aggregates of snRNPs and splicing factors (Spector, 1993). Coiled bodies are round structures 0.1 to 1.0 μm in diameter containing coiled threads (Monneron and Bernhard, 1969). They contain spliceosomal small nuclear ribonucleoproteins (snRNPs), U3 small nucleolar ribonucleoprotein (snoRNP), fibrillarin, NOPP140, and an autoantigen called p80-coilin (Lamond and Carmo-Fonseca, 1993; Gall et al., 1995). Antibodies against p80-

coilin have been widely used as a marker for coiled bodies (Andrade et al., 1993). Coiled bodies vary in number from 1 to 10, and the number increases remarkably in cancer cells (Spector et al., 1992). Gems (gemini of coiled bodies) are very similar to coiled bodies in number, size, and distribution pattern (Liu and Dreyfuss, 1996). However, gems and coiled bodies appear to be indistinguishable in some cell types (Matera and Frey, 1998; Bechade et al., 1999). Gems contain the survival of motor neurons (SMN) protein (Liu and Dreyfuss, 1996), which is encoded by the gene responsible for spinal muscular atrophy (Lefebvre et al., 1995), and an interacting protein, SIP1 (Liu et al., 1997). SMN and SIP1 have been shown to interact with several spliceosomal snRNP proteins and play an essential role in snRNP biogenesis and pre-mRNA splicing (Fischer et al., 1997; Liu et al., 1997; Pellizzoni et al., 1998). PML nuclear bodies, also known as PODs (PML oncogenic domain), are nuclear matrix-associated structures altered by oncoproteins and viruses (Dyck et al., 1994; Weis et al., 1994; Doucas et al., 1996). Anti-PML antibodies serve as a marker for PML nuclear bodies. Generally, each mammalian cell contains 10-20 PML nuclear bodies, with sizes range from 0.3 to 1.0 μm. In acute promyelocytic leukemia cells, PML nuclear bodies are disorganized into micropunctate structures (Dyck et al., 1994; Koken et al., 1994; Weis et al., 1994). The functions of coiled bodies, gems, and PML nuclear bodies remain unknown.

Here we demonstrate that Sam68, SLM-1, and SLM-2 are concentrated in distinct nuclear bodies. These nuclear structures show dynamic changes in response to transcriptional inhibitors and during mitosis. Sam68 nuclear bodies contain nucleic acids, as visualized by electron microscopy, but they are not the site of ongoing transcription, as determined by bromo (Br)-UTP incorporation. Double immunostaining studies demonstrate that the Sam68 nuclear bodies are distinct from coiled bodies, gems, and PML nuclear bodies. Amino acid substitutions or deletions within the Sam68 GSG domain concentrate these mutant proteins in nuclear bodies, in the cytoplasm, and on microtubules. These data suggest that the GSG domain plays a role in localizing Sam68/SLM to different cellular compartments, including a novel nuclear structure.

MATERIALS AND METHODS

DNA Constructs. The green fluorescent protein (GFP)-Sam68 fusion constructs, except for GFP-Sam68:I→N-nuclear localization signal (NLS) and GFP-Sam68ΔC, were generated by subcloning the *Eco*RI fragments of corresponding myc-Sam68 constructs into pEGFP-C1 (Clontech, Palo Alto, CA). myc-Sam68f, Sam68Δ1-67 (initially called p62), Sam68ΔN, Sam68ΔKH, Sam68ΔL1, Sam68ΔL4, Sam68:G→D (GLD-1 missense mutation), Sam68:I→N (human FMR1 missense mutation), Sam68:E→G (mouse Qk1 missense mutation), and Sam68:2G→R (GLD-1 missense mutation) have been described previously (Richard et al., 1995; Chen et al., 1997). myc-Sam68ΔKH isoform (Sam68 splice variant) , Sam68:NF→DL, Sam68:R→C (Who/How missense mutation), Sam68:R→A (putative Sam68 NLS disruption), Sam68:D→N (GLD-1 missense mutation), Sam68:A→T (GLD-1 missense mutation), Sam68:P→L (GLD-1 missense mutation), and Sam68ΔL1-NF were generated by inverse PCR using myc-Sam68f as the DNA template and the following oligonucleotide pairs as primers: 5'-GAG GAA GAG TTG CGC AAG GGT-3' and 5'-CTT TGG ATA CTG CTT GAC AGG-3' (for myc-Sam68ΔKH isoform); 5'-GAC CTT GTG GGG AAG ATT CTT GGA-3' and 5'-GAA CTT TGG ATA CTG CTT GAC-3' (for myc-Sam68:NF→DL); 5'-TGC GAC AAA GCC AAG GAG GAA-3' and 5'-CAT TGA ACC CTT CCC CAA GAC-3' (for myc-Sam68:R→C); 5'-TCC GCA CCA GTG AAG GGA GCA TAC-3' and 5'-TCC AGG AGC CTT CAG TGA TGG CCT-3' (for myc-Sam68:R→A); 5'-AAG CAG TAT CTA AAG TTC AAT TTT-3' and 5'-GAC AGG TAT CAG CAC GCG TTC-3' (for myc-Sam68:P→L); 5'-ACT ATG GAA GAA GTC AAG-3' and 5'-ACG GGC CAT AAG AGC ATA-3' (for myc-Sam68:A→T); 5'-AAT ATC TGT CAG GAG CAG-3' and 5'-ATC CAT CAT ATC TGG TAC-3' (for myc-Sam68:D→N); and 5'-AAT TTT GTG GGG AAG ATT-3' and 5'-GAC AGG TAT CAG CAC GCG TTC-3' (for myc-Sam68ΔL1NF). (Underlined nucleotides denote changes introduced.)

GFP-Sam68:I→N-NLS was constructed by a two-step subcloning strategy. The NLS

sequence was first introduced into a DNA fragment encoding the C-terminal 113 amino acids of Sam68 by PCR amplification with myc-Sam68f as the DNA template and the following oligonucleotides as primers: 5'-CTT GAA TTC AGT ACC TGA ACC CTC TCG-3' and 5'-TCA GAA TTC *ACA CCT TAC GCT TCT TCT TTG* GAT GCT CTC TGT ATG CTC CCT TCA CTG G-3'. (The *Eco*RI site is underlined and the sequence encoding the SV40 large T antigen NLS, PKKKRKV [Kalderon et al., 1984], is italicized.) The amplified DNA fragment was digested with *Eco*RI and *Xho*I (an internal *Xho*I site was used) and subcloned into the corresponding sites of pEGFP-C1, generating GFP-NLS. The *Xho*I fragment of GFP-Sam68:I→N was then inserted into the *Xho*I site of GFP-NLS, generating GFP-Sam68:I→N-NLS. GFP-Sam68ΔC was constructed by subcloning the *Sac*I fragment of myc-Sam68f into the corresponding *Sac*I site in pEGFP-C1. GFP-SLM-1 and GFP-SLM-2 were described previously (Di Fruscio et al., 1999). The construct encoding GFP-polypyrimidine-tract binding protein (PTB) was generated by subcloning the PTB cDNA (Patton, 1991; kindly provided by Nahum Sonenberg, McGill University) into the *Bgl*II and *Sma*I sites of pEGFP-C1. All GFP fusion constructs produce proteins with GFP fused to their N termini. The identities of the plasmid constructs were verified by dideoxynucleotide sequencing with Sequenase (United States Biochemical, Cleveland, OH).

Antibodies. The Sam68 mAb 7-1 (sc-1238), the Sam68 rabbit polyclonal antibody C-20 (sc-333), the PML mAb PG-M3 (sc-966), and the SMRT goat polyclonal antibody N-20 (sc-1610) were purchased from Santa Cruz Biotech (Santa Cruz, CA). The anti-tubulin mAb B-5-1-2 was from Sigma Chemical (St. Louis, MO). The anti-SMN mAb was purchased from Transduction Laboratories (Lexington, KY). Anti-bromodeoxyuridine (BrdU) mAb (BMC9318) was purchased from Boehringer Mannheim (Indianapolis, IN). Anti-SC35 mAb was purchased from PharMingen (San Diego, CA). Anti-coilin rabbit polyclonal antibody was kindly provided by Edward Chan (Andrade et al., 1993), the anti-fibrillarin mAb 72B9 was kindly provided by Michael Pollard (Scripps Institute, La Jolla, CA), and the Y12 mouse mAb was kindly provided by Mark Bedford (Harvard University,

Cambridge, MA). We generated a new rabbit anti-Sam68 antibody (AD1). Rabbits were immunized with a peptide encompassing amino acids 330-348 of Sam68 (RGVPPPPTVRGAPTPRAR) covalently coupled to keyhole limpet hemacyanin. The 9E10 anti-myc mAb was from the American Type Culture Collection (Rockville, MD). Goat anti-mouse and goat anti-rabbit rhodamine-conjugated antibodies were purchased from Pierce (Rockford, IL). Goat anti-mouse and goat anti-rabbit FITC-conjugated antibodies were from Jackson Laboratories (Bar Harbor, ME).

Cell Culture and Transient Transfection. HeLa cells were maintained in DMEM (Life Technologies-BRL, Grand Island, NY), supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, 100 U/ml penicillin, 100 µg/ml streptomycin (all from ICN, Costa Mesa, CA), and 10% bovine calf serum (Hyclone, Logan, UT). Cells were maintained at 37° C in 5% CO_2. The day before transfection, the cells were plated on glass coverslips at a density of 10^5 cells per 22-mm^2 coverslip (Fisher Scientific, Nepean, Ontario, Canada). DNA transfections of GFP plasmids were performed with the use of LipofectAMINE PLUS reagent (Life Technologies-BRL) according to the manufacturer's instructions. The transfection of myc-Sam68 into HeLa cells was carried out with the T7 vaccinia virus system as previously described (Richard et al., 1995). The other cell lines used in this study were maintained as suggested by American Type Culture Collection.

Labeling of Transcription Sites in Permeabilized Cells. Labeling of RNA synthesis sites was done according to published methods (Jackson et al., 1993; Pombo et al., 1998) with modifications. Briefly, HeLa cells grown on coverslips were transfected with GFP-Sam68, GFP-Sam68ΔL1, or GFP-PTB. Twelve hours after transfection, the cells were washed once with 1x PBS, once with a "physiological" buffer (100 mM potassium acetate, 30 mM KCl, 10 mM Na_2HPO_4, 1 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, 0.2 mM PMSF, 10 U/ml RNAguard RNase inhibitor, and 100 µg/ml BSA), and then permeabilized with 50 µg/ml saponin in physiological buffer for 5 min at 4°C. The permeabilized cells were rinsed

with ice-cold physiological buffer and immediately incubated with physiological buffer supplemented with 0.1 mM ATP, CTP, GTP and Br-UTP for 20 min at 33°C. After rinsing with ice-cold physiological buffer, the cells were fixed with 4% paraformaldehyde, permeabilized with 1% Triton X-100, and immunostained with anti-BrdU mAb, which also cross-reacts with Br-UTP, followed by a rhodamine-conjugated goat anti-mouse secondary antibody.

Immunofluorescence and Confocal Microscopy. Transfected or untransfected cells were fixed with 4% paraformaldehyde in 1x PBS for 5 min at room temperature and permeabilized with either 1% Triton X-100 in PBS for 5 min at room temperature or 50% methanol:50% acetone for 15 min at -20°C. If the cells were to be visualized only for GFP, then the glass cover slide was mounted onto glass slides with glycerol containing 3 μg/ml DAPI to stain the nuclei. If the cells required antibody staining, the permeabilized cells were blocked with 10% calf serum (Hyclone) in PBS for 30 min at room temperature, the primary antibodies were incubated at room temperature for 1 hr in PBS containing 3% BSA, anti-SMN (1:20), anti-PML (1:50), anti-coilin (1:500), anti-tubulin (1:200), anti-myc ascites (1:500), anti-fibrillarin tissue culture supernatant (1:20), anti-Y12 tissue culture supernatant (1:1), anti-SC35 (1:100), anti-BrdU (6 μg/ml), anti-Sam68 C-20 and AD1 (1:200), and anti-Sam68 7-1 (1:20). The cells were washed extensively with PBS and incubated with the appropriate secondary antibodies (1:200) in PBS containing 3% BSA for 20 min. The cells were washed and mounted onto glass slides as described above. The cells were visualized with a Leitz (Wetzlar) Aristoplan fluorescence microscope or by confocal microscopy.

Correlative Microscopy and Electron Microscopy. Detailed descriptions of the electron microscopy procedure are presented elsewhere (Bazett-Jones and Hendzel, 1999). Briefly, HeLa cells were grown on plastic caps, fixed with 1% paraformaldehyde in PBS, permeabilized with 0.5% Triton X-100 in PBS, and labeled with immunofluorescence using anti-Sam68 AD1 rabbit polyclonal antibodies. The cells were then refixed in 2%

glutaraldehyde in PBS for 5 min and dehydrated in ethanol, starting at 30%, before embedding in Quetol 651 resin. Sections of ~30 nm thickness were obtained by ultramicrotomy with a diamond knife (Drukker, Cuijk, The Netherlands). The sections were placed directly onto finder grids. Electron micrographs were obtained with a Gatan (Pleasanton, CA) 14-bit slow-scan cooled charge-coupled device detector on a Carl Zeiss (Thornwood, NY) EM 902 transmission electron microscope equipped with an imaging spectrometer.

RESULTS

Sam68, SLM-1, and SLM-2 Localize in Nuclear Dots. Sam68 has been shown to be present in membranes and the nucleus of NIH 3T3 cells (Wong et al., 1992; Ishidate et al., 1997). To further examine the subcellular localization of Sam68, we carried out indirect immunofluorescence studies in HeLa cells with three different anti-Sam68 antibodies. The Sam68 antibodies were mouse mAb 7-1 as well as rabbit polyclonal antibodies C-20 (Taylor and Shalloway, 1994) and AD1. mAb 7-1 is Sam68 specific, whereas C-20 recognizes Sam68, SLM-1, and SLM-2 (Di Fruscio et al., 1999). AD1 is a Sam68-specific antibody that we generated against a peptide containing amino acids 330-348 of mouse Sam68. The specificity of this antibody is depicted in Fig. 4-1A. HeLa cells expressing GFP fusion proteins of Sam68, SLM-1, and SLM-2 were lysed, divided equally, and analyzed by immunoblotting with either rabbit C-20 or AD1 antibody. C-20 recognized GFP-Sam68, GFP-SLM-1, and GFP-SLM2, whereas AD1 recognized only GFP-Sam68. A band at 68 kDa was observed in both C-20 and AD1 immunoblots that correspond to endogenous Sam68/SLM proteins. The band at 68 kDa in the AD1 immunoblots was less intense than the corresponding band in the C-20 lanes, consistent with our immunodepletion studies demonstrating that C-20 recognizes other proteins at 68 kDa that may be SLM-1, SLM-2, or yet unidentified SLMs (Di Fruscio et al., 1999).

To examine the localization of endogenous Sam68, the antibodies were used in indirect immunofluorescence studies in HeLa cells. All three anti-Sam68 antibodies gave a similar pattern, a diffuse nucleoplasmic staining with approximately two to five prominent nuclear dots, that was not observed with normal rabbit serum (Fig. 4-1B). C-20 antibody gave a stronger diffuse nucleoplasmic signal, which most likely reflected the immunostaining of both Sam68 and SLM proteins. The strong nucleoplasmic staining made the nuclear dots less prominent, which may be one of the reasons that the nuclear dots were not detected in previous studies (see DISCUSSION). Plasmids expressing GFP alone or GFP-Sam68 were transfected in HeLa cells, and the cells were visualized live using fluorescence microscopy. GFP-Sam68 displayed an expression pattern identical to that observed with

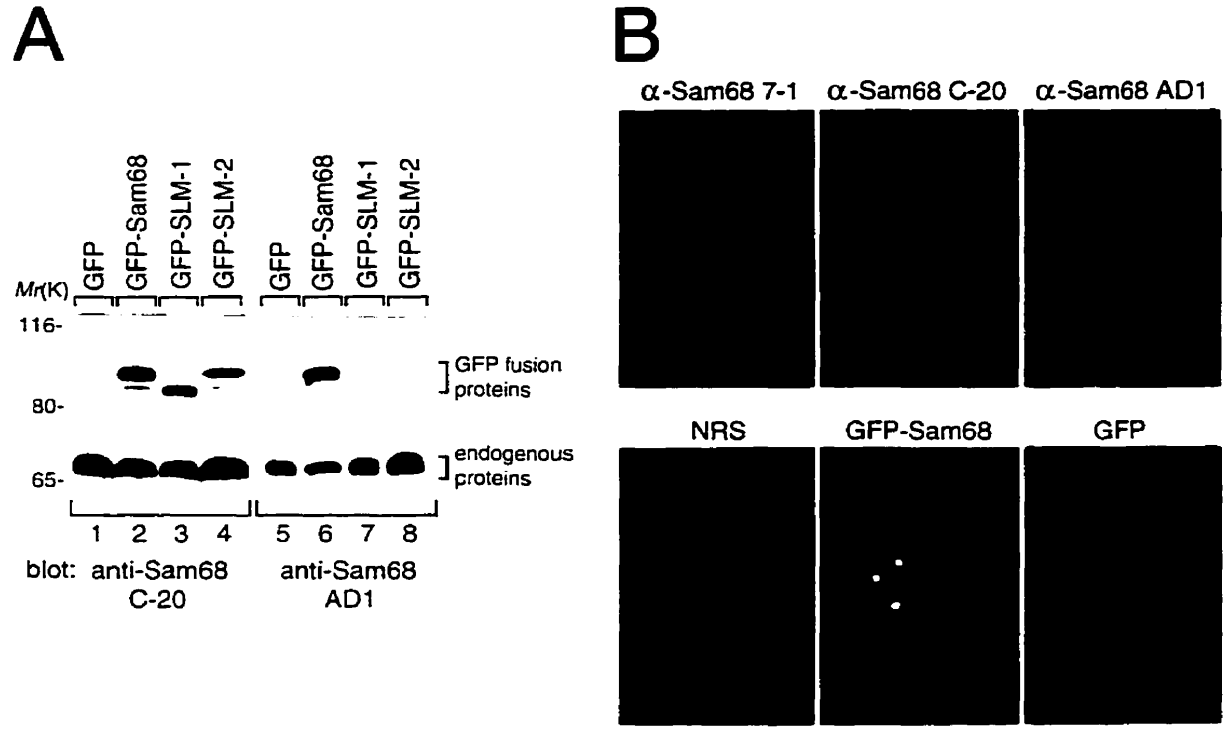


FIG. 4-1. Sam68 localizes in nuclear bodies. (A) Anti-Sam68 AD1 recognizes Sam68 but not SLM-1 and SLM-2. HeLa cells transfected with GFP alone, GFP-Sam68, GFP-SLM-1, or GFP-SLM-2 were lysed and the cell lysates were divided equally, resolved by SDS-PAGE, transferred to nitrocellulose membranes, and immunoblotted with anti-Sam68 C-20 (lanes 1-4) or AD1 (lanes 5-8) antibodies. The migration of endogenous and GFP-Sam68/SLM proteins is indicated on the right, and the positions of molecular mass markers (in kilodaltons) are shown on the left. (B) HeLa cells were fixed, permeabilized, and immunostained with anti-Sam68 mAb 7-1, rabbit polyclonal Sam68/SLM C-20 antibody, rabbit polyclonal antibody AD1, or normal rabbit serum (NRS), followed by a rhodamine-conjugated goat anti-mouse or anti-rabbit secondary antibody, and visualized by fluorescence microscopy. Expression vectors encoding GFP-Sam68 or GFP alone were transfected into HeLa cells, and 12 h after transfection, live cells were visualized by fluorescence microscopy.

endogenous Sam68, whereas GFP alone was expressed diffusely throughout the cell (Fig. 4-1B). Taken together, these findings demonstrate that the localization of Sam68 into nuclear dots was not an artifact of overexpression, fixation and/or immunostaining procedures.

We next investigated whether the closely related Sam68 family members SLM-1 and SLM-2 also concentrated in nuclear dots. HeLa cells transfected with GFP-SLM-1 and GFP-SLM-2 were immunostained with anti-Sam68 AD1 antibody, and colocalization was determined by confocal microscopy. GFP-SLM-1 displayed a localization pattern similar to that of Sam68: diffuse extranucleolar staining with one to five distinct nuclear dots (Fig. 4-2a), and these dots colocalized with endogenous Sam68 nuclear dots (Fig. 4-2c, arrows). GFP-SLM-2 displayed three major localization patterns: ~40% of transfected cells exhibited a pattern similar to that of Sam68 and SLM-1 (Fig. 4-2d), another ~40% of transfected cells had GFP-SLM-2 accumulated in the nucleoli with a relatively weak nucleoplasmic staining (Fig. 4-2g), and the remaining ~20% of transfected cells showed evenly diffuse staining throughout the nucleus, including both the nucleoplasm and the nucleoli (Fig. 4-2j). Nuclear dots were observed in all three localization patterns of GFP-SLM-2, and they always colocalized with Sam68 nuclear dots (Fig. 4-2, f, i and l, arrows). These data suggest that Sam68 nuclear dots also contain SLM-1 and SLM-2.

A panel of human, mouse, and rat cell lines was examined for the presence of Sam68 nuclear dots using indirect immunofluorescence with anti-Sam68 antibodies. Sam68 nuclear dots were predominantly present in some, but not all, human transformed cell lines. Approximately 50-90% of HeLa (human cervical cancer), BT-20 (human breast cancer), and Hs578T (human breast cancer) cells contained Sam68 nuclear dots (Table 4-1). Sam68 nuclear dots were rarely found (~1-7%) in immortalized cells such as NIH 3T3 or other transformed cell lines examined, such as v-Src-transformed 3T3 cells, SK-N-MC, MCF-7, C6 glioma, GH3, Neuro-2A, AtT-20, and 293. Normal fibroblasts, including HF-7650, MEF, Rat1, and REF-52 cells, did not have Sam68 nuclear dots (<1%). Overexpression of GFP-Sam68 in NIH 3T3 and REF-52 cells did not induce the formation of nuclear dots

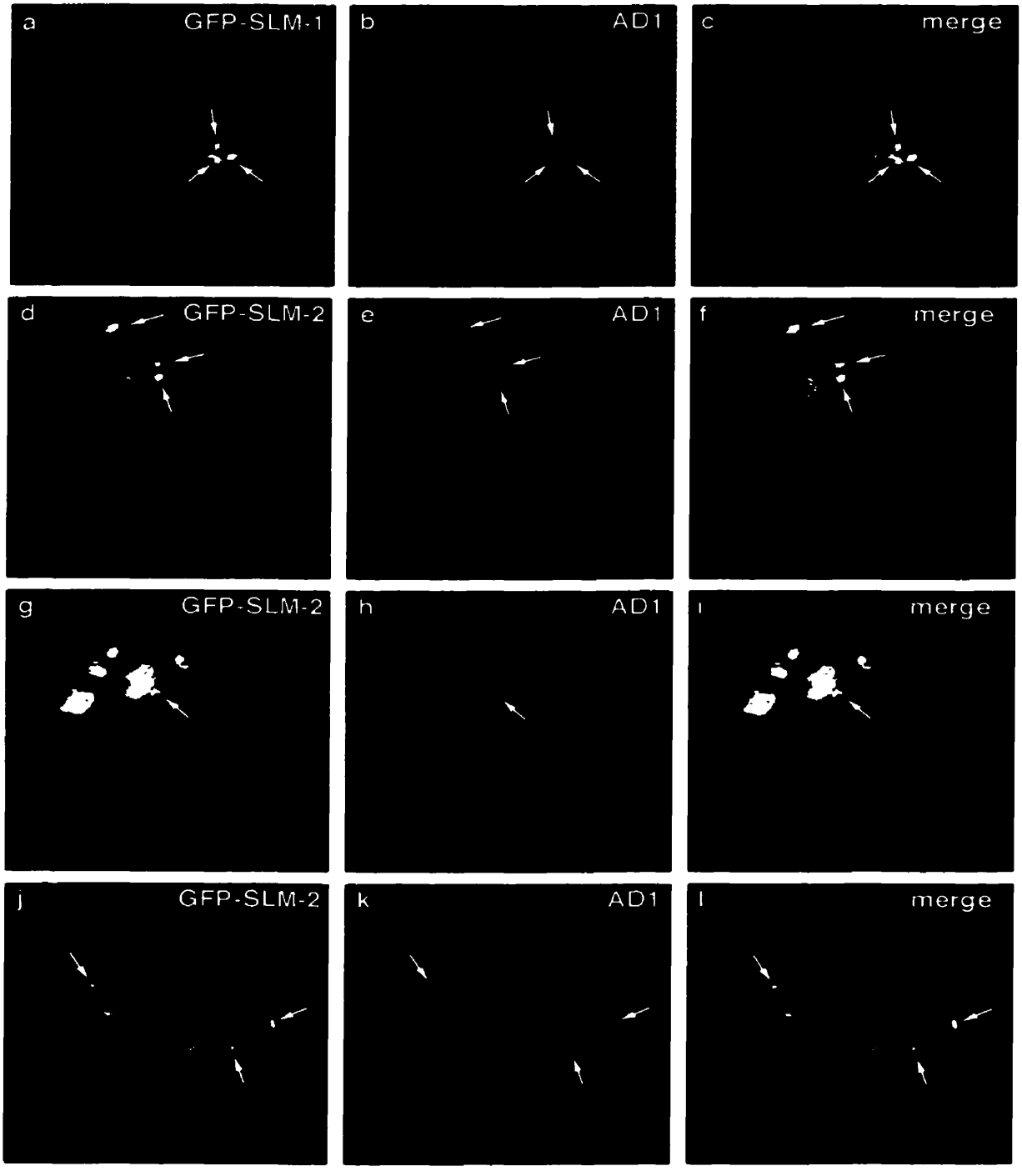


FIG. 4-2. GFP-SLM-1 and GFP-SLM-2 colocalize with Sam68 in nuclear bodies. GFP-SLM-1 (a-c) and GFP-SLM-2 (d-l) were individually transfected in HeLa cells, and the cells were fixed, permeabilized, and immunostained with anti-Sam68 AD1 antibody followed by a rhodamine-conjugated goat anti-rabbit secondary antibody (b, e, h, and k). Colocalization was determined by confocal microscopy, and the merged images are shown on the right (c, f, i, and l). The arrows point to nuclear bodies that colocalize. Three major localization patterns were observed with GFP-SLM-2: extranucleolar staining (d), accumulation in nucleoli (g), and diffuse staining in both nucleoplasm and nucleoli (j).

Table 4-1. Prevalence of Sam68 nuclear bodies in different cell lines

Cell line	Description	SNB prevalence[a] (%)
HeLa	human cervical carcinoma epithelial cell	88.7
BT-20	human breast carcinoma epithelial cell	89.1
Hs 578T	human breast carcinoma epithelial cell	48.6
MCF7	human breast adenocarcinoma epithelial cell	4.8
SK-N-MC	human neuroblastoma cell	0.7
293	human embryo kidney cell transformed with adenovirus 5	4.5
HF-7650	human fibroblast cell	0
AtT-20	mouse pituitary tumor cell	1.3
Neuro-2A	mouse neuroblastoma cell	0.4
NIH 3T3	mouse embryo fibroblast cell	5.6
Src 3T3	NIH 3T3 cell transformed with v-src	7.2
MEF	primary mouse embryo fibroblast cell	0.9
C6 glioma	rat glioma cell	1.1
GH3	rat pituitary tumor cell	0.8
Rat1	rat connective tissue epithelial cell	0.5
REF-52	rat embryo fibroblast cell	0.5

[a]Different cell lines grown on coverslips were fixed, permeabilized, and immunostained with anti-Sam68 AD1 antibody followed by a rhodamine-conjugated secondary antibody. For each cell line, an average of 250-300 cells from two independent experiments were counted, and SNB prevalence is defined as the percentage of cells that contain one or more SNBs.

(data not shown), suggesting that the presence or absence of Sam68 nuclear dots is not due to different levels of Sam68 expression in different cell lines. These findings imply that immortalization and/or transformation is a prerequisite for the appearance of Sam68 nuclear dots but that Sam68 nuclear dots are not a general marker for transformation or cancer.

Sam68 Nuclear Bodies Are Novel Nuclear Structures. The Sam68 nuclear dots resembled nuclear bodies in number and size. Confocal microscopy studies were initiated to investigate whether Sam68 nuclear dots were coiled bodies, gems, or PML nuclear bodies. HeLa cells were double immunostained with anti-Sam68 antibodies and anti-coilin, anti-SMN, or anti-PML antibodies. Colocalization of rhodamine-stained (red) Sam68 nuclear dots and fluorescein-stained (green) coiled bodies, gems, or PML nuclear bodies was analyzed by confocal microscopy. Sam68 nuclear dots did not colocalize with coiled bodies, gems, or PML nuclear bodies (Fig. 4-3). Occasionally, Sam68 nuclear dots were adjacent to or even partially overlapped coiled bodies or gems. The large arrow in Fig. 4-3f indicates partial overlapping with a gem. The significance of this overlap is unknown; the overlap may be a chance occurrence, given that overlapping was infrequently observed. Taken together, these studies suggest that Sam68 nuclear dots are novel nuclear structures distinct from coiled bodies, gems, and PML nuclear bodies. The Sam68 nuclear dots were also distinct from the perinucleolar compartment (PNC), as visualized with GFP-PTB (Huang et al., 1997) and anti-Sam68 antibodies (data not shown). Sam68 nuclear dots as well as nucleoplasmic Sam68 did not colocalize with fibrillarin (data not shown), a marker for coiled bodies and the nucleolus (Lamond and Earnshaw, 1998). Based on the uniqueness of Sam68 nuclear dots, we named them Sam68/SLM nuclear bodies (SNBs).

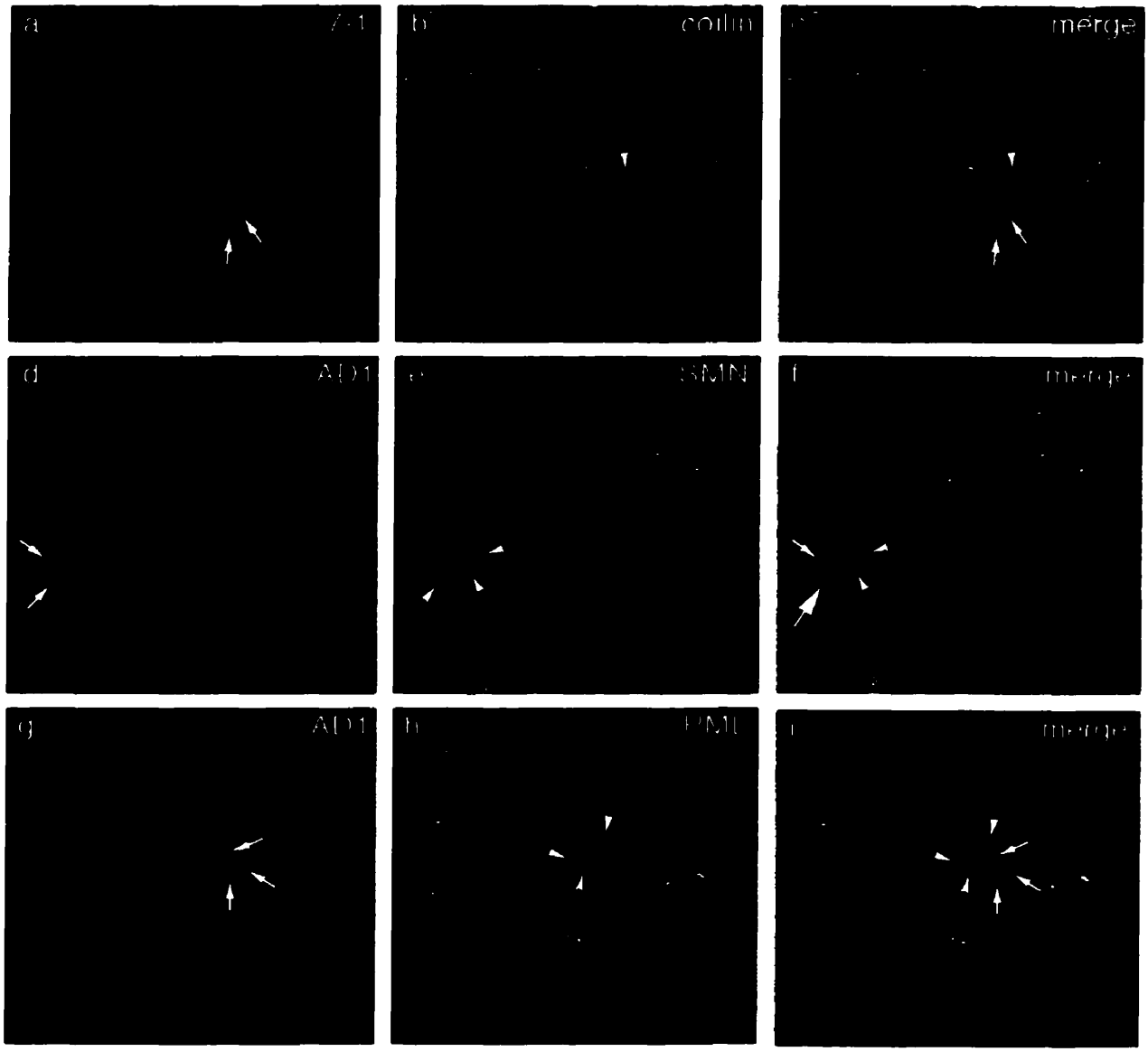


FIG. 4-3. Sam68 nuclear bodies are novel nuclear structures. HeLa cells were fixed, permeabilized, and double immunostained with anti-Sam68 7-1 and anti-coilin (a-c), anti-Sam68 AD1 and anti-SMN (d-f), or anti-Sam68 AD1 and anti-PML (g-i). Anti-Sam68 antibodies were followed by rhodamine-conjugated secondary antibodies (left panels), and the other antibodies were followed by FITC-conjugated secondary antibodies (middle panels). Colocalization was determined by merging the confocal images of the left and middle panels using confocal microscopy (right panels). Sam68 nuclear bodies are indicated by arrows, and coiled bodies, gems, or PML nuclear bodies are indicated by arrowheads. The large arrow in panel f indicates partial overlap between a Sam68 nuclear body and a gem.

SNBs Are Dynamic Structures. The dynamic nature of SNBs and their cell cycle regulation were investigated. Unsynchronized HeLa cells were immunostained with anti-Sam68 antibodies, and cells in interphase and various stages of mitosis, identified by chromatin morphology with DAPI staining, were examined for the presence of SNBs (Fig. 4-4A). In interphase cells, Sam68 exhibited a diffuse extranucleolar pattern with several SNBs in each nucleus (Fig. 4-4Aa, lower cell). In early prophase, Sam68 remained confined to the nucleus and SNBs were visible (Fig. 4-4A, a and f). SNBs disassembled in mid to late prophase (Fig. 4-4A, b and g) before the breakdown of the nuclear membrane. In metaphase (Fig. 4-4A, c and h) and anaphase (Fig. 4-4A, d and i), Sam68 was diffusely localized throughout the cell and no SNBs were visible. Of note, Sam68 did not colocalize with the chromosomes. During late telophase and/or early G1 phase, SNBs reappeared in the daughter cells (Fig. 4-4A, e and j). These findings demonstrate that SNBs are dynamic structures that disassemble during mitosis.

Transcriptional inhibition has been shown to influence the integrity of several nuclear structures, including coiled bodies and gems (Carmo-Fonseca et al., 1993; Liu and Dreyfuss, 1996). The effect of transcriptional inhibition on SNBs was examined. After 3 h of treatment with 5 μg/ml actinomycin D, a concentration known to inhibit RNA polymerases I and II, Sam68 immunostaining became less diffuse in the nucleoplasm and SNBs disassembled (Fig. 4-4Bb). SNBs also disassembled when HeLa cells were treated with 50 μg/ml α-amanitin, another transcription inhibitor, for 5 h (data not shown). The treatment of HeLa cells with the protein synthesis inhibitor cycloheximide had no effect on the structural integrity of SNBs (Fig. 4-4Bc). These observations suggested that the integrity of SNBs was dependent on transcription but not on ongoing protein synthesis. Because cycloheximide also inhibits the transcriptional activity of RNA polymerase I (Willems et al., 1969), the fact that SNBs were resistant to cycloheximide treatment suggested that the structure of SNBs was most likely dependent on the activity of RNA polymerase II but not RNA polymerase I. To exclude the interference of the fixation procedures, we examined GFP-Sam68 in living cells. Actinomycin D treatment caused the

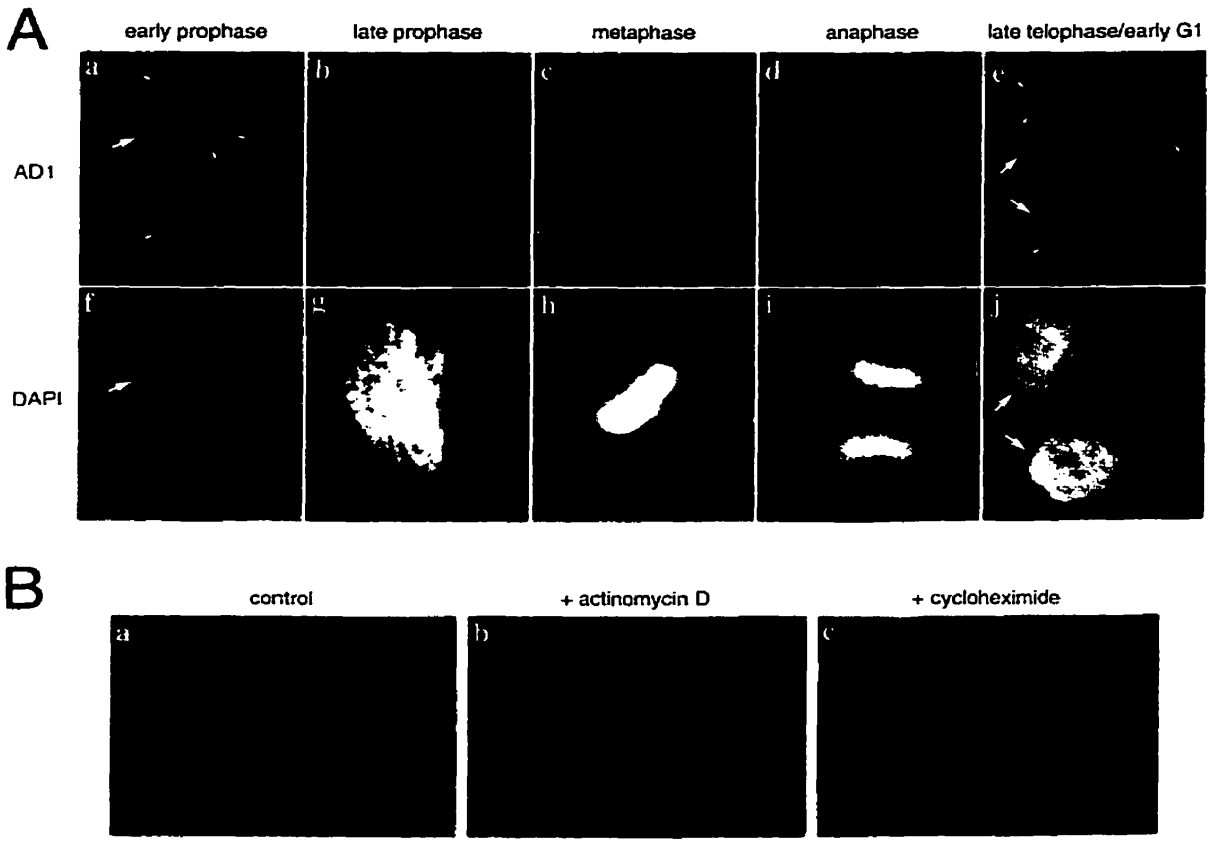


FIG. 4-4. Sam68 nuclear bodies are dynamic structures. (A) HeLa cells were fixed, permeabilized, and immunostained with anti-Sam68 AD1 antibody followed by rhodamine-conjugated goat anti-rabbit secondary antibody, and the cell nuclei were stained with DAPI. Fluorescence of representative cells in interphase (a and f, lower cells; e and j, right cells), early prophase (a and f, large arrow), late prophase (b and g), metaphase (c and h), anaphase (d and i), and late telophase/early G1 phase (e and j, large arrows) are shown. Sam68 nuclear bodies are indicated by small arrows. (B) HeLa cells were mock-treated or treated with 5 mg/ml actinomycin D (b) or 20 mg/ml cycloheximide (c) for 3 h, then fixed, permeabilized, and immunostained with anti-Sam68 AD1 antibody.

GFP-Sam68 nuclear bodies to disassemble (data not shown). Taken together, these results further indicated that SNBs, like nuclear bodies, are dynamic structures whose organization changes with metabolic states.

GSG Domain Mutations and Deletions Alter Sam68 Localization. Because many genetic mutations have been identified in the GSG and KH domains, we investigated the role of the Sam68 GSG/KH domain in protein localization. GFP-Sam68 fusion proteins containing missense mutations or deletions within the GSG domain were examined for altered cellular localization. Interestingly, different Sam68 localization patterns were observed (Fig. 4-5A), and the Sam68 mutants were grouped into six different classes based on their localization patterns (Fig. 4-5B).

In class I, all transfected cells displayed wild-type Sam68 expression pattern, a diffuse nucleoplasmic staining with several SNBs (pattern A). Sam68 alanine 247 in the KH domain and aspartic acid 262 at the C-terminal-to-KH-domain region (CK region; Fig. 4-5B) are conserved amino acids in the GSG domain, and missense mutations for these amino acids have been isolated in GLD-1 (Jones and Schedl, 1995). Mutation of Sam68:A247T and Sam68:D262N displayed wild-type cellular Sam68 distribution, suggesting that these amino acid changes had no effect on Sam68 protein localization.

In class II, ~90% of the Sam68:P168L-transfected cells displayed a wild-type Sam68 localization pattern (pattern A) and ~10% of cells concentrated Sam68:P168L in 10 to 30 SNBs (pattern B). Sam68 proline 168, located in loop 1 of the Sam68 KH domain, is fully conserved in the GSG domain, and this point mutation isolated in GLD-1 reverses the gain-of-function sex determination phenotype of the glycine 248-to-arginine mutation and displays the loss-of-function defective oogenesis (Jones and Schedl, 1995). The GFP-Sam68:P168L protein implicates the KH domain loop 1 in Sam68 cellular localization.

In class III, ~90% of the cells transfected with Sam68ΔL1, Sam68:NF→DL, Sam68ΔL1NF or Sam68:R204C displayed exclusive SNBs (pattern B) and ~10% of the cells showed a wild-type Sam68 pattern (pattern A). In addition to the six-amino acid

deletion in KH domain loop 1, asparagine 171 and phenylalanine 172 were substituted to aspartic acid and leucine, respectively, in Sam68ΔL1. Therefore, to determine whether it was the deletion or the substitutions that altered Sam68 localization, we constructed a protein containing only the deletion (Sam68ΔL1NF) and another harboring the two amino acid substitutions (Sam68:NF→DL). Either the substitution of fully conserved amino acids 171 and 172 (Sam68:NF→DL) or the deletion of loop 1 (Sam68ΔL1NF) was sufficient to concentrate the proteins exclusively in SNBs. These data demonstrate that Sam68 KH domain loop 1 is involved in protein localization. We have shown that Sam68ΔL1 has impaired self-association capabilities, with no effect on RNA binding (Chen et al., 1997). Because the RNA-binding properties of Sam68ΔL1 are normal, this suggests that the protein localization property of the Sam68 GSG domain is separate from its RNA-binding activity. The substitution of R204C in Sam68 KH domain loop 4 also resulted in ~90% of the cells containing only SNBs. This missense mutation identified in Who/How has been shown to result in flies with the wings-held-out phenotype (Baehrecke, 1997). Interestingly, this missense mutation, when introduced in Who/How, did not alter RNA binding and self-association (Chen and Richard, 1998), suggesting that the molecular defect of this missense mutation may be in protein localization.

In class IV, cells transfected with Sam68ΔKH, Sam68ΔKH isoform, Sam68ΔL4, or Sam68:2G→R displayed three major expression patterns: ~30% of the cells displayed only SNBs (pattern B), ~50% of cells showed a punctate staining throughout the entire cell (pattern C), and the remaining ~20% had distinct cytoplasmic punctate staining (pattern D). The cytoplasmic punctate staining observed may represent Sam68 localization to specific organelles. Using double staining with Lysotracker Red, we have observed that the Sam68 punctate staining does not represent lysosomes (data not shown). Sam68ΔKH contains a 60-amino acid deletion that removes a large portion of the KH domain. This deletion completely altered the localization of Sam68, as none of the cells expressing Sam68ΔKH displayed wild-type Sam68 expression pattern. A rare natural isoform of Sam68, Sam68ΔKH isoform, has been discovered and contains a 39-amino acid deletion in the

Sam68 KH domain (Barlat et al., 1997). A GFP fusion protein containing this isoform had similar expression patterns as those observed with GFP-Sam68ΔKH. The deletion of KH domain extended loop 4 (Sam68ΔL4) or the replacement of glycines 199 and 201 in loop 4 with arginines (Sam68:2G→R) also exhibited expression patterns identical to those observed with Sam68ΔKH. These data further implicate the KH domain in protein localization, with a specific requirement for loop 4. Our findings are consistent with previous results showing that deletion of the Sam68 KH domain results in accumulation in nuclear dots (McBride et al., 1998).

In class V, the cellular localization of two RNA-binding defective Sam68 proteins was examined. One mutation altering the equivalent substitution of GLD-1 glycine 227 to aspartic acid (Jones and Schedl, 1995) and the second altering the equivalent substitution of FMR1 isoleucine 304 to asparagine (DeBoulle et al., 1993) were introduced in Sam68. GFP-Sam68:G→D and GFP-Sam68:I→N displayed expression patterns similar to those observed with GFP-Sam68ΔKH, except a larger percentage of cells showed a cytoplasmic distribution (pattern D) and a smaller percentage of cells showed an entire cell distribution (pattern C). In addition, a considerable number of cells (~25% for Sam68:G→D and ~7% for Sam68:I→N) displayed a cytoplasmic fibrous pattern characteristic of microtubules that was unique to these two Sam68 proteins (pattern E). The GLD-1 glycine 227-to-aspartic acid point mutation has been shown to result in germline tumors (Jones and Schedl, 1995). The human FMR1 isoleucine 304-to-asparagine results in severe mental retardation (DeBoulle et al., 1993). FMR1:I304N has been shown to alter the structure of the KH domain (Musco et al., 1996), impair the ability of FMR1 to associate with RNA (Siomi et al., 1994), and prevent FMR1 from associating with polyribosomes (Feng et al., 1997). We have previously shown that the introduction of these point mutations in Sam68 alters its ability to bind homopolymeric and cellular RNA (Chen et al., 1997). Because both Sam68:G→D and Sam68:I→N self-associated (Chen et al., 1997), we reasoned that the altered protein localization patterns observed were independent of self-association.

A

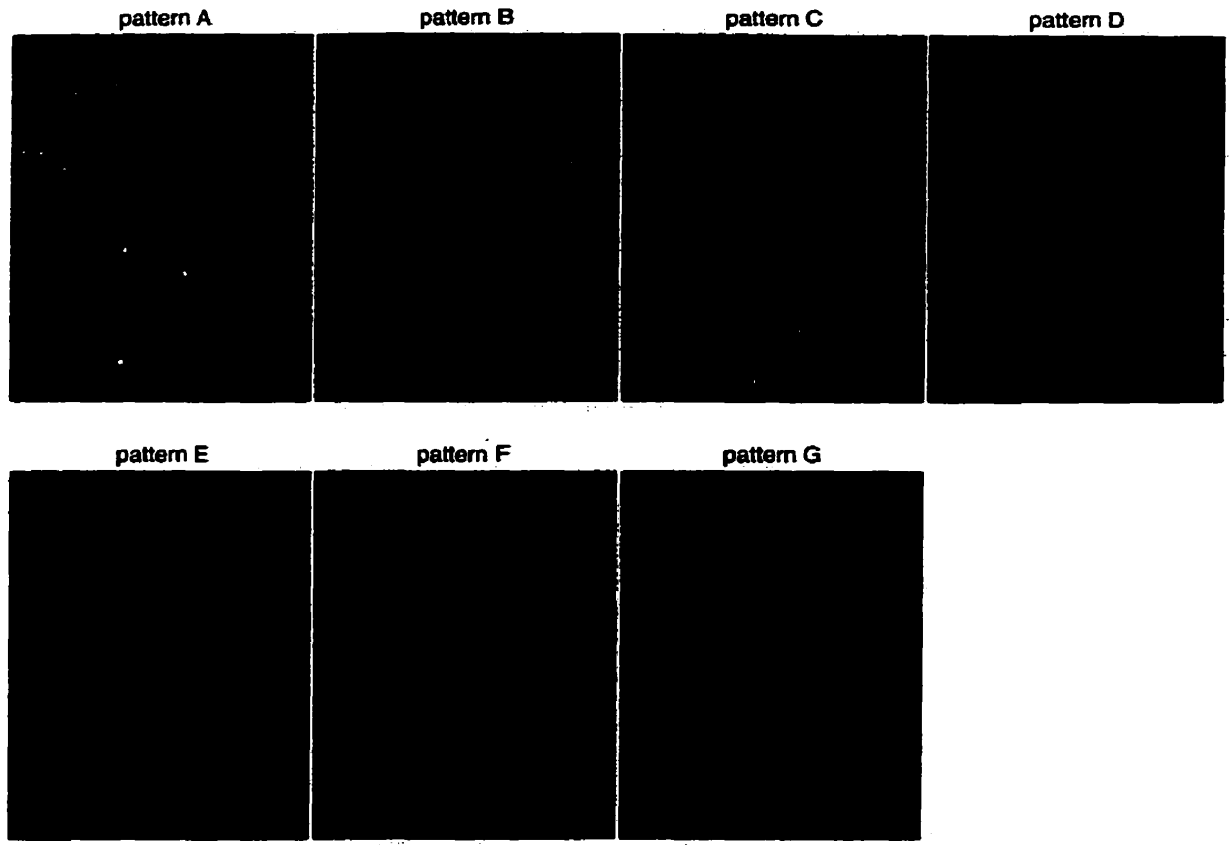
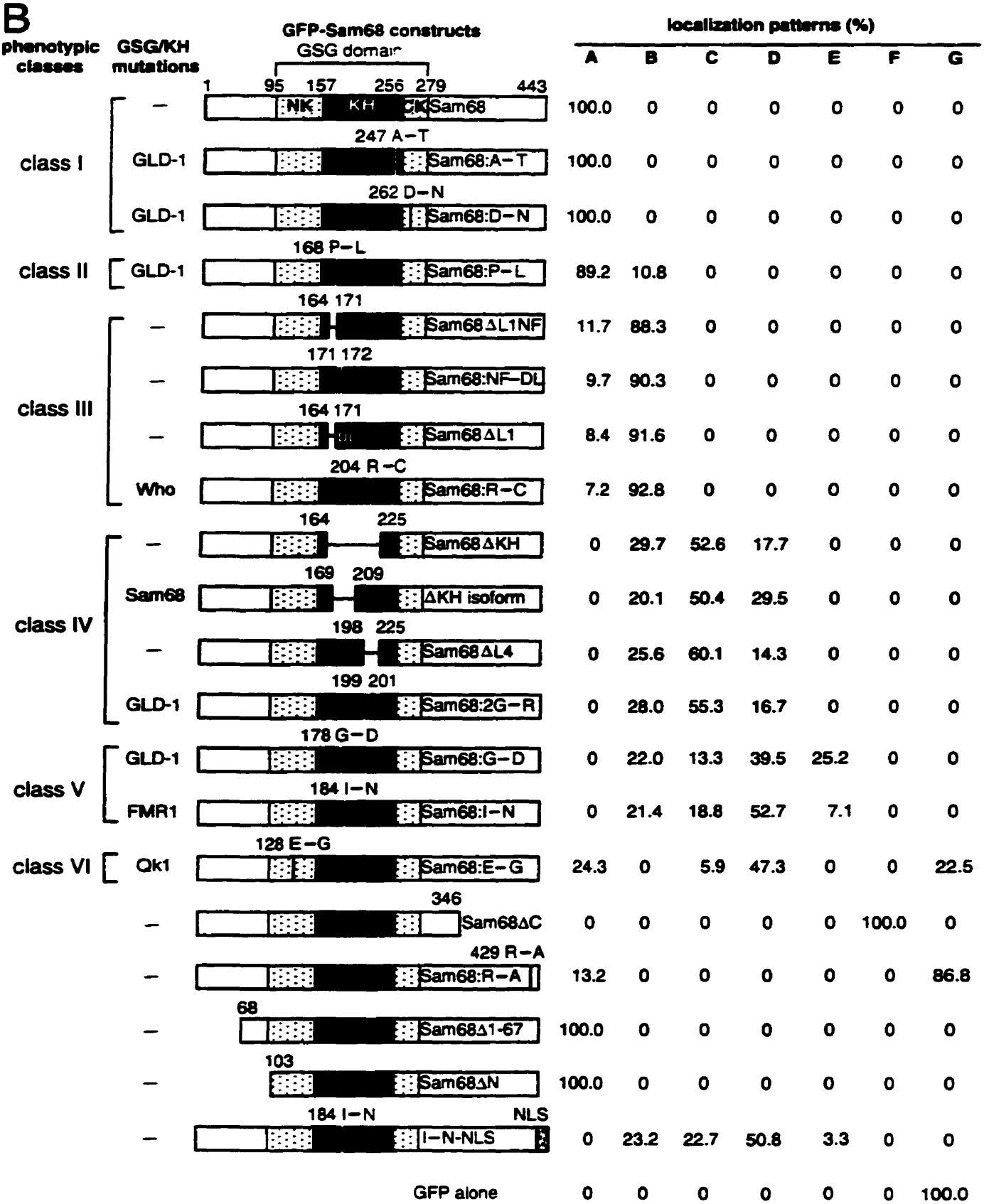


FIG. 4-5. Amino acid substitutions or deletions in the Sam68 GSG domain alter Sam68 localization. (A) GFP-Sam68 constructs were transfected into HeLa cells, and 12 h after transfection the protein expression patterns were analysed by fluorescence microscopy. The GFP fusion partner was located at the N terminus. The typical localization patterns observed are shown: diffuse nuclear staining (excluding nucleolus) with several SNBs (pattern A); exclusive SNB accumulation (pattern B); punctate nuclear and cytoplasmic staining (pattern C); punctate cytoplasmic staining (pattern D); fibrous structure (pattern E); diffuse cytoplasmic staining (pattern F); and diffuse nuclear and cytoplasmic staining (pattern G). (B) (next page) Schematic diagram of GFP-Sam68 constructs and the quantitation of different localization patterns. In the diagrams of the constructs, the GSG domain is depicted with a bracket. The contains an extended KH domain (denoted by the black box). The regions in the GSG domain at the N terminus of the KH domain and at the C terminus of the KH domain were called the NK and CK regions, respectively. The genes in which the genetic missense mutations were identified that alter amino acids in the GSG or KH domain are shown on the left. Δ indicates a deletion; ➔ represents an amino acid substitution; and a horizontal or vertical thin line denotes the position of a deletion or point mutation. The NLS represents the SV40 large T antigen nuclear localization sequence (PKKKRKV). For each construct, an average of 250-300 transfected (green) cells from three separate experiments were counted, and the localization patterns were expressed as percentages. The Sam68 proteins harboring amino acid substitutions or deletions in their GSG domains were grouped into six phenotypic classes based on their localization patterns.

B



phenotypic classes	GSG/KH mutations	localization patterns (%) A	B	C	D	E	F	G
class I	–	100.0	0	0	0	0	0	0
	GLD-1	100.0	0	0	0	0	0	0
	GLD-1	100.0	0	0	0	0	0	0
class II	GLD-1	89.2	10.8	0	0	0	0	0
class III	–	11.7	88.3	0	0	0	0	0
	–	9.7	90.3	0	0	0	0	0
	–	8.4	91.6	0	0	0	0	0
	Who	7.2	92.8	0	0	0	0	0
class IV	–	0	29.7	52.6	17.7	0	0	0
	Sam68	0	20.1	50.4	29.5	0	0	0
	–	0	25.6	60.1	14.3	0	0	0
	GLD-1	0	28.0	55.3	16.7	0	0	0
class V	GLD-1	0	22.0	13.3	39.5	25.2	0	0
	FMR1	0	21.4	18.8	52.7	7.1	0	0
class VI	Qk1	24.3	0	5.9	47.3	0	0	22.5
	–	0	0	0	0	0	100.0	0
	–	13.2	0	0	0	0	0	86.8
	–	100.0	0	0	0	0	0	0
	–	100.0	0	0	0	0	0	0
	–	0	23.2	22.7	50.8	3.3	0	0
	GFP alone	0	0	0	0	0	0	100.0

In class VI, the lethal point mutation identified in the Qk1 GSG domain was introduced in the N-terminal-to-KH-domain (NK) region of Sam68 (Sam68:E→G). GFP-Sam68:E→G also exhibited several patterns of expression. A large fraction of cells (~47%) showed punctate cytoplasmic staining (pattern D) and a small fraction (~6%) showed punctate staining throughout the entire cell (pattern C). Additionally, many cells (~24%) displayed a wild-type Sam68 expression pattern (pattern A) and an equal number of cells (~22%) displayed a diffuse staining in both the nucleus and the cytoplasm (pattern G). Because the Sam68:E→G protein has wild-type RNA binding and self-association properties (Chen et al., 1997), the localization property of the Sam68 GSG domain is separate from its ability to bind RNA and self-associate.

The Sam68 C Terminus and the GSG Domain Are Both Required for Nuclear Localization. The primary amino acid sequence of Sam68 does not reveal any known NLS or nuclear export sequence. A putative NLS has been identified at the C terminus of Sam68, and the substitution of Sam68 arginine 429 with alanine in the putative NLS has been shown to result in a diffuse distribution throughout the entire cell (Ishidate et al., 1997). To examine the role of this putative NLS in Sam68 protein localization, the C terminus of Sam68 was truncated to remove the tyrosine-rich region and the putative NLS (Sam68ΔC). The plasmid encoding this protein was transfected in HeLa cells, and GFP-Sam68ΔC was visualized by fluorescence microscopy. Transfected cells expressing GFP-Sam68ΔC displayed diffuse cytoplasmic staining (Fig. 4-5A, pattern F). Similar results were obtained when only the last 4 or 11 amino acids were deleted at the C terminus of Sam68 (data not shown). Moreover, the substitution of Sam68 arginine 429 with alanine (Sam68:R→A) resulted in a distribution pattern similar to GFP alone (Fig. 4-5A, pattern G), consistent with previous studies (Ishidate et al., 1997). In contrast, deletion of 67 or 102 amino acids at the N terminus of Sam68 (Sam68Δ1-67 and Sam68ΔN) showed no effect on Sam68 localization (Fig. 4-5B). These results demonstrate that the C-terminal region of the Sam68, but not the N-terminal region, plays a role in Sam68 protein localization. To

determine whether the GSG domain played a dominant role in protein localization, a sequence encoding the SV40 large T antigen NLS was introduced at the C terminus of Sam68:I→N (Sam68:I→N-NLS). Interestingly, this protein was not localized to the nucleus but displayed the typical pattern observed with Sam68: I→N (Fig. 4-5B). As a control, the introduction of the SV40 large T antigen NLS at the C terminus of GFP targeted the fusion protein to the nucleus (data not shown). These data suggest that the GSG domain plays a dominant role in Sam68 localization but also requires the C terminus of Sam68 for proper localization.

Sam68:G→D and Sam68:I→N Associate with Microtubules. The fibrous pattern observed with Sam68:G→D and Sam68:I→N was characteristic of microtubules (Fig. 4-5A, pattern E). To verify that these proteins were associated with microtubules, HeLa cells transfected with Sam68:G→D were immunostained with anti-tubulin antibodies and colocalization of Sam68 fibers with tubulin was analysed with confocal microscopy. As shown in Fig. 4-6A, some Sam68 fibers colocalized with tubulin. The tubulin staining was much weaker than the Sam68 staining, which may explain the partial colocalization with tubulin. To further confirm that the Sam68 fibers associated with microtubules, the effect of the microtubule-disrupting drug nocodazole on the integrity of these fibers was examined. Cells transfected with Sam68:G→D displaying a fibrous pattern were localized and photographed live before and after nocodazole treatment. After the addition of nocodazole for 15 and 30 min, the fibers completely disassembled (Fig. 4-6B), suggesting that Sam68:G→D was associated with the microtubules or microtubule-associated proteins.

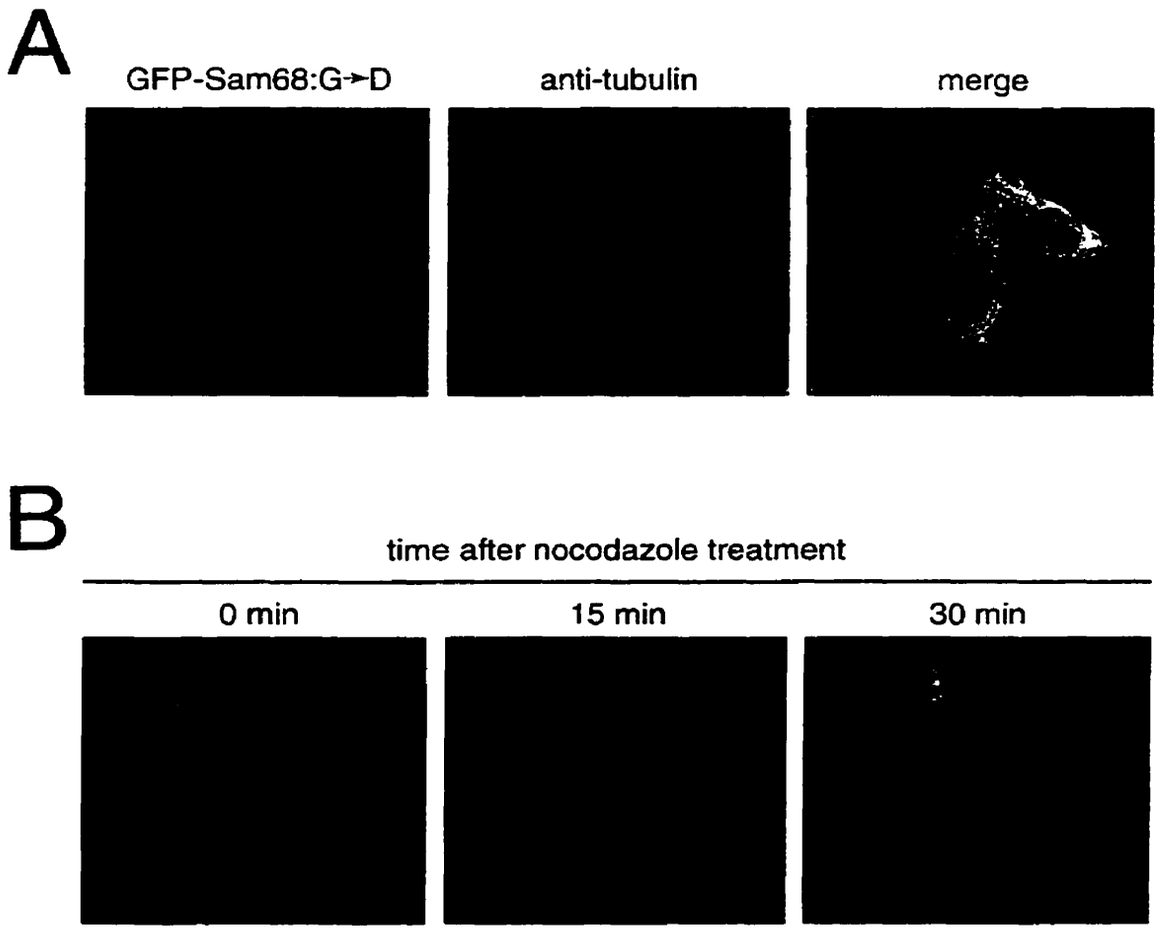


FIG. 4-6. Sam68:G→D and Sam68:I→N associate with microtubules. (A) HeLa cells transfected with GFP-Sam68:G→D were fixed, permeabilized and immunostained with anti-tubulin antibody followed by a rhodamine-conjugated goat-anti-mouse secondary antibody, and then analysed by confocal microscopy. Co-localization of Sam68 fibers (green) with microtubule fibers (red) results in yellow color when the confocal images of GFP-Sam68:G→D and anti-tubulin immunostaining were merged. (B) HeLa cells transfected with GFP-Sam68:G→D were incubated with 40 ng/ml nocodazole and cells with fibrous phenotype were photographed live before (0 min) and 15 or 30 min after the addition of nocodazole.

Characterization of the GFP-Sam68ΔL1 Nuclear Bodies. We wanted to verify that the nuclear bodies observed with GFP-Sam68ΔL1 and other Sam68 proteins harboring GSG mutations/deletions behaved like endogenous SNBs. Initially, we examined whether the GFP-Sam68ΔL1 bodies were nuclear structures. Difference interference contrast microscopy of HeLa cells transfected with GFP-Sam68ΔL1 demonstrated that the nuclear bodies generated were bona fide nuclear structures (Fig. 4-7A). The colocalization of GFP-Sam68ΔL1 nuclear bodies with other nuclear bodies was examined. Confocal microscopy revealed that GFP-Sam68ΔL1 nuclear bodies did not colocalize with coiled bodies, gems, PML nuclear bodies, BCL6/SMRT nuclear structures, and fibrillarin, nor did they cause any of these structures to reorganize (data not shown).

The dynamics of the GFP-Sam68ΔL1 nuclear bodies were examined. As has been shown with endogenous SNBs, the GFP-Sam68ΔL1 nuclear bodies disassembled with actinomycin D treatment (Fig. 4-7B). Coiled bodies and gems have been shown to increase in size at lower temperatures (Carmo-Fonseca et al., 1993; Liu and Dreyfuss, 1996). HeLa cell transfected with GFP-Sam68ΔL1 were incubated at 32°C, and the nuclear bodies were clearly larger in size (Fig. 4-7B), suggesting that GFP-Sam68ΔL1 nuclear bodies behave like other nuclear bodies, such as coiled bodies and gems. It is interesting to note that proteins such as GFP-Sam68ΔL1 that gave a Sam68-exclusive nuclear body pattern were not expressed diffusely in the nucleoplasm 12-18 h after transfection. To further characterize the dynamic nature of GFP-Sam68ΔL1 nuclear bodies, cells transfected with GFP-Sam68ΔL1 were examined for 60 h (Fig. 4-7C). At 12 h after transfection, GFP-Sam68ΔL1-transfected cells contained ~30 GFP-Sam68ΔL1 nuclear bodies. The size of these nuclear bodies increased with time, whereas the number decreased to <20 at 24-36 h after transfection. At 48-60 h, the GFP-Sam68ΔL1 nuclear bodies gradually disappeared and the nuclei displayed a diffuse extranucleolar staining, as observed with wild-type Sam68 (Fig. 4-7C). These findings further suggested that there exists a Sam68 equilibrium between SNBs and the nucleoplasm. This equilibrium is shifted in favor of the SNBs when Sam68 has certain amino acid substitutions or deletions in the GSG domain. Taken

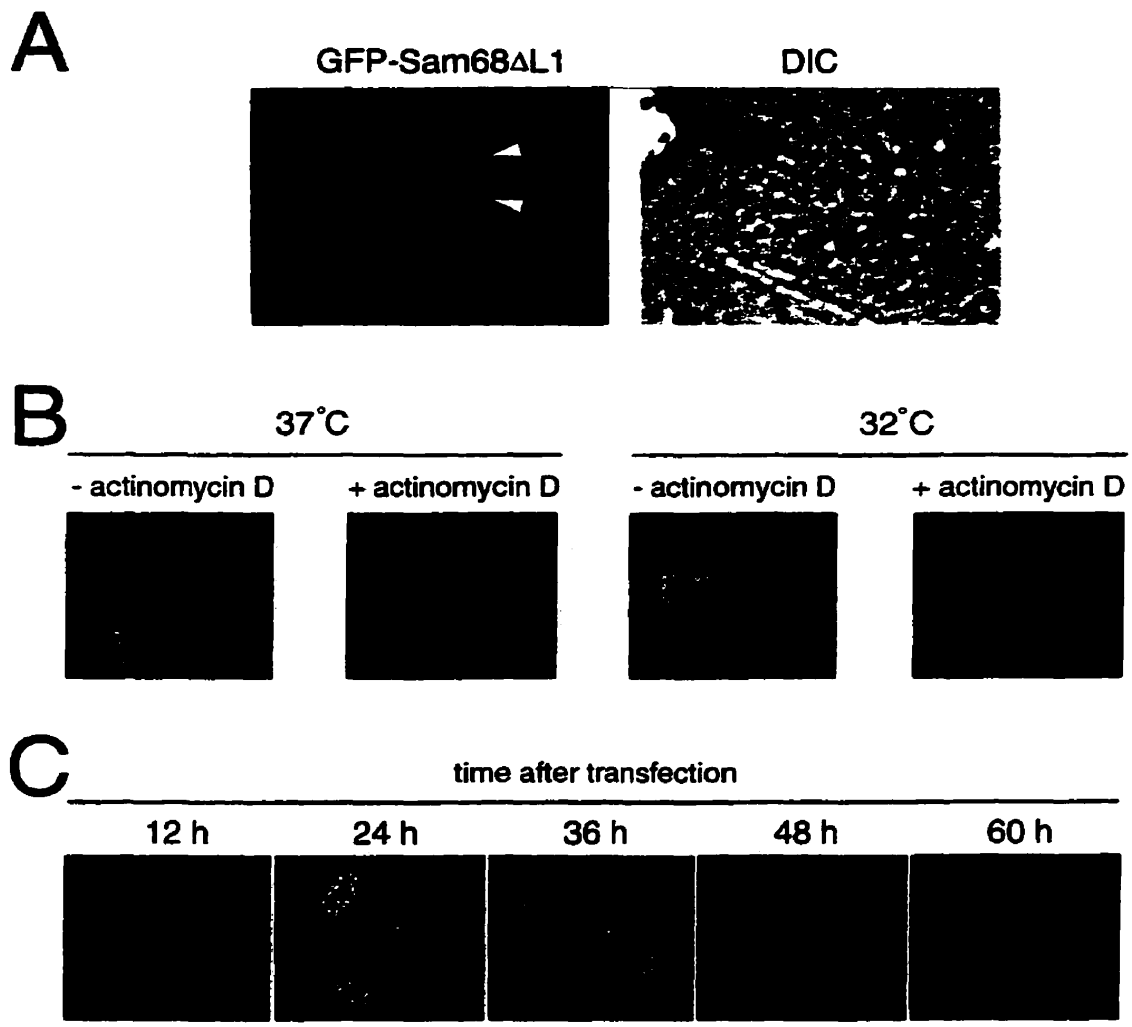


FIG. 4-7. Characterization of GFP-Sam68ΔL1 nuclear bodies. (A) HeLa cells transfected with GFP-Sam68ΔL1 were fixed and analyzed with fluorescence and difference interference contrast (DIC) microscopy. (B) GFP-Sam68ΔL1 nuclear bodies are dynamic structures. HeLa cells were transfected with GFP-Sam68ΔL1 and then incubated at 37°C or 32°C for 24 h; the last 3 h of incubation was performed with or without 5 mg/ml actinomycin D. The cells were fixed and visualized by fluorescent microscopy. (C) HeLa cells transfected with GFP-Sam68DL1 were fixed and observed 12, 24, 36, 48, or 60 h after transfection.

together, these data demonstrate that Sam68ΔL1 nuclear bodies share identical properties to endogenous SNBs.

SNBs Do Not Contain Nascent RNAs and snRNPs. To determine whether SNBs are involved in transcription, we performed in situ Br-UTP incorporation experiments (Fig. 4-8A). HeLa cells transfected with GFP-Sam68 (Fig. 4-8A, a-c), GFP-Sam68ΔL1 (Fig. 4-8A, d-f), or GFP-PTB (Fig. 4-8A, g-i) were briefly permeabilized and incubated in a transcription mixture containing Br-UTP. Then, the cells were fixed and the sites of transcription were detected by immunolabeling with a mAb that recognizes Br-UTP followed by a rhodamine-conjugated goat anti-mouse secondary antibody. Confocal microscopy revealed that both wild-type Sam68 and Sam68ΔL1 nuclear bodies did not colocalize with the major sites of Br-UTP incorporation (Fig. 4-8A, a-c and d-f). As a positive control, we used GFP-PTB, which has been shown to colocalize in the perinucleolar compartment with newly synthesized RNA (Huang et al., 1998). As expected, GFP-PTB colocalized with sites of Br-UTP incorporation (Fig. 4-8A, g-i; colocalization indicated by arrows in i). The specificity of the Br-UTP incorporation assay was demonstrated by the fact that treatment of cells with actinomycin D prevented Br-UTP incorporation (data not shown). These findings indicate that SNBs do not represent major sites of transcription.

Coiled bodies have been found to contain spliceosomal snRNPs (Carmo-Fonseca et al., 1993). Although gems do not contain snRNPs (Liu and Dreyfuss, 1996), as observed with immunofluorescence with the anti-Sm antibody Y12 (Petterson et al., 1984), SMN has been shown to interact directly with Sm core proteins (Liu et al., 1997). Moreover, a truncation of the N-terminal 27 amino acids of SMN causes the reorganization of snRNPs in the nucleus (Pellizzoni et al., 1998). To determine whether SNBs contain splicing factors, HeLa cells were double immunostained with anti-Sam68 AD1 (Fig. 4-8Ba) and anti-SC35 antibodies (Fig. 4-8Bb). The merged confocal image revealed that SNBs did not colocalize with SC35 (Fig. 4-8Bc). HeLa cells expressing GFP-Sam68ΔL1 were also immunostained with the anti-SC35 and analyzed by confocal microscopy (Fig. 4-8B, d-f). The Sam68ΔL1 nuclear

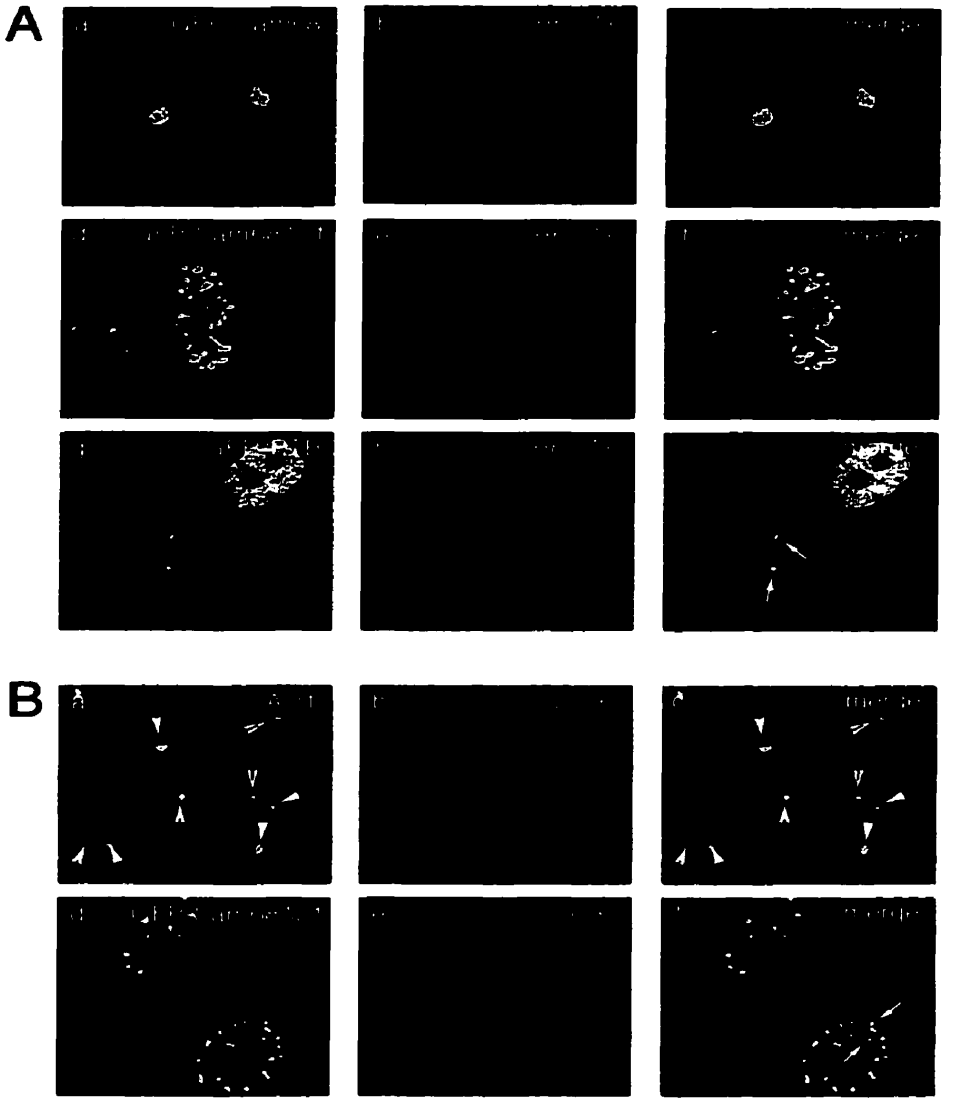


FIG. 4-8. Sam68 nuclear bodies do not contain nascent RNAs and snRNPs. (A) HeLa cells transfected with GFP-Sam68 (a-c), GFP-Sam68ΔL1 (d-f), or GFP-PTB (g-i) were permeabilized with 50 mg/ml saponin for 5 min at 4°C and then incubated with a transcription cocktail containing Br-UTP for 20 min at 33°C. The cells were fixed, and the sites of Br-UTP incorporation were detected by immunolabeling using anti-BrdU antibody (which also recognizes Br-UTP) followed by a rhodamine-conjugated secondary antibody (b, e, and h). Colocalization was determined by confocal microscopy (c, f, and i). Arrows in panel i indicate co-localization of PNC with sites of Br-UTP incorporation. The arrows in panel i indicate colocalization of PNC with sites of Br-UTP incorporation. (B) HeLa cells (a-c) or HeLa cells transfected with GFP-Sam68ΔL1 (d-f) were fixed, permeabilized, and immunostained with anti-SC35 antibody followed by a secondary antibody conjugated to rhodamine (b and e). Endogenous Sam68 nuclear bodies were detected by immunolabeling using anti-Sam68 AD1 antibody followed by FITC-conjugated goat anti-rabbit antibody (a). Colocalization of SNBs and SC35 was analyzed by confocal microscopy (c and f). Endogenous SNBs in panels a and c are indicated by arrowheads. The arrows in panel f indicate partial overlap of Sam68ΔL1 SNBs with SC35, most likely random overlapping.

bodies did not colocalize with SC35, although a few bodies seemed to partially overlap with SC35 (Fig. 4-8B; indicated by arrows in f), which we suspect represent random overlapping because of the numerous bodies in both fields (Fig. 4-8Bf). Moreover, both wild-type Sam68 and Sam68ΔL1 nuclear bodies were not stained with the anti-Sm antibody Y12 (data not shown). Taken together, these findings suggest that SNBs do not contain snRNPs. In addition, overexpression of Sam68ΔL1 did not change the pattern of anti-SC35 (Fig. 4-8B; compare b and e), anti-coilin, and anti-SMN (data not shown) immunostaining. These findings suggest that Sam68ΔL1, in spite of inducing more SNBs, has no effect on the distribution and organization of coiled bodies, gems, and snRNPs.

Structure of the Sam68/SLM Nuclear Bodies by Correlative Microscopy. We used the technique of electron spectroscopic imaging to map protein-based and nucleic acid-based regions in and around SNBs. This technique allows comparison of nitrogen and phosphorus maps, providing the ability to directly distinguish protein from nucleic acids in situ based on elemental composition, without specific stains or labels (Hendzel et al., 1999). To identify the nuclear bodies in the electron microscope images, we used a correlative fluorescence and electron microscopy approach. The sections of fluorescently labeled cells were first imaged in the fluorescence microscope to identify the nuclear bodies. Subsequently, the same sections were imaged by electron spectroscopic imaging (Hendzel et al., 1999). We have found that Sam68 nuclear bodies are large, spherical or ovoidal structures of ~0.6 x 1 μm (Fig. 4-9). The SNBs are enriched in phosphorus-rich and nitrogen-rich fibers and granules, indicating the presence of nucleic acids (Fig. 4-9, Net P and Net N images). These data further suggest that Sam68/SLM structures are nuclear bodies and that these structures are enriched in nucleic acids that may represent RNA.

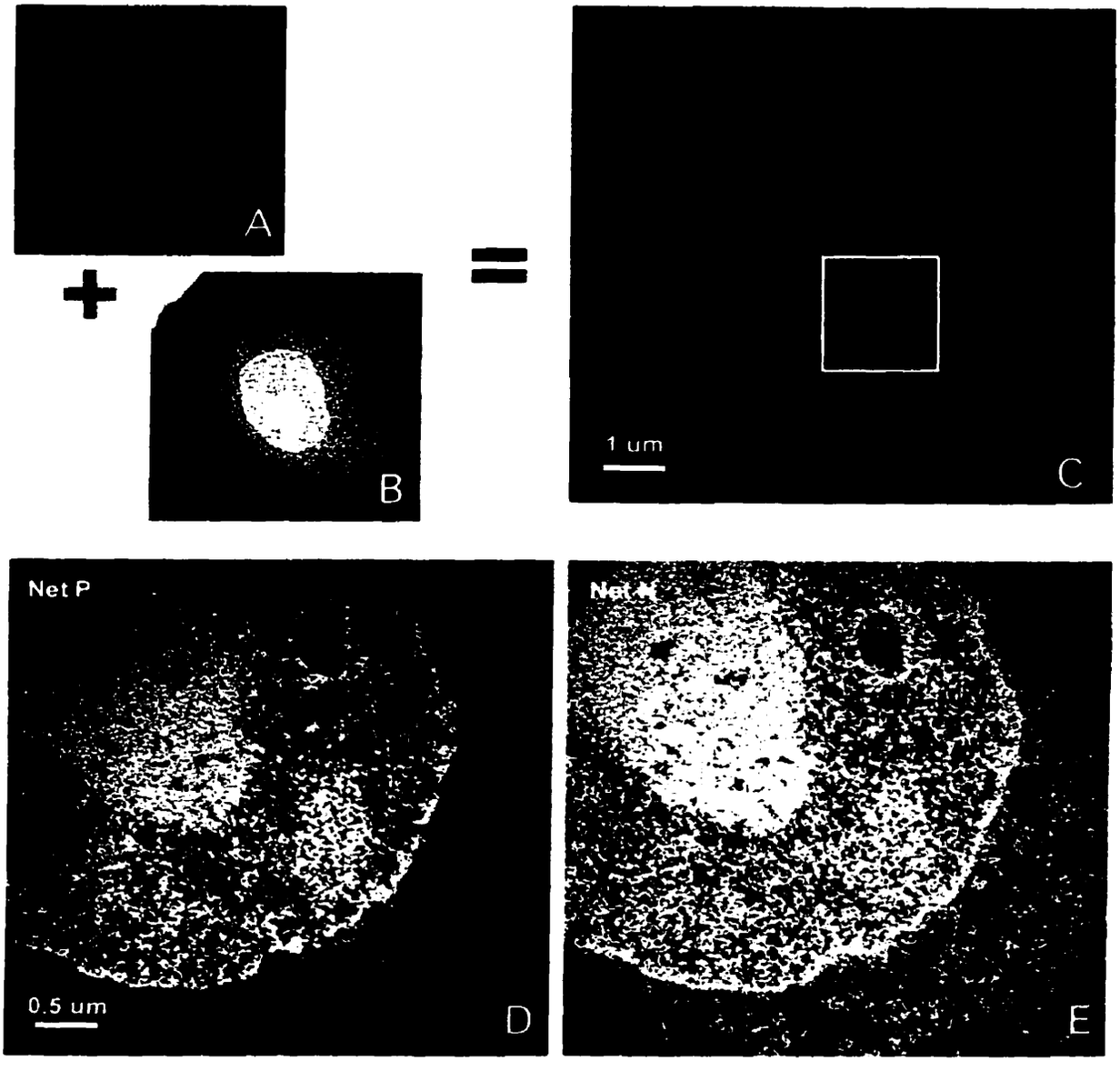


FIG. 4-9. Correlative microscopy of Sam68 nuclear bodies. An ultrathin section (30 nm) of HeLa cells, previously labeled with anti-Sam68 AD1 antibody and embedded for electron microscopy, was examined under an immunofluorescence microscope (A) and then an electron microscope (B). The respective images were resized and rotationally aligned before being merged (C). It was then possible to identify the locations of structures labeled by immunofluorescence and to characterize them, by electron spectroscopic imaging, for phosphorus content (D) and nitrogen content (E). (Magnification: B and C, x3,000; D and E, x12,000).

DISCUSSION

We demonstrate that endogenous Sam68 localizes into novel nuclear structures that we have named SNBs for Sam68/SLM nuclear bodies. Alterations in the GSG domain resulted in several Sam68 cellular patterns, including exclusive SNB accumulation, microtubule association, diffuse cytoplasmic staining, and whole cell and cytoplasmic punctate staining. These observations implicate the Sam68 GSG domain in protein localization. The protein localization property was separate from the other GSG properties such as RNA binding and self-association, because Sam68 proteins defective in RNA binding (e. g. Sam68:G→D and Sam68:I→N; Chen et al., 1997), self-association (e. g. Sam68ΔL1; Chen et al., 1997), or both (e. g. Sam68ΔKH, Sam68ΔL4, and Sam68:2G→R; Chen et al., 1997) displayed altered localization. The localization property of the GSG domain is a bona fide role for this protein module. Thus, genetic mutations in the GSG domain can alter RNA binding, self-association, and/or the localization of a protein. The C-terminal tyrosine-rich region of Sam68 was required for proper Sam68 localization, suggesting that tyrosine phosphorylation of the C terminus of Sam68 may regulate its localization as it regulates RNA binding and self-association (Wang et al., 1995; Chen et al., 1997). Although RNA binding, self-association, and protein localization are separable properties of the GSG domain, they most likely function together to transport RNAs as protein multimers. Recently, Sam68 has been shown to function like HIV-1 Rev and may be its cellular homologue (Reddy et al., 1999). These findings are consistent with our data that Sam68 may be involved in transport of cellular RNAs.

It is not understood why mutations in the Sam68 GSG domain localize the mutant proteins to different cellular locations. The most likely explanation and the one we favor is that some Sam68 proteins containing amino acid substitutions or deletions are trapped in certain locations or 'travel' slowly inside the cell. This allows these mutant Sam68 proteins to concentrate in areas where Sam68 would normally be found transiently. The patterns would then represent 'snapshots' of the different journeys of Sam68 within the cell. It is also possible that the location of Sam68 mutants may be cell cycle regulated and may

represent different locations of Sam68 during the cell cycle. The presence of Sam68 in both the cytoplasm and the nucleus suggest that it shuttles like other RNA-binding proteins, including hnRNP K and A1 (Pinol-Roma and Dreyfuss, 1992; Michael et al., 1997). A high-affinity RNA-binding motif (UAAA), present in most if not all mRNAs, has been identified for Sam68 (Lin et al., 1997). This information, coupled with the possibility that Sam68 shuttles between the nucleus and the cytoplasm, raises the question of whether Sam68 functions as a general transporter protein for RNAs. The presence of Sam68 in specialized nuclear structures suggests that Sam68 also have a specialized function in RNA metabolism, such as snRNP biogenesis or pre-mRNA processing. The localization of Sam68 to microtubules also raises the possibility that Sam68 may be transiently associated with the mitotic spindle or that Sam68 bridges RNAs to microtubules or travels along the microtubules. A bridging function was recently shown for the *Xenopus* KH domain protein Vg1 RBP that bridges the Vg1 mRNA to microtubules (Havin et al., 1998).

The expression of GFP-SLM-1 and GFP-SLM-2 in HeLa cells resulted in colocalization with Sam68 in SNBs, suggesting that SLM-1 and SLM-2 are also components of SNBs. Interestingly, unlike Sam68 and SLM-1, GFP-SLM-2 was also localized or even concentrated in the nucleoli of ~60% of the transfected cells. This raises the possibility that SLM-2 as well as SNBs may be involved in some aspects of nucleolar functioning, such as rRNA and ribosome biogenesis. The observation that SNBs are frequently found in close proximity to nucleoli also supports this possibility.

The function of most nuclear bodies is still largely unknown. Many components have been identified in coiled bodies and gems to suggest that these nuclear structures function in some aspect of snRNP biogenesis or pre-mRNA processing (Lamond and Earnshaw, 1998). The absence of SC35 and Sm proteins in SNBs, as detected by immunofluorescence with anti-SC35 and the Y12 anti-Sm antibodies (data not shown), suggests that snRNPs are not present in SNBs. Because SMN has been shown to associate directly with Sm core proteins (Liu et al., 1997), we performed coimmunoprecipitation experiments with individual Sm core proteins to further examine the possibility that Sm proteins associated with Sam68.

No convincing association was observed between Sam68 and individually overexpressed Sm proteins, including HA epitope-tagged Sm-B, Sm-D1, Sm-D3, Sm-E, Sm-F, and Sm-G (data not shown). These Sm proteins are known to be the major protein components of the snRNPs (Mattaj and DeRobertis, 1985). Thus, we conclude that Sam68 does not associate directly with Sm core proteins and that SNBs, unlike coiled bodies, do not contain snRNPs.

SNBs are primarily present in cancer cells, suggesting that they may be linked to cell transformation. Either SNBs contribute to the transformed phenotype or they are a consequence of transformation. The fact that SNB prevalence varies significantly among transformed cells indicates that the presence of SNBs is cell type specific and/or may be related to the degree of malignancy. The latter possibility is supported by the observation that SNB prevalence correlated with the differentiation status and tumorigenicity of the three breast cancer cell lines examined. BT-20 cells, a poorly differentiated cell line that induces tumors in nude mice, has a much higher SNB prevalence (~90%) than Hs578T cells (~50%), which are not tumorigenic in nude mice. On the other hand, MCF-7 cells, a well-differentiated cell line, had the lowest SNB prevalence, with ~5% of the cells containing SNBs. SNBs may be a useful marker for certain cancers.

The importance of nuclear bodies is highlighted by their linkage to certain diseases. The number of coiled bodies or PNCs is significantly increased in transformed cells (Spector et al., 1992; Huang et al., 1997). Gems contain SMN (Liu and Dreyfuss, 1996), the protein responsible for spinal muscular atrophy (Bussaglia et al., 1995; Lefebvre et al., 1995). PML nuclear bodies are disorganized in patients with acute promyelocytic leukemia (Dyck et al., 1994; Weis et al., 1994; Doucas et al., 1996). *BRCA1* and *BRCA2*, known tumor suppressor genes that account for most cases of familial, early-onset breast and/or ovarian cancer, localize into nuclear bodies (Miki et al., 1994; Wooster et al., 1995). The oncoprotein LAZ3/BCL6 and the corepressor SMRT have been shown to colocalize to nuclear bodies (Dhordain et al., 1997). Disruption of the *LAZ3/BCL6* gene because of the chromosomal translocation 3q27 causes diffuse large cell lymphomas (Kerckaert et al., 1993; Ye et al., 1993). Intranuclear inclusions composed of mutant proteins have been

found in cells affected by several neurodegenerative diseases, including huntington's disease and type 1 spinocerebellar ataxia (Sisodia, 1998).

HeLa cells transfected with GFP-Sam68 have a diffuse nuclear staining with 2-5 SNBs. On the other hand, HeLa cells transfected with GFP-Sam68ΔL1 have 10-30 SNBs with no diffuse nuclear staining. Because the Sam68 GFP fusion proteins were expressed to similar levels, as detected by Western blotting (data not shown), this suggests that all the GFP-Sam68ΔL1 concentrates in SNBs. The question that arises is whether new SNBs are created or existing undetectable SNBs increase in size and become visible. The fact that SNBs generated by GFP-Sam68ΔL1 relocalize to diffuse nuclear staining after 48-60 h suggests that SNBs are in equilibrium with the nucleoplasm. The labeling with Br-UTP demonstrates that SNBs are not the site of active transcription, as has been shown for the perinucleolar compartment (Huang et al., 1998). However, SNBs do contain nucleic acids that may be RNA, as detected using correlative electron microscopy (Fig. 4-9). Because we have shown that Sam68 oligomerization is RNA dependent (Chen et al., 1997), it is possible that SNBs form only in cells that contain specific Sam68/SLM RNA targets.

Previous immunofluorescence studies on Sam68 did not reveal SNBs in NIH 3T3 cells (Ishidate et al., 1997). Although there are several factors that can influence the presence and detection of SNBs, such as the antibody, fixation, and cell type, we believe the major reason in this study was the use of NIH 3T3 cells. We found that overexpression of GFP-Sam68 in NIH 3T3, REF-52 or COS cells does not result in the appearance of SNBs, consistent with the data reported by Ishidate et al. (1997). Another Sam68 immunofluorescence study using HeLa cells and NIH 3T3 cells did not detect SNBs (McBride et al., 1998). In that study, the cells were fixed with methanol prior to staining with anti-Sam68 antibodies. We found that such a fixation procedure results in a granular nuclear staining pattern and made SNBs hard to detect (data not shown). Furthermore, McBride et al. (1998) showed that Sam68 localized in punctate nuclear structures in HeLa cells upon treatment with transcription inhibitors. These data contradict our observations that transcription inhibitors, including actinomycin D and α-amanitin, disassemble SNBs. Because we show that SNBs

disassemble with transciption inhibitors in several situations, including in live cells, we believe the difference is due to their fixation-permeabilization procedure. Recently, another study demonstrated that overexpression of T-STAR/ETOILE (human SLM-2) in HeLa cells, but not in COS cells, resulted in nuclear bodies similar to SNBs (Venables et al., 1999). These nuclear structures did not colocalize with coiled bodies, the PNC, and Y12-stained structures (Venables et al., 1999), consistent with our observations. We have observed that not all nuclear GSG proteins localize in SNBs in HeLa cells (data not shown); thus, SNBs may not be a general site for all nuclear GSG proteins.

In conclusion, we have identified a novel nuclear structure that contains Sam68, SLM-1, and SLM-2. We have termed these structures Sam68/SLM nuclear bodies. SNBs are dynamic structures primarily present in cancer cells that are distinct from coiled bodies, gems, and PML nuclear bodies. Alterations in the GSG domain of Sam68 gave six separate classes of Sam68 cellular patterns, including exclusive SNB localization and association with the microtubules. Our data imply that the GSG domain is a multifunctional protein module involved in protein localization and define a new cellular compartment in transformed cells for Sam68/SLM proteins.

ACKNOWLEDGMENTS

We thank Hugo Dilhuydy at the Centre de Recherche Institut Universitaire de Gériatrie de Montréal for excellent help and assistance with the confocal microscopy. We thank Manfred Herfort for excellent technical assistance and Mark Bedford for the SmB antibodies, for critically reading the manuscript, and for helpful discussions. We thank Edward Chan for the anti-coilin antibodies, Michael Pollard for the anti-fibrillarin antibodies, and Yuri Svitkin and Nahum Sonenberg for the PTB cDNA. We are grateful to members of the Wilson Miller lab for helpful discussions. This work was supported by research grants from the Medical Research Council of Canada and the Cancer Research Society to S. R. and D. P. B.-J. T. C. is a recipient of the Doctoral Research Award from the Medical Research Council of Canada, and S.R. is a Scholar of the Medical Research Council of Canada.

REFERENCES

Andrade, L. E. C., Tan, E. M. and Chan, E. K. L. 1993. Immunocytochemical analysis of the coiled body in the cell cycle and during cell proliferation. Proc. Natl. Acad. Sci. USA, 90:1947-1951.

Arning, S., Gruter, P., Bilbe, G. and Kramer, A. 1996. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. RNA, 2:794-810.

Baehrecke, E. H. 1997. who encodes a KH RNA binding protein that functions in muscle development. Development, 124:1323-1332.

Barlat, I., Maurier, F., Duchesne, M., Guitard, E., Tocque, B. and Schweighoffer, F. 1997. A role for Sam68 in cell cycle progression antagonized by a spliced variant within the KH domain. J. Biol. Chem., 272:3129-3132.

Bazett-Jones, D. P. and Hendzel, M. J. 1999. Electron spectroscopic imaging of chromatin. Methods, 17:188-200.

Bechade, C., Rostaing, P., Cisterni, C., Kalusch, R., La Bella, V., Pettmann, B. and Triller, A. 1999. Subcellular distribution of survival motor neuron (SMN) protein: possible involvement in nucleocytoplasmic and dendritic transport. Eur. J. Neurosci., 11:293-304.

Berglund, J. A., Chua, K., Abovich, N., Reed, R. and Rosbash, M. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branch-point sequence UACUAAC. Cell, 89:781-787.

Bussaglia, E., Clermont, O., Tizzano, E., Lefebvre, S., Burglen, L., Cruaud, C. and Urtizberea, J. A. 1995. A frame-shift deletion in the survival motor neuron gene in Spanish spinal muscular atrophy patients. Nature Genet., 11:335-337.

Carmo-Fonseca, M., Ferreira, J. and Lamond, A. I. 1993. Assembly of snRNP-containing coiled bodies is regulated in interphase and mitosis-evidence that the coiled body is a kinetic nuclear structure. J. Cell Biol., 120:841-852.

Chen, T., Damaj, B., Herrerra, C., Lasko, P. and Richard, S. 1997. Self-association of the single-KH domain family members Sam68, GRP33, GLD-1 and Qk1: Role of the KH

domain. Mol. Cell. Biol., 17:5707-5718.

Chen, T. and Richard, S. 1998. Structure-function analysis of Qk1: A lethal point mutation in mouse quaking prevents homodimerization. Mol. Cell. Biol., 18:4863-4871.

Cruz-Alvarez, M. and Pellicer, A. 1987. Cloning of a full-length complementary cDNA for as *Artemia salina* glycine-rich protein. J. Biol. Chem., 262:13377-13380.

DeBoulle, K., Verkerk, A. J. M. H., Reyniers, E., Vits, L., Hendrickx, J., Roy, B. V., Bos, F. V. D., DeGraaff, E., Oostra, B. A. and Willems, P. J. 1993. A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nature Genetics, 3:31-35.

Dhordain, P., Albagli, O., Lin, R. J., Ansieau, S., Quief, S., Leutz, A., Kerckaert, J. P., Evans, R. M. and Leprince, D. 1997. Corepressor SMRT binds the BTB/POZ repressing domain of the LAZ3/BCL6 oncoprotein. Proc. Natl. Acad. Sci. USA, 94:10762-10767.

Di Fruscio, M., Chen, T., Bonyadi, S., Lasko, P. and Richard, S. 1998. The identification of two Drosophila KH domain proteins: KEP1 and SAM are members of the Sam68 family of GSG domain proteins. J. Biol. Chem., 273:30122-30130.

Di Fruscio, M., Chen, T. and Richard, S. 1999. Two novel Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis. Proc. Natl. Acad. Sci. USA, 96:2710-2715.

Doucas, V., Ishov, A. M., Romo, A., Juguilon, H., Weitzman, M. D., Evans, R. M. and Maul, G. G. 1996. Adenovirus replication is coupled with the dynamic properties of the PML nuclear structure. Genes Dev., 10:196-207.

Dyck, J. A., Maul, G. G., Miller, W. H., Chen, J. D., Kakizuka, A. and Evans, R. M. 1994. A novel macromolecular structure is a target of the promyelocyte-retinoic acid receptor oncoprotein. Cell, 76:333-343.

Ebersole, T. A., Chen, Q., Justice, M. J. and Artzt, K. 1996. The *quaking* gene unites signal transduction and RNA binding in the developing nervous system. Nature Genetics, 12:260-265.

Feng, Y., Absher, D., Eberhart, D. E., Brown, V., Malter, H. E. and Warren, S. T. 1997. FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe

fragile X syndrome abolishes this association. Mol. Cell, 1:109-118.

Fischer, U., Liu, Q. and Dreyfuss, G. 1997. The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. Cell, 90:1023-1029.

Francis, R., Barton, M. K., Kimbel, J. and Schedl, T. 1995a. Control of oogenesis, germline proliferation and sex determination by the *C. elegans* gene *gld-1*. Genetics, 139:579-606.

Francis, R., Maine, E. and Schedl, T. 1995b. *gld-1*, a cell-type specific tumor suppressor gene in *Caenorhabditis elegans*. Genetics, 139:607-630.

Fu, X. D. and Maniatis, T. 1990. Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. Nature, 343:437-441.

Fumagalli, S., Totty, N. F., Hsuan, J. J. and Courtneidge, S. A. 1994. A target for Src in mitosis. Nature, 368:871-874.

Fyrberg, C., Becker, J., Barthmaier, P., Mahaffey, J. and Fyrberg, E. 1997. A Drosophia muscle-specific gene related to the mouse quaking locus. Gene, 197:315-323.

Fyrberg, C., Becker, J., Barthmaier, P., Mahaffey, J. and Fyrberg, E. 1998. A family of Drosophila genes encoding quaking-related Maxi-KH domains. Biochem. Genetics, 36:51-64.

Gall, J. G., Tsvetkov, A., Wu, Z. A. and Murphy, C. 1995. Is the sphere organelle/coiled body a universal nuclear component? Dev. Genet., 16:25-35.

Gibson, T. J., Thompson, J. D. and Heringa, J. 1993. The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. FEBS Letters, 324:361-366.

Havin, L., Git, A., Elisha, Z., Oberman, F., Yaniv, K., Schwartz, S. P., Standart, N. and Yisraeli, J. K. 1998. RNA-binding protein conserved in both microtubule- and microfilament-based RNA localization. Genes & Dev., 12:1593-1598.

Hendzel, M. J., Boisvert, F.-M. and Bazett-Jones, D. P. 1999. Direct visualization of a protein nuclear architecture. Mol. Biol. Cell, 10:2051-2062.

Hogan, E. L. and Greenfield, S. 1984. Animal models of genetic disorders of myelin. Myelin. P. Morells. Plenum, New York. 489-534.

Huang, S., Deerinck, T. J., Ellisman, M. H. and Spector, D. L. 1997. The dynamic organization of the perinucleolar compartment in the cell nucleus. J. Cell Biol., 137:965-974.

Huang, S., Deerinck, T. J., Ellisman, M. H. and Spector, D. L. 1998. The perinucleolar compartment and transcription. J. Cell Biol., 143:35-47.

Ishidate, T., Yoshihara, S., Kawasaki, Y., Roy, B. C., Toyoshima, K. and Akiyama, T. 1997. Identification of a novel nuclear localization signal in Sam68. FEBS Lett., 409:237-241.

Jackson, D., Hassan AB, Errington RJ, and Cook PR. 1993. Visualization of focal sites of transcription within human nuclei. EMBO J, 12:1059-1065.

Jones, A. R. and Schedl, T. 1995. Mutations in GLD-1, a female germ cell-specific tumor suppressor gene in *C.elegans*, affect a conserved domain also found in Sam68. Genes Dev., 9:1491-1504.

Justice, M. J. and Bode, V. C. 1988. Three ENU-induced alleles of the murine quaking locus are recessive embryonic lethal mutations. Genet. Res. Camb., 51:95-102.

Kalderon, D., Roberts, B. L., Richardson, W. D. and Smith, A. E. 1984. A short amino acid sequence able to specify nuclear location. Cell, 39:499-509.

Kerckaert, J. P., Deweindt, C., Tilly, H., Quief, S., Lecocq, G. and Bastard, C. 1993. LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3q27 translocations in human lymphomas. Nature Genet., 5:66-70.

Koken, M. H. M., Puvion-Dutilleul, F., Guillemin, M. C., Viron, A., Linares-Cruz, G., Stuurman, N., de Jong, L., Szostecki, C., Calvo, F., Chomienne, C., Degos, L., Puvion, E. and de Thé, H. 1994. The t(15;17) translocation alters a nuclear body in a retinoic acid-reversible fashion. EMBO J., 13:1073-1083.

Lamond, A. I. and Carmo-Fonseca, M. 1993. The coiled body. Trends Cell Biol., 3:198-204.

Lamond, A. I. and Earnshaw, W. C. 1998. Structure and function in the nucleus. Science, 280:547-553.

Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M. and al., e. 1995. Identification and characterization of a spinal muscular atrophy-determining gene. Cell, 89:155-165.

Lin, Q., Taylor, S. J. and Shalloway, D. 1997. Specificity and determinants of Sam68 RNA binding. J. Biol. Chem., 272:27274-27280.

Liu, Q. and Dreyfuss, G. 1996. A novel nuclear structure containing the survival of motor neurons protein. EMBO J., 15:3555-3565.

Liu, Q., Fischer, U., Wang, F. and Dreyfuss, G. 1997. The spinal muscular atrophy gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. Cell, 90:1013-1021.

Matera, A. G. and Frey, M. R. 1998. Coiled bodies and gems: janus or gemini? Am. J. Hum. Genet., 63:317-321.

Mattaj, I. W. and DeRobertis, E. M. 1985. Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. Cell, 40:111-118.

McBride, A. E., Taylor, S. J., Shalloway, D. and Kirkegaad, K. 1998. KH domain integrity is required for wild-type localization of Sam68. Exp. Cell Res., 241:84-95.

Michael, W. M., Eder, P. S. and Dreyfuss, G. 1997. The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein. EMBO J., 16:3587-3598.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A. and Harshman, K., et al. 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science, 266:66-71.

Monneron, A. and Bernhard, W. 1969. Fine structural organization of the interphase nucleus in some mammalian cells. J. Ultrastruct. Res., 27:266-288.

Musco, G., Stier, G., Joseph, c., Morelli, M. A. and Pastore, A. 1996. Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. Cell, 85:237-245.

Patton, J., Mayer, SA, Tempst, P, Nadal-Ginard, B. 1991. Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. Genes Dev, 5:1237-1251.

Pellizzoni, L., Kataoka, N., Charroux, B. and Dreyfuss, G. 1998. A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. Cell, 95:615-624.

Petterson, I., Hinterberger, M., Mimori, T., Gottlieb, E. and Steitz, J. A. 1984. The structure of mammalian small nuclear ribonucleoproteins: identification of multiple protein components reactive with anti-(U1) ribonucleoprotein and anti-Sm autoantibodies. J. Biol. Chem., 259:5907-5914.

Pinol-Roma, S. and Dreyfuss, G. 1992. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. Nature, 355:730-732.

Pombo, A., Cuello, P., Schul, W., Yoon, J. B., Roeder, R. G., Cook, P. R. and Murphy, S. 1998. Regional and temporal specialization in the nucleus: a transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. EMBO J, 17:1768-1778.

Rain, J. C., Rafi, Z., Rhani, Z., Legrain, P. and Kramer, A. 1998. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. RNA, 4:551-565.

Reddy, T. R., Xu, W., Mau, J. K. L., Goodwin, C. D., Suhasini, M., Tang, H., Frimpong, K., Rose, D. W. and Wong-Staal, F. 1999. Inhibition of HIV replication by dominant negative mutants of Sam68, a fuctional homolog of HIV-1 Rev. Nat. Med., 5:635-642.

Richard, S., Yu, D., Blumer, K. J., Hausladen, D., Olszowy, M. W., Connelly, P. A. and Shaw, A. S. 1995. Association of p62, a multi-functional SH2- and SH3-binding protein, with src-family tyrosine kinases, Grb2, and phospholipase Cγ-1. Mol. Cell. Biol., 15:186-197.

Siomi, H., Choi, M., Siomi, M. C., Nussbaum, R. L. and Dreyfuss, G. 1994. Essential role for KH domain in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. Cell, 77:33-39.

Sisodia, S. S. 1998. Nuclear inclusions in glutamine repeat disorders: are they pernicious, coincidental, or beneficial? Cell, 95:1-4.

Spector, D. L. 1993. Nuclear organization of pre-mRNA processing. Curr. Opin. in Cell Biol., 5:442-447.

Spector, D. L., Lark, G. and Huang, S. 1992. Differences in snRNP localization between transformed and nontransformed cells. Mol. Biol. Cell., 3:555-569.

Tanaka, H., Abe, K. and Kim, C. H. 1997. Cloning and expression of the quaking gene in the Zebrafish embryo. Mech. Develop., 69:209-213.

Taylor, S. J., Anafi, M., Pawson, T. and Shalloway, D. 1995. Functional interaction between c-src and its mitotic target, Sam68. J. Biol. Chem., 270:10120-10124.

Taylor, S. J. and Shalloway, D. 1994. An RNA-binding protein associated with src through its SH2 and SH3 domains in mitosis. Nature, 368:867-871.

Venables, J. P., Vernet, C., Chew, S. L., Elliot, D. J., Cowmeadow, R B., Wu, J., Cooke, H. J., Artzt, K. and Eperon, I. C. 1999. T-STAR/ETOILLE: a novel relative of Sam68 that interacts with an RNA-binding protein implicated in spermatogenesis. Hum. Mol. Genet., 8:959-969.

Vernet, C. and Artzt, K. 1997. STAR, a gene family involved in signal transduction and activation of RNA. Trends in Genet., 13:479-484.

Wang, L. L., Richard, S. and Shaw, A. S. 1995. p62 association with RNA is regulated by tyrosine phosphorylation. J. Biol. Chem, 270:2010-2013.

Weis, K., Rambaud, S., Lavau, C., Jansen, J., Carvalho, T., Carmo-Fonseca, M., Lamond, A. and Dejean, A. 1994. Retinoic acid regulates aberrant nuclear localization of PML-RARa in acute promyelocytic leukemia cells. Cell, 76:345-356.

Willems, M., Penman, M. and Penmen, S. 1969. The regulation of RNA synthesis and processing in the nucleolus during inhibition of protein synthesis. J. Cell Biol., 41:177-187.

Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M. F., Polakis, P. and McCormick, F. 1992. Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. Cell, 69:551-558.

Wooster, R., Bignell, G., Lancaster, J., Swift, S. and Seal, S. e. a. 1995. Identification of the breast cancer susceptibility gene BRCA2. Nature, 378:789-792.

Ye, B. H., Lista, F., Lo Coco, F., Knowles, D. M., Offit, K., Chaganti, R. S. K. and Dalla-Favera, R. 1993. Alterations of a zinc finger-encoding gene, BCL-6, in diffuse large-cell lymphoma. Science, 262:747-750.

Zaffran, S., Astier, M., Gratecos, D. and Semeriva, M. 1997. The held out wings (how) Drosophila gene encodes a putative RNA binding protein involved in the control of muscular and cardiac activity. Development, 124:2087-2098.

Zorn, A. M., Grow, M., Patterson, K. D., Ebersole, T. A., Chen, Q., Artzt, K. and Krieg, P. A. 1997. Remarkable sequence conservation of transcripts encoding amphibian and mammalian homologues of *quaking* a KH domain RNA-binding protein. Gene, 188:199-206.

Zorn, A. M. and Krieg, P. A. 1997. The KH domain protein encoded by quaking functions as a dimer and is essential for notochord development in Xenopus embryos. Genes Dev., 11:2176-2190.

Chapter 5
GENERAL DISCUSSION

5.1 THE GSG DOMAIN IS A MULTIFUNCTIONAL PROTEIN MODULE

The GSG domain, which contains a KH RNA-binding module, has been thought to be involved in RNA binding since its discovery in 1995 (Jones and Schedl, 1995). The evidence presented in this thesis demonstrates that the GSG domain does mediate RNA binding. As shown in chapters 2 and 3, the GSG domain of Sam68 and Qk1 is required for RNA binding. While the minimum region of Sam68 that is capable of binding RNA is the GSG domain plus ~50 amino acids immediately C-terminal to it (Chen et al., 1997), the Qk1 GSG domain alone is sufficient for RNA binding (Chen and Richard, 1998). In addition, we demonstrate that the GSG domain has other functions, including mediating protein-protein interactions such as self-association (Chen et al., 1997; Chen and Richard, 1998; Di Fruscio et al., 1998; Di Fruscio et al., 1999) and directing protein localization (Chen et al., 1999). All these functions are dependent on an intact GSG domain, and the embedded KH domain is required but not sufficient, suggesting that the entire GSG domain functions as an inseparable unit. RNA-binding proteins usually contain one or more RNA-binding domains and other separate functional regions, such as protein-protein interaction motifs and localization signals (Siomi and Dreyfuss, 1997). With these properties combined in a single protein module, the GSG family thus represents a special group of RNA-binding proteins.

Oligomerization is a common property of GSG domain-containing proteins as most family members examined so far are able to self-associate (Chen et al., 1997; Zorn and Krieg, 1997; Chen and Richard, 1998; Di Fruscio et al., 1998; Di Fruscio et al., 1999). This suggests that oligomerization must be functionally important for this protein family. Consistent with this view is the finding that failure to dimerize is most likely the underlying mechanism for the lethal phenotype caused by the mouse Qk1 E48G point mutation (Chen and Richard, 1998). While the precise role of oligomerization for the functions of GSG

proteins is unknown, it is possible that oligomerization enhances RNA binding. This possibility is supported by several observations: 1) Sam68 multimers can associate with RNA *in vitro* (Chen et al., 1997), 2) mutations that affect Sam68 self-association usually also impair its RNA-binding capability (Chen et al., 1997), and 3) phosphorylation of Sam68 by Src-family kinases such as $p59^{fyn}$ inhibits both self-association (Chen et al., 1997) and RNA binding (Wang et al., 1995). KH domain-containing proteins usually have multiple copies of KH domains (Gibson et al., 1993; Musco et al., 1996), which function cooperatively in achieving optimal RNA binding (Siomi et al., 1994). Thus, oligomerization of GSG proteins, which contain a single KH domain, may fulfill the multi-KH requirement for RNA binding. Alternatively, some GSG proteins may need to form multimers in order to bind their RNA substrates at multiple sites. The *C. elegans* GSG protein GLD-1 has recently been shown to repress translation of *tra-2*, a sex determination gene, by binding to a direct repeat located in the *tra-2* 3'UTR (Jan et al., 1999). This direct repeat consists of two regulatory elements called *tra-2* and *GLI* elements (TGEs) (Goodwin et al., 1993). Interestingly, in situ chemical cross-linking experiments in HeLa cells demonstrated that GLD-1 forms mainly homodimers (Chen et al., 1997). This raises the possibility that each subunit of the GLD-1 dimer binds to one TGE, similar to the association of steroid receptors with their specific response elements located in the promoter region of target genes. Despite all the above evidence, it cannot be ruled out that some GSG proteins may bind RNA as monomers. The Qk1 E48G point mutation abolishes Qk1 dimerization but has little effect on its ability to associate with total cellular RNA (Chen and Richard, 1998). However, it was recently shown that Qk1-6, like GLD-1, binds the *tra-2* TGEs in vitro, and the binding is reduced with the introduction of the E48G point mutation (Saccomanno et al., 1999). These observations suggest that although Qk1 monomers have the capability to bind RNA, dimerization is required for stable association with specific RNA targets.

Although the GSG domain is required and sufficient for self-association of GSG proteins (Chen et al., 1997; Chen and Richard, 1998; Di Fruscio et al., 1998; Di Fruscio et al., 1999), the regions within the GSG domain that mediate self-association differ among

GSG family members. For Sam68, the entire GSG domain is required for self-association and within this region the KH domain loops 1 and 4 are essential (Chen et al., 1997), consistent with the prediction that these extended loops are involved in protein-protein interactions (Musco et al., 1996). Qk1 self-association is mediated by the NK region, which is predicted to form coiled-coil interactions (Chen and Richard, 1998). The coiled-coil motif is conserved in Qk1-related family members but not in other GSG proteins (Chen and Richard, 1998), suggesting that coiled-coil interaction may be a common mechanism of self-association for Qk1-related GSG proteins. The Qk1 KH domain is not required for dimerization (Chen and Richard, 1998), but the extended loops 1 and 4 may be involved in interaction with other proteins. The different mechanisms by which GSG proteins self-associate may explain certain differences observed among GSG family members. For instance, Sam68 and GRP33 require cellular RNA for self-association (Chen et al., 1997; Chen and Richard, 1998) whereas Qk1 self-association is RNA independent (Chen and Richard, 1998).

The sequence of the GSG domain suggests that this protein module may be derived from the KH domain during evolution. While KH domains are present in proteins from both prokaryotes and eukaryotes (Gibson et al., 1993; Musco et al., 1996), the GSG domain has been found only in eukaryotic proteins (Vernet and Artzt, 1997). Therefore, it is likely that GSG domain-containing proteins have evolved to function in certain processes specific to eukaryotic cells. Recent evidence has demonstrated the involvement of GSG proteins in various processes of eukaryotic RNA metabolism. SF1/BBP binds the branchpoint sequence of introns (Berglund et al., 1997) and facilitates spliceosome assembly during pre-mRNA splicing (Arning et al., 1996; Abovich and Rosbash, 1997). GLD-1 functions as a specific translational repressor (Jan et al., 1999). How is involved in tendon cell differentiation by regulating *stripe* mRNA nuclear export (Nabel-Rosen et al., 1999).

Binding to specific RNA targets is essential for the functions of GSG proteins. What then determines the RNA-binding specificity of these proteins? Our deletion analyses demonstrated that the GSG domain is the major determinant of RNA binding in GSG

proteins (Chen et al., 1997; Chen and Richard, 1998; Di Fruscio et al., 1998; Di Fruscio et al., 1999), consistent with a number of recent studies (Lin et al., 1997; Berglund et al., 1998; Rain et al., 1998). This indicates that the GSG domain binds RNA with sequence and/or structure specificity. Although the embedded KH domain is absolutely required for RNA binding (Chen et al., 1997; Chen and Richard, 1998), it alone may not confer specificity given the fact that it contains highly conserved amino acid residues in the predicted RNA-binding surface (Musco et al., 1996). Since GSG proteins form homo- and hetero-multimers (Chen et al., 1997; Di Fruscio et al., 1999), which bring multiple KH domains in close proximity, the spatial arrangement and organization of these KH domains may favor the binding of RNA with specific structures. Recently, the three-dimensional structure of a KH domain-RNA complex has shown that the third KH domain of Nova-2 binds to a high-affinity ligand as dimers (Lewis et al., 2000). Thus, in the GSG domain, the KH domain forms the contact site for RNA and the newly evolved sequence elements, including the extended loops and the NK and CK regions, play a role in holding two or more KH domains together and presenting them in a specific, RNA binding-competent conformation. Alternatively, the NK and CK regions may directly contact RNA and contribute to RNA-binding specificity. Sequence elements outside of the GSG domain may also play accessory roles in enhancing RNA-binding affinity and affecting specificity. It has been shown that the GSG domain of SF1/BBP mediates specific recognition of the pre-mRNA branchpoint sequence and the zinc knuckles increase RNA-binding affinity (Berglund et al., 1998). We demonstrated that the minimum region of Sam68 that is sufficient to bind poly(U) is the GSG domain plus an ~50-amino-acid region that contains multiple arginine/glycine (RG) repeats (Chen et al., 1997). It remains to be determined whether this RG-rich region plays a role in maintaining the structure of the GSG domain or it represents a functional RNA-binding motif that works cooperatively with the GSG domain.

In addition to mediating self-association and RNA binding, the Sam68 GSG domain plays a role in protein localization (Chen et al., 1999). This property is probably not unique to the Sam68 GSG domains, because a recent report showed that a deletion in the GSG

domain of Qk1-5 also alters protein localization patterns (Wu et al., 1999). Thus, some phenotypes of genetic mutations in the GSG domain may arise from mislocalization of a GSG protein. A point mutation identified in the GSG domain of *Drosophila* How, altering arginine 185 to cysteine (Baehrecke, 1997), results in embryonic lethality, but it does not alter self-association or RNA binding (Chen and Richard, 1998). Introduction of the analogous mutation (R204C) in Sam68 results in dramatic alteration of protein localization (Chen et al., 1999), suggesting that mislocalization of How may cause the lethal phenotype. Indeed, the subcellular localization of How is developmentally regulated and plays a critical role in tendon cell differentiation (Nabel-Rosen et al., 1999). Several GSG proteins, including Qk1, Sam68, and How, have been implicated in RNA transport (Chen and Richard, 1998; Nabel-Rosen et al., 1999; Reddy et al., 1999). The GSG domains of these proteins may play an important role in the movement of the ribonucleoprotein complexes. While the localization property of the GSG domain appears to be separate from RNA binding and self-association (Chen et al., 1999; Wu et al., 1999), it may involve GSG domain-mediated interactions with specific proteins.

5.2 A MODEL FOR THE FUNCTION OF THE GSG DOMAIN

Several GSG proteins have been shown to regulate gene expression post-transcriptionally by interacting with specific RNA targets (Jan et al., 1999; Nabel-Rosen et al., 1999; Reddy et al., 1999). Based on the evidence presented in this thesis and a number of recent studies, a model is proposed to explain how the GSG domain mediates protein-RNA interactions (Fig. 5-1). While GSG proteins are involved in different stages of RNA metabolism, they interact with RNA in a similar fashion and the general process can be divided into several sequential steps. The process starts with GSG domain-mediated oligomerization of GSG proteins. In some cases, monomeric GSG proteins loosely associate with specific or nonspecific RNAs, and these RNA molecules may bridge the protein monomers and facilitate their initial interactions during oligomerization. Then, GSG protein multimers tightly associate with their RNA substrates. The KH domain is the major

RNA contact site and its elongated loops 1 and 4 as well as the newly evolved NK and CK regions mediate protein-protein interactions, including self-association and association with other proteins. Oligomerization of GSG proteins forms multivalent RNA-binding sites and thus stabilizes protein-RNA interactions. In addition, homo- and hetero-oligomerization of GSG proteins may be a major determinant for RNA-binding specificity. GSG domain-mediated interactions with other proteins facilitate the assembly of RNP complexes and target the complexes to proper cellular compartments. Binding of specific RNA sequences or structures by GSG proteins may induce changes in RNA conformation, localization, stability, or accessibility to other factors, resulting in modulation of RNA functions. These include mRNA stability, splicing, polyadenylation, RNA transport, and translation. Lastly, GSG proteins dissociate from their RNA substrates and are recycled. Since GSG proteins usually contain signaling motifs, such as SH2, SH3, and WW domain binding motifs and phosphorylation sites, dissociation of protein-RNA complexes may be regulated by signal transduction pathways. Indeed, the RNA-binding ability of Sam68 is inhibited by tyrosine phosphorylation (Wang et al., 1995; Derry, submitted) or binding to SH3 domains (Taylor et al., 1995). Some of the regulatory mechanisms may act at the level of protein oligomerization. For instance, phosphorylation of Sam68 prevents self-association (Chen et al., 1997), which may account for the inhibition of RNA binding (Wang et al., 1995; Derry, submitted).

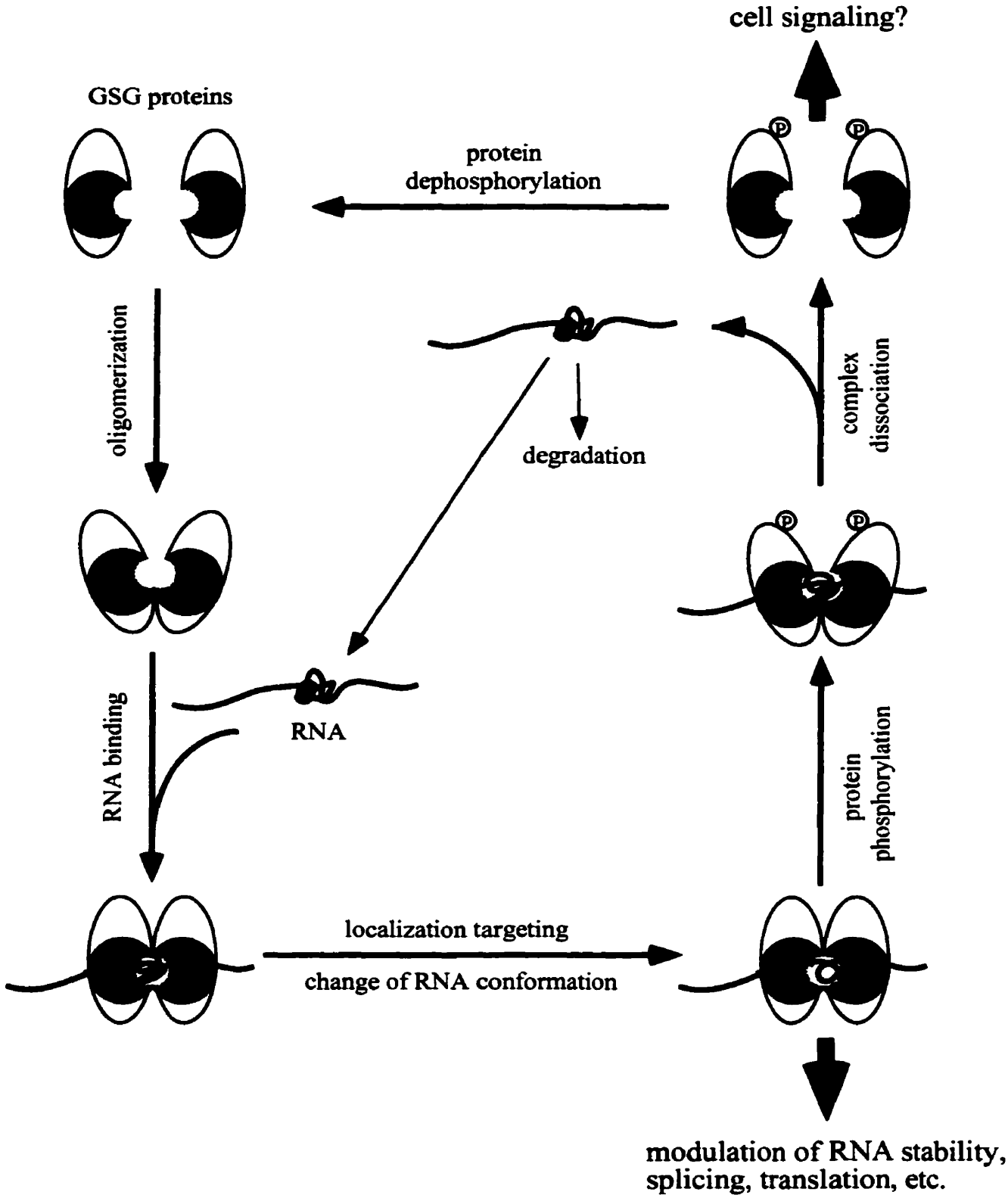


FIG. 5-1. A model for the function of the GSG domain. The shaded area in GSG proteins represents the GSG domain. Ⓟ = phosphorylation

5.3 SPECULATIONS ON THE FUNCTION OF Qk1

Qk1 have been implicated in myelination and early embryogenesis, but its precise function is still unknown. Although Qk1 contains potential signaling motifs, including a cluster of tyrosine residues close to the C terminus and a few proline-rich regions (Ebersole et al., 1996b), no evidence has been provided that Qk1 is involved in signal transduction. We demonstrated that Qk1 binds RNA via its GSG domain (Chen and Richard, 1998), indicating that Qk1 may play a role in some aspects of RNA metabolism. Indeed, recent studies have shown that Qk1 homologs in various species interact with specific RNA targets and post-transcriptionally regulate gene expression. GLD-1, the *C. elegans* Qk1 homolog, represses the translation of *tra-2*, a sex determination gene (Jan et al., 1999), and Qk1-6 has similar functions when expressed in *C. elegans* (Saccomanno et al., 1999). How, the *Drosophila* Qk1 homolog, functions in the nuclear export of *stripe* mRNA and controls tendon cell differentiation (Nabel-Rosen et al., 1999).

In *quaking* viable mice, Qk1-6 and Qk1-7 are absent in oligodendrocytes (myelin-forming cells) whereas the expression of Qk1-5 is not significantly altered (Hardy et al., 1996). The dysmyelination observed in these mutants may arise from certain defects in RNA metabolism. Several consequences could be envisioned from altered expression of Qk1 isoforms in oligodendrocytes. First, specific RNAs could be mislocalized. We demonstrated that Qk1 dimerizes both in vitro and in vivo (Chen and Richard, 1998), suggesting that Qk1 may bind its RNA targets as dimers. Given the fact that Qk1-5 localizes primarily in the nucleus and Qk1-6 and Qk1-7 localize mainly in the cytoplasm (Hardy et al., 1996), heterodimerization between Qk1 isoforms may play a critical role in the transport and localization of specific RNAs. In fact, heterodimerization between two How isoforms, How(L) and How(S), has been implicated in the nuclear export of *stripe* mRNA (Nabel-Rosen et al., 1999). Second, nuclear RNA processing such as mRNA splicing could be altered. Indeed, the level of the mRNAs and polypeptides for the small and large myelin-associated glycoprotein (MAG) is altered in *quaking* viable mice (Frail and Braun, 1985; Braun et al., 1990; Fujita et al., 1990; Bartoszewicz et al., 1995). It would be interesting to

determine whether Qk1 proteins are directly involved in the regulation of alternative splicing of MAG pre-mRNA. Third, cytoplasmic RNA processing and function could be affected. In oligodendrocytes, which send out many long processes that wrap around axons and form the myelin sheath, certain myelin proteins, such as myelin basic protein (MBP), are synthesized locally in the processes at sites of myelin assembly. To achieve this, mRNAs encoding these proteins need to be transported from the cell body to the processes (Colman et al., 1982). Alterations in the localization of MBP mRNA and proteins have been observed in cultured oligodendrocytes from *quaking* viable mice (Barbarese, 1991), suggesting that Qk1-6 and Qk1-7 may play a role in the trafficking of MBP mRNA or, alternatively, in maintaining its stability and/or suppressing its translation during the transport process.

It is not clear why altered expression of Qk1 isoforms in *quaking* viable mutants occurs in oligodendrocytes but not in other glial cells such as astrocytes (Hardy et al., 1996). The genomic structure of the *quaking* gene shows a single transcription initiation site (Cox et al., 1999; Kondo et al., 1999), suggesting that differential expression of Qk1 isoforms is regulated by alternative splicing. Since the genetic lesion of the *quaking* viable mutation is a large (>1Mb) deletion on chromosome 17 (Ebersole et al., 1996a), other genes may also contribute to the phenotype. One possibility is that a factor essential for Qk1 pre-mRNA splicing in oligodendrocytes is missing but the defect may be compensated in other glial cells. Interestingly, the 6 kb and 7 kb *qkI* transcripts share the same 3'UTR and show similar expression patterns during development (Ebersole et al., 1996a), suggesting that the 3'UTR may play a regulatory role in *qkI* expression.

Four ethylnitrosourea (ENU)-induced *quaking* alleles, qk^{l-1}, qk^{kt1}, qk^{k2}, and $qk^{kt3/kt4}$, are lethal during early embryogenesis (Bode, 1984; Justice and Bode, 1988; Shedlovsky et al., 1988). qk^{kt3} and qk^{kt4} alter glutamic acid 48 to a glycine (E48G) in the NK region of the GSG domain (Ebersole et al., 1996a; Cox et al., 1999). We demonstrated that Qk1:E48G is unable to dimerize (Chen and Richard, 1998), suggesting that dimerization is required for the normal function of Qk1 proteins. Moreover, we showed that expression of either wild-type Qk1-7 or Qk1-7:E48G in NIH 3T3 cells induces apoptosis

and Qk1-7:E48G is consistently more potent than wild-type Qk1-7 (Chen and Richard, 1998). These findings indicate that Qk1 may play a role in apoptosis, and unregulated apoptotic cell death due to the absence of GSG domain-mediated dimerization may be the underlying mechanism for the embryonic lethality observed in $qk^{kt3/kt4}$ mice. Although the exact role of Qk1 in apoptosis is unknown, it is likely that Qk1 post-transcriptionally regulates the expression of apoptotic or cell survival genes.

5.4 SPECULATIONS ON THE FUNCTION OF SAM68

Because Sam68 contains several potential signaling motifs (tyrosine-rich domain, proline-rich regions, etc.) (Wong et al., 1992; Richard et al., 1995) and it is tyrosine phosphorylated by Src when cells are arrested in mitosis by nocodazole (Fumagalli et al., 1994; Taylor and Shalloway, 1994), previous studies were focused on the role of Sam68 in signal transduction. In certain cellular systems, Sam68 has been shown to be a substrate for Src family kinases (Fumagalli et al., 1994; Taylor and Shalloway, 1994; Weng et al., 1994; Richard et al., 1995), ZAP-70 (Lang et al., 1997), the insulin receptor (Sanchez-Margalet and Najib, 1999), the nuclear Src-like kinase Sik/BRK (Derry et al., submitted), and the serine/threonine kinase Cdc2 (Resnick et al., 1997). Moreover, Sam68 is able to associate with a number of signaling molecules via SH2 and/or SH3 domain-mediated interactions (Richard et al., 1995). These findings are consistent with the hypothesis that Sam68 functions as an adaptor molecule for tyrosine kinases (Richard et al., 1995; Taylor et al., 1995). However, a physiological role for Sam68 in signal transduction has yet to be established.

One puzzling question is how Sam68, a nuclear protein, gets access to its interacting partners, which are mostly cytoplasmic. It has been hypothesized that Sam68 interacts with Src family kinases during mitosis when the nuclear envelope breaks down (Fumagalli et al., 1994; Taylor and Shalloway, 1994). Indeed, we observed that cytoplasmic Sam68 can only be detected in mitotic cells by indirect immunofluorescence studies (Chen et al., 1999). Since the nuclear envelope remains intact until late prophase, Sam68 phosphorylation

presumably occurs after mitotic entry. This suggests that Sam68 phosphorylation may play a role in mitotic progression. Our finding that some mutant Sam68 associates with microtubules (Chen et al.. 1999) raises the possibility that Sam68 may be involved in mitotic spindle formation and function. Besides Src, several signaling proteins including components of the MAP kinase pathway are activated during mitosis (Bagrodia et al., 1991; Kaech et al., 1991; Laird et al., 1995; Roche et al., 1995; Shapiro et al., 1998; Zecevic et al., 1998), and active MAK kinases have been shown to localize at kinetochores (Shapiro et al., 1998; Zecevic et al., 1998). It would be interesting to determine whether Sam68 acts as an adaptor protein linking Src to the MAP kinase pathway during mitosis.

Regardless of its role, if any, in signal transduction, Sam68 is an RNA-binding protein and may be involved in posttranscriptional gene expression. Sam68 is localized predominantly, if not exclusively, in the nucleus during interphase (Wong et al., 1992; Ishidate et al., 1997; McBride et al., 1998; Chen et al., 1999). Although it has been shown to relocalize to the cytoplasm in cells infected with poliovirus (McBride et al., 1996), no direct evidence has been provided that Sam68 normally shuttles between the nucleus and the cytoplasm. Moreover, we have shown that Sam68 is concentrated in distinct nuclear structures called Sam68/SLM nuclear bodies (SNBs) (Chen et al., 1999). These observations suggest that Sam68 may function in certain nuclear steps of RNA processing.

SNBs are dynamic nuclear structures that contain nucleic acids (most likely representing RNA) (Chen et al., 1999), suggesting that these structures may be directly involved in RNA processing. We initially identified Sam68, SLM-1, and SLM-2 in SNBs (Chen et al., 1999). Ever since, two additional protein components, Sik/BRK and YT521-B, have been identified in these structures (Hartmann et al., 1999; Derry et al, submitted). Sik/BRK is a nuclear Src-like kinase that has been shown to phosphorylate Sam68 and negatively regulate its RNA-binding ability (Derry et al., submitted). YT521-B is a nuclear phosphoprotein implicated in alternative splicing. YT521-B has been shown to associate with Sam68 and the association is negatively regulated by tyrosine phosphorylation of Sam68 (Hartmann et al., 1999). These findings indicate that Sam68 and SNBs may play a

role in pre-mRNA splicing and their function may be regulated by tyrosine phosphorylation. We showed that there exists a Sam68 equilibrium between SNBs and the nucleoplasm (Chen et al., 1999). One scenario is that Sam68 acts as a "molecular carrier" and takes splicing factors and/or RNAs into SNBs for processing. Phosphorylation of Sam68 by Sik/BRK may trigger the dissociation of Sam68-cargo complexes and facilitate the delivery of cargo in SNBs. SNBs are present primarily in cancer cells, implying a possible link between these structures and cell transformation.

Additional evidence for the involvement of Sam68 in pre-mRNA splicing comes from several recent studies. Using the rat β-tropomyosin gene as a model system, Grossman and coworkers identified Sam68 as a component of a complex that assembles on the same intron bound by PTB (Grossman et al., 1998). We also observed that Sam68 is able to interact with PTB in vitro, as demonstrated by GST pull-down assay (Chen and Richard, unpublished data). PTB has been implicated in the regulation of a number of alternatively spliced pre-mRNAs (Valcarcel and Gebauer, 1997). It binds pyrimidine-rich sequences and represses nearby splice sites, possibly by competing with U2AF in getting access to the polypyrimidine tract (Chabot, 1996; Valcarcel and Gebauer, 1997; Lopez, 1998). Since Sam68 binds preferentially to poly(U) in vitro (Taylor and Shalloway, 1994; Wang et al., 1995; Chen et al., 1997), it is possible that Sam68 may also interfere with the binding of U2AF to some 3' splice sites. Alternatively, Sam68 may play a regulatory role in the assembly or activity of the repressor complex. In agreement with the possibility that Sam68 acts as a negative regulator of splicing, a recent study showed that Sam68 is a functional homolog of the HIV-1 Rev protein. It was shown that overexpressed Sam68 can substitute for Rev in mediating HIV-1 gene expression and viral replication (Reddy et al., 1999). Rev binds to the Rev response element (RRE) in the HIV-1 pre-mRNA to block splicing and facilitates the export of unspliced pre-mRNA (Kjems et al.. 1991; Fischer et al., 1995). Although Sam68 may play a role in RNA export, a more likely possibility is that it inhibits HIV-1 pre-mRNA splicing.

If Sam68 is indeed involved in alternative splicing, it would be interesting to determine

whether Sam68 modulates the alternative splicing of its own pre-mRNA. A rare alternatively spliced Sam68 isoform (Sam68ΔKH) has been identified that contains a 39-amino-acid deletion in the KH domain (Barlat et al., 1997). This isoform has reduced RNA-binding capability (Barlat et al., 1997) and altered cellular localization (Chen et al., 1999). Overexpression of Sam68ΔKH isoform in NIH 3T3 cells inhibits G1/S transition, and the effect can be rescued by coexpression of Sam68 (Barlat et al., 1997). One scenario is that Sam68 represses the 3' splice site of the intron immediately upstream of the alternative exon, resulting in exon skipping and the production of Sam68ΔKH isoform. The balance between the two isoforms of Sam68 may control cell proliferation by posttranscriptionally regulating the expression of genes involved in cell cycle progression.

Another nuclear step of RNA processing in which Sam68 may be involved is polyadenylation. Sam68 has been shown to bind RNA sequences containing a UAAA motif with high affinity (Lin et al., 1997). This raises the possibility that Sam68 may bind the polyadenylation signal (AAUAAA) of pre-mRNAs, like the cleavage and polyadenylation specificity factor (CPSF). Moreover, some polyadenylation factors localize in nuclear structures called cleavage bodies (Schul et al., 1996), which are similar to SNBs in size, number, and distribution pattern. It would be interesting to determine the relationship between cleavage bodies and SNBs. While Sam68 may not be an essential polyadenylation factor, it may play a role in defining the polyadenylation site or regulating the length of poly(A) tail.

5.5 CONCLUDING REMARKS

The GSG protein family has been expanding over the last several years. There is accumulating evidence that the GSG domain mediates specific RNA binding. In the future, more work needs to be done to elucidate the functions of GSG proteins in various aspects of RNA metabolism. To achieve this, it is essential to identify the physiological RNA targets for individual GSG family members. Since many GSG proteins have potential signaling properties, another important issue that needs to be addressed is whether these proteins play

a role in signal transduction and, if so, whether they provide a link between signal transduction and RNA metabolism.

REFERENCES (FOR CHAPTERS 1 AND 5)

Abovich, N. and Rosbash, M. 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell, 89:403-412.

Arning, S., Gruter, P., Bilbe, G. and Kramer, A. 1996. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. RNA, 2:794-810.

Arts, G.-J., Fornerod, M. and Mattaj, I. W. 1998. Identification of a nuclear export receptor for tRNA. Curr. Biol., 8:305-314.

Ashburner, M., Thomson, P., Roote, J., Lasko, P. F., Grau, Y., El Messal, M., Roth, S. and Simpson, P. 1990. The genetics of a small autosomal region of *Drosophila melanogaster* containing the structural gene for alcohol dehydrogenase. VII. Characterization of the region around the *snail* and *cactus* loci. Genetics, 126:679-694.

Ashiya, M. and Grabowski, P. J. 1997. A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. RNA, 3:996-1015.

Ashley, C. T., Wilkinson, K. D., Reines, D. and Warren, S. T. 1993. FMR1 protein: conserved RNP family domains and selective RNA binding. Science, 262:563-566.

Baber, J. L., Libutti, D., Levens, D. and Tjandra, N. 1999. High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. J. Mol. Biol., 289:949-962.

Baehrecke, E. H. 1997. who encodes a KH RNA binding protein that functions in muscle development. Development, 124:1323-1332.

Bagrodia, S., Chackalaparampil, I., Kmiecik, T. E. and Shalloway, D. 1991. Altered tyrosine 527 phosphorylation and mitotic activation of p60c-src. Nature, 349:172-175.

Barbarese, E. 1991. Spatial distribution of myelin basic protein mRNA and polypeptide in *Quaking* oligodendrocytes in culture. J. Neursci. Res., 29:271-281.

Barlat, I., Maurier, F., Duchesne, M., Guitard, E., Tocque, B. and Schweighoffer, F. 1997. A

role for Sam68 in cell cycle progression antagonized by a spliced variant within the KH domain. J. Biol. Chem., 272:3129-3132.

Bartoszewicz, Z. P., Noronha, A. B., Fujita, N., Sato, S., Bo, L., Trapp, B. D. and Quarles, R. H. 1995. Abnormal expression and glycosylation of the large and small isoforms of myelin-associated glycoprotein in dysmelinating *quaking* mutants. J. Neurchem. Res., 41:27-38.

Bedford, M. T., Chan, D. C. and Leder, P. 1997. FBP WW domains and the Abl SH3 domain bind a specific class of proline-rich ligands. EMBO J., 16:2376-2383.

Bechade, C. R. P., Cisterni, C., Kalusch, R., La Bella, V., Pettmann, B. and Triller, A. 1999. Subcellular distribution of survival motor neuron (SMN) protein: possible involvement in nucleocytoplasmic and dendritic transport. Eur. J. Neurosci., 11:293-304.

Beggs, J. D., Teigelkamp, S. and Newman, A. J. 1995. The role of PRP8 protein in nuclear pre-mRNA splicing in yeast. J. Cell Sci., 19:101-105.

Bell, M. V., Hirst, M. C., Nakanori, Y., MacKinnon, R. N., Roche, A., Flint, T. J., Jacobs, P. A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U., Kerr, B., Turner, G., Lindenbaum, R. H., Winter, R., Pembrey, M., Thibodeau, S. and Davies, K. E. 1991. Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell, 64:861-866.

Bennett, M., Michaud, S., Kingston, J. and Reed, R. 1992. Protein components specifically associated with prespliceosome and spliceosome complexes. Genes Dev., 6:1986-2000.

Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G. and Reed, R. 1992. Differential binding of heterogeneous ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. Mol. Cell. Biol., 12:3165-3175.

Berg, J. M. 1986. Potential metal-binding domains in nucleic acid binding proteins. Science, 232:485-487.

Berglund, J. A., Chua, K., Abovich, N., Reed, R. and Rosbash, M. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell, 89:781-787.

Berglund, J. A., Abovich, N. and Rosbash, M. 1998a. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes Dev., 12:858-867.

Berglund, J. A., Fleming, M. L. and Rosbash, M. 1998b. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. RNA, 4:998-1006.

Bienroth, S., Wahle, E., Satler-Crazzolara, C. and Keller, W. 1991. Purification of the cleavage and polyadenylation factor involved in 3' processing of mRNA precursors. J. Biol. Chem., 266:19768-19776.

Bischoff, F. R. and Pontingl, H. 1995. Catalysis of guanine nucleotide exchange of Ran by RCC1 and stimulation of hydrolysis of Ran-bound GTP by Ran-GAP1. Methods Enzymol., 257:135-144.

Bode, V. C. 1984. Ethylsnitrosourea mutagenesis and the isolation of mutant alleles for specific genes located in the t region of the mouse chromosome 17. Genetics, 108:457-470.

Bogerd, H. P., Fridell, R. A., Benson, R. E., Hua, J. and Cullen, B. R. 1996. Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel in vivo randomization selection assay. Mol. Cell. Biol., 16:4207-4214.

Bonifaci, N., Moroianu, J., Radu, A. and Blobel, G. 1997. Karyopherin beta2 mediates nuclear import of a mRNA binding protein. Proc. Natl. Acad. Sci. USA, 94:5055-5060.

Bourne, H. R., Sanders, D. R. and McCormick, F. 1991. The GTPase superfamily: conserved structure and molecular mechanism. Nature, 349:117-127.

Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Haendler, B. and Jacob, M. 1982. U2 RNA shares a structural domain with U1, U4 and U5 RNAs. EMBO J., 1:1259-1265.

Braun, P. E., Horvath, E. and Edwards, A. M. 1990. Two isoforms of myelin associated glycoprotein accumulate in *quaking* mice: only the large polypeptide is phosphorylated. Dev. Neurosci., 12:286-292.

Bray, M., Prasad, S., Dubay, J. W., Hunter, E., Jeang, K.-T., Rekosh, D. and Hammarskjold,

M.-L. 1994. A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. Proc. Natl. Acad. Sci. USA, 91:1256-1260.

Brosi, R., Groning, K., Behrens, S. E., Luhrmann, R. and Kramer, A. 1993a. Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. Science, 262:102-105.

Brosi, R., Hauri, H. P. and Kramer, A. 1993b. Separation of splicing factor SF3 into two components and purification of SF3a activity. J. Biol. Chem., 268:17640-17646.

Bunnell, S. C., Henry, P. A., Kolluri, R., Kirchhausen, T., Rickles, R. J. and Berg, L. J. 1996. Identification of Itk/Tsk Src homology 3 domain ligands. J. Biol. Chem., 271:25646-25656.

Burd, C. G. and Dreyfuss, G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. Science, 265:615-621.

Cajal, S. R. y. 1903. Un sencillo metodo de coloracion selectiva del reticulo protoplasmatico y sus efectos en los diversos organos nerviosos de vertebrados e invertebrados. Trab. Lab. Invest. Biol., 2:129-221.

Carpino, N., Wisniewski, D., Strife, A., Marshak, D., Kobayashi, R., Stillman, B. and Clarkson, B. 1997. p62^{dok}: a constitutively tyrsoine phosphorylated, GAP-associated protein in chronic myelogenous leukemia progenitor cells. Cell, 88:197-204.

Caslini, C., Spinelli, O., Cazzaniga, G., Golay, J., De Gioia, L., Pedretti, A., Brevario, F., Amaru, R., Barbui, R., Biondi, A., Introna, M. and Rambaldi, A. 1997. Identification of two novel isoforms of the ZNF162 gene: a growing family of signal transduction and activator of RNA proteins. Genomics, 42:268-277.

Cavaloc, Y., Popielarz, M., Fuchs, J. P., Gattoni, R. and Stevenin, J. 1997. Characterization and cloning of the human splicing factor 9G8: A novel 35 kDa factor of the serine/arginine protein family. EMBO J., 13:2639-2649.

Chabot, B., Bisotto, S. and Vincent, M. 1995. The nuclear matrix phosphoprotein p225 associates with splicing complexes as part of the U4/U6.U5 tri-snRNP particle. Nucleic

Acids Res., 23:3206-3213.

Chabot, B. 1996. Directing alternative splicing: cast and scenarios. Trends Genet., 12:472-478.

Champion-Arnaud, P. and Reed, R. 1994. The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. Genes Dev., 8:1974-1983.

Charroux, B., Pellizzoni, L., Perkinson, R A., Shevchenko, A., Mann, M. and Dreyfuss, G. 1999. Gemin3: a novel DEAD box protein that interacts with SMN, the spinal muscular atrophy gene product, and is a component of gems. J. Cell Biol., 147:1181-1193.

Chen, T., Damaj, B., Herrerra, C., Lasko, P. and Richard, S. 1997. Self-association of the single-KH domain family members Sam68, GRP33, GLD-1 and Qk1: Role of the KH domain. Mol. Cell. Biol., 17:5707-5718.

Chen, T. and Richard, S. 1998. Structure-function analysis of Qk1: A lethal point mutation in mouse quaking prevents homodimerization. Mol. Cell. Biol., 18:4863-4871.

Chen, T., Boisvert, F.-M., Bazett-Jones, D. P. and Richard, S. 1999. A role for the GSG domain in localizing Sam68 to novel nuclear structures in cancer cell lines. Mol. Biol. Cell, 10:3015-3033.

Chen, T. and Richard, S. Unpublished data.

Cheng, S. C. and Abelson, J. 1987. Spliceosome assembly in yeast. Genes Dev., 1:1014-1027.

Chi, N. C., Adam, E. J. H. and Adam, S. A. 1995. Sequence and characterization of cytoplasmic nuclear import factor p97. J. Cell Biol., 130:265-274.

Chi, N. C., Adam, E. J. H. and Adam, S. A. 1997. Different binding domains for Ran-GTP and Ran-GDP/RanBP1 on nuclear import factor p97. J. Biol. Chem., 272:

Chi, N. C. and Adam, S. A. 1997. Functional domains in nuclear import factor p97 for binding the nuclear localization sequence receptor and the nuclear pore. Mol. Biol. Cell, 8:945-956.

Chiara, M. D., Palandjian, L., Kramer, R. F. and Reed, R. 1997. Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. EMBO J., 16:4746-4759.

Cho, E.-J., Takagi, T., Moore, C. R. and Buratowski, S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev., 11:3319-3326.

Chook, Y. M. and Blobel, G. 1999. Structure of the nuclear transport complex karyopherin-β2-Ran.GppNHp. Nature, 399:230-237.

Cingolani, G., Petosa, C. Weis, K. and Muller, C. W. 1999. Structure of importin-β bound to the IBB domain of importin-α. Nature, 399:221-229.

Cole, C. N. and Hammell, C. M. 1998. Nucleocytoplasmic transport: Driving and directing transport. Curr. Biol., 8:R368-R372.

Colman, D. R., Kreibich, G., Frey, A. B. and Sabatini, D. D. 1982. Synthesis and incorporation of myelin polypeptides into CNS myelin. J. Cell Biol., 95:598-608.

Connelly, S. and Manley, J. L. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev., 2:440-452.

Conti, E., Uy, M., Leighton, L., Blobel, G. and Kuriyan, J. 1998. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α. Cell, 94:193-204.

Cooke, C. and Alwine, J. C. 1996. Cap and 3' splice site similarly affect polyadenylation efficiency. Mol. Cell. Biol., 16:2579-2584.

Coppola, J. A., Field, A. S. and Luse, D. S. 1983. Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. Proc. Natl. Acad. Sci. USA, 80:1251-1255.

Corbett, A. and Silver, P. A. 1997. Nucleocytoplasmic transport of macromolecules. Microbiol. Mol. Biol. Rev., 61:193-211.

Corbin, F., Bouillon, M., Fortin, A., Morin, S. and Rousseau, F. 1997. The fragile X mental

retardation protein is associated with poly(A)+ mRNA in actively translating polyribosomes. Hum. Mol. Genet., 6:1465-1472.

Cox, R. D., Hugill, A., Shedlovsky, A., Noveroske, J. K., Best, S., Justice, M. J., Lehrach, H. and Dove, W. F. 1999. Contrasting effects of ENU induced embryonic lethal mutations of the *quaking* gene. Genomics, 57:333-341.

Cruz-Alvarez, M. and Pellicer, A. 1987. Cloning of a full-length complementary cDNA for as *Artemia salina* glycine-rich protein. J. Biol. Chem., 262:13377-13380.

Cullen, B. R. 1998. Retroviruses as model systems for the study of nuclear RNA export pathways. Virology, 249:203-210.

Dahanukar, A. and Wharton, R. P. 1996. The Nanos gradient in Drosophila embryos is generated by translational regulation. Genes Dev., 10:2610-2620.

Dahlberg, J. E. and Lund, E. 1998. Functions of the GTPase Ran in RNA export from the nucleus. Curr. Opin. Cell Biol., 10:400-408.

Darlix, J. L., Lapadat-Tapolsky, M., de Rocquigny, H. and Roques, B. 1995. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. J. Mol. Biol., 254:523-537.

Das, A. 1993. Control of transcription termination by RNA-binding proteins. Annu. Rev. Biochem., 62:893-930.

Davis, L. I. 1995. The nuclear pore complex. Annu. Rev. Biochem., 64:865-896.

de The, H., Lavau, C., Marchio, A., Chomienne, C., Degos, L. and Dejean, A. 1991. The PML-RARα fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR. Cell, 66:675-684.

DeBoulle, K., Verkerk, A. J. M. H., Reyniers, E., Vits, L., Hendrickx, J., Roy, B. V., Bos, F. V. D., DeGraaff, E., Oostra, B. A. and Willems, P. J. 1993. A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nature Genetics, 3:31-35.

DelPriore, V., Snay, C. A., Bahr, A. and Cole, C. N. 1996. The product of the Saccharomyces cerevisiae RSS1 gene, identified as a high-copy suppressor of the rat7-1

temperature-sensitive allele of the Rat7/NUP159 nucleoporin, is required efficient mRNA export. Mol. Biol. Cell, 7:1601-1621.

Derry, J. J., Richard, S., Chen, T., Ye, X., Vasioukhin, V. and Tyner, A. L. Sik/BRK phosphorylates Sam68 in the nucleus and negatively regulates its RNA binding ability. Submitted.

Di Fruscio, M., Chen, T., Bonyadi, S., Lasko, P. and Richard, S. 1998. The identification of two Drosophila KH domain proteins: KEP1 and SAM are members of the Sam68 family of GSG domain proteins. J. Biol. Chem., 273:30122-30130.

Di Fruscio, M., Chen, T. and Richard, S. 1999. Two novel Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis. Proc. Natl. Acad. Sci. USA, 96:2710-2715.

Dingwall, C. and Laskey, R. A. 1991. Nuclear targeting sequences: a consensus? Trends Biochem. Sci., 16:178-181.

Doye, V. and Hurt, T. 1997. From nucleoporins to nuclear pore complexes. Curr. Opin. Cell Biol., 9:401-411.

Draper, B. W., Mello, C. C., Bowerman, B., Hardin, J. and Priess, J. R. 1996. MEX-3 is a KH domain protein that regulates blastomere identity in early C. elegans embryos. Cell, 87:205-216.

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. and Burd, C. G. 1993. hnRNP proteins and the biogenesis of mRNA. Annu. Rev. Biochem., 62:289-321.

Drivas, G. T. 1990. Characterization of four novel ras-like genes expressed in a human teratocarcinoma cell line. Mol. Cell. Biol., 10:1793-1798.

Dyck, J. A., Maul, G. G., Miller, W. H., Chen, J. D., Kakizuka, A. and Evans, R. M. 1994. A novel macromolecular structure is a target of the promyelocyte-retinoic acid receptor oncoprotein. Cell, 76:333-343.

Eberhart, D. E., Malter, H. E., Feng, Y. and Warren, S. T. 1996. The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear

export signals. Hum. Mol. Genet., 5:1083-1091.

Ebersole, T. A., Chen, Q., Justice, M. J. and Artzt, K. 1996. The *quaking* gene unites signal transduction and RNA binding in the developing nervous system. Nature Genetics, 12:260-265.

Edery, I. and Sonenberg, N. 1985. Cap-dependent RNA splicing in a HeLa nuclear extract. Proc. Natl. Acad. Sci. USA, 82:7590-7594.

Edwalds-Gilbert, G., Veraldi, K. L. and Milcarek, C. 1997. Alternative poly(A) site selection in complex transcription units: means to an end? Nucleic Acids Res., 25:2547-2561.

Ephrussi, A., Dickinson, L. K. and Lehmann, R. 1991. Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. Cell, 66:37-50.

Ephrussi, A. and Lehmann, R. 1992. Induction of germ cell formation by oskar. Nature, 358:387-392.

Fakan, S. and Puvion, E. 1980. The ultrastructural visualization of nucleolar and extranucleolar RNA synthesis and distribution. Int. Rev. Cytol., 65:255-299.

Fakans, S. 1994. Perichromatin fibrils are in situ forms of nascent transcripts. Trends Cell Biol., 4:86-90.

Feng, Y., Absher, D., Eberhart, D. E., Brown, V., Malter, H. E. and Warren, S. T. 1997a. FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. Mol. Cell, 1:109-118.

Feng, Y., Gutekunst, C. A., Eberhart, D. E., Yi, H., Warren, S. T. and Hersch, S. M. 1997b. Fragile X mental retardation protein: nucleocytoplasmic shuttling and association with somatodendritic ribosomes. J. Neurosci., 17:1539-1547.

Feng, Y., Zhang, F., Lokey, L. K., Chastain, J. L., Lakkis, L., Eberhart, D. and Warren, S. T. 1995. Translational suppression by trinucleotide repeat expansion at FMR1. Science, 268:731-4.

Fischer, U., Sumpter, V., Sekine, M., Satoh, T. and Luhrmann, H. 1993. Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that

binds a transport receptor independently of the m_3G cap. EMBO J., 12:573-583.

Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W. and Luhrmann, R. 1995. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. Cell, 82:475-483.

Fischer, U., Liu, Q. and Dreyfuss, G. 1997. The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. Cell, 90:1023-1029.

Fischer, U. and Luhrmann, R. 1990. An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. Science, 249:786-790.

Flach, J., Bossie, M., Vogel, J., Corbett, A., Jinks, T., Willins, D. A. and Silver, P. A. 1994. A yeast RNA-binding protein shuttles between the nucleus and the cytoplasm. Mol. Cell. Biol., 14:8399-8407.

Fornerod, M., Ohno, M., Yoshida, M. and Mattaj, I. W. 1997. CRM1 is an export receptor for leucine-rich nuclear export signals. Cell, 90:1051-1060.

Frail, D. E. and Braun, P. E. 1985. Abnormal expression of the myelin-associated glycoprotein in the central nervous system of dysmyelinating mutant mice. J. Neurochem., 45:1071-1075.

Francis, R., Barton, M. K., Kimbel, J. and Schedl, T. 1995a. *gld-1*, a tumor suppressor gene required for oocyte development in *Caenorhabditis elegans*. Genetics, 139:579-606.

Francis, R., Maine, E. and Schedl, T. 1995b. Analysis of multiple roles of *gld-1* in germline development: interactions with the sex determination cascade and the *glp-1* signalling pathway. Genetics, 139:607-630.

Frank, D. and Guthrie, C. 1992. An essential splicing factor, *SLU7*, mediates 3' splice-site choice in yeast. Genes Dev., 6:2112-2124.

Fridell, R. A., Benson, R. E., Hua, J., Bogerd, H. P. and Cullen, B. R. 1996. A nuclear role for the fragile X mental retardation protein. EMBO J., 15:5408-5414.

Fridell, R. A., Fischer, U., Luhrmann, R., Meyers, B. E., Meinkoth, J. L., Malim, M. H. and

Cullen, B. R. 1996. Amphibian transcription factor IIIA proteins contain a sequence elements functionally equivalent to the nuclear export signal of human immunodeficiency virus type 1 Rev. Proc. Natl. Acad. Sci. USA, 93:2936-2940.

Fridell, R. A., Truant, R., Thorne, L., Benson, R. E. and Cullen, B. R. 1997. Nuclear import of hnRNP A1 is mediated by a novel cellular cofactor related to karyopherin-beta. J. Cell Sci., 110:1325-1331.

Fromont-Racine, M., Rain, J. C. and Legrain, P. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nature Genet., 16:277-282.

Fu, X.-D. 1995. The superfamily of arginine/serine-rich splicing factors. RNA, 1:663-680.

Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Pieretti, M., Sutcliffe, J., Richards, S., Verkerk, A. J. M. H., Holden, J. J. A., Fenwick, R. G., Jr., Warren, S. T., Oostra, B. A., Nelson, D. L. and Caskey, C. T. 1991. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell, 67:1047-1058.

Fujita, N., Sato, S., Ishiguro, H., Inuzuka, T., Baba, H., Kurihara, T., Takahashi, Y. and Miyatake, T. 1990. The large isoform of the myelin-associated glycoporotein is scarcly expressed in the quaking mouse brain. J. Neurochem., 55:1056-1059.

Fukuda, M., Asano, S., Nakamura, T., Adachi, M., Yoshida, M., Yanagida, M., and Nishida, E. 1997. CRM1 is responsible for intracellular transport mediated by the nuclear export signal. Nature, 390:308-311.

Fumagalli, S., Totty, N. F., Hsuan, J. J. and Courtneidge, S. A. 1994. A target for Src in mitosis. Nature, 368:871-874.

Fung, E. T., Lanahan, A., Worley, P. and Huganir, R. L. 1998. Identification of a Torpedo homolog of Sam68 that interacts with the synapse organizing protein rapsyn. FEBS Lett. 437:29-33.

Fusaki, N., Iwamatsu, A., Iwahima, M. and Fujisawa, J. 1997. Interaction between Sam68 and Src family tyrosine kinases, fyn and lck, in T cell receptor signaling. J. Biol. Chem., 272:6214-6219.

Fyrberg, C., Becker, J., Barthmaier, P., Mahaffey, J. and Fyrberg, E. 1997. A *Drosophia*

muscle-specific gene related to the mouse quaking locus. Gene, 197:315-323.

Fyrberg, C., Becker, J., Barthmaier, P., Mahaffey, J. and Fyrberg, E. 1998. A family of Drosophila genes encoding quaking-related Maxi-KH domains. Biochem. Genetics, 36:51-64.

Gall, J. G., Bellini, M., Wu, Z., and Murphy, C. 1999. Assembly of the nuclear transcription and procesing machinery: Cajal bodies (coiled bodies) and transcriptosomes. Mol. Biol. Cell, 10:4385-4402.

Gary, J. D. and Clarke, S. 1998. RNA and protein interactions modulated by protein arginine methylation. Prog. Nuc. Acid Res. Mol. Biol., 61:65-131.

Gavis, E. R., Lunsford, L., Bergsten, S. E. and Lehmann, R. 1996. A conserved 90 nucleotide element mediates translational repression of nanos RNA. Development, 122:2791-2800.

Gedeon, A. K., Baker, E., Robinson, H., Partington, M. W., Gross, B., Manca, A., Korn, B., Poustka, A., Yu, S., Sutherland, G. R. and Mulley, J. C. 1992. Fragile X syndrome without CCG amplification has an FMR1 deletion. Nat. Genet., 1:341-344.

Ghetti, A., Pinol-Roma, S., Michael, W. M., Morandi, C. and Dreyfuss, G. 1992. hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. Nucleic Acids Res., 20:3671-3678.

Gibson, T. J., Rice, P. M., Thompson, J. D., and Heringa, J. 1993. KH domains within the FMR1 sequence suggest that fragile X syndrome stems from a defect in RNA metabolism. Trends Biochem. Sci., 18:331-333.

Gibson, T. J., Thompson, J. D. and Heringa, J. 1993. The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. FEBS Lett., 324:361-366.

Gilmartin, G. M. and Nevins, J. R. 1989. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. Genes Dev., 3:2180-2189.

Gilmartin, G. M. and Nevins, J. R. 1991. Molecular analyses of two poly(A) site processing

factors that determine the recognition and efficiency of cleavage of the pre-mRNA. Mol. Cell. Biol., 11:2432-2438.

Goodwin, E. B., Okkema, P. G., Evans, T. C. and Kimble, J. 1993. Translational regulation of *tra-2* by its 3' untranslated region controls sexual identity in *C. elegans*. Cell, 75:329-339.

Gorbalenya, A. E. and Koonin, E. V. 1993. Helicases: amino acid sequence comparisons and structure-function relationships. Curr. Biol., 3:419-429.

Gorlich, D., Prehn, S., Laskey, R. A. and Hartmann, E. 1994. Isolation of a protein that is essential for the first step of nuclear import. Cell, 79:767-778.

Gorlich, D., Kostka, S., Kraft, R., Dingwall, C., Laskey, R. A., Hartmann, E. and Prehn, S. 1995. Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope. Curr. Biol., 5:383-392.

Gorlich, D., Henklein, P., Laskey, R. A. and Hartmann, E. 1996a. A 41 amino acid motif in importin-α confers binding to importin-β and hence transits into the nucleus. EMBO J., 15:1810-1817.

Gorlich, D., Kraft, R., Kostka, S., Vogel, F., Hartmann, E., Laskey, R., Mattaj, I. W. and Izzaurralde, E. 1996b. Importin provides a link between nuclear protein import and U snRNA export. Cell, 87:21-32.

Gorlich, D. 1997. Nuclear protein import. Curr. Opin. Cell Biol., 9:412-419.

Gozani, O., Feld, R. and Reed, R. 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev., 10:233-243.

Gozani, O., Potashkin, J. and Reed, R. 1998. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. Mol. Cell. Biol., 18:4752-4760.

Grossman, J. S., Meyer, M. I., Wang, Y. C., Mulligan, G. J., Kobayashi, R. and Helfman, D. M. 1998. The use of antibodies to the polypyrimidine tract binding protein (PTB) to analyze the protein components that assemble on alternatively spliced pre-mRNAs that use distant branch points. RNA, 4:613-625.

Gustavson, K.-H., Blomquist, H. and Holmgren, G. 1986. Prevalence of fragile-X syndrome in mentally retarded children in a Swedish county. Am. J. Med. Genet., 23:581-588.

Haghighat, A., Mader, S., Pause, A. and Sonenberg, N. 1995. Repression of cap-dependent translation by 4E-binding protein 1: competition with p220 for binding to eukaryotic initiation factor-4E. EMBO J., 14:5701-5709.

Hamm, J., Darzynkiewicy, E., Tahara, S. M. and Mattaj, I. W. 1990. The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. Cell, 62:569-577.

Hamm, J. and Lamond, A. I. 1998. Spliceosome assembly: The unwinding role of DEAD-box proteins. Curr. Biol., 8:R532-R534.

Hamm, J. and Mattaj, I. W. 1990. Monomethylated cap structures facilitate RNA export from the nucleus. Cell, 63:109-118.

Hardy, R. J., Loushin, C. L., Friedrich Jr., V. L., Chen, Q., Ebersole, T. A., Lazzarini, R. A. and Artzt, K. 1996. Neural cell type-specific expression of QKI proteins is altered in the *quaking* viable mutant mice. J. Neuroscience, 16:7941-7949.

Harrison, C. J., Jack, E. M. and Allen, T. D. 1983. The fragile X: a scanning electron microscopy study. J. Med. Genet., 20:280-285.

Hartmann, A. M., Nayler, O., Schwaiger, F. W., Obermeier, A. and Stamm, S. 1999. The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59^{fyn}. Mol. Biol. Cell, 10:3909-3926.

Hazelrigg, T. 1998. The destinies and destinations of RNAs. Cell, 95:451-460.

Hellen, C. U., Pestova, T. V., Litterst, M. and Wimmer, E. 1994. The cellular polypeptide p57 (pyrimidine tract-binding protein) binds to multiple sites in the poliovirus 5' nontranslated region. J. Virol., 68:941-950.

Hershey, J. W. B., Mathews, M. B. and Sonenberg, N. 1996. Translational control. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Hirst, M., Grewal, P., Flannery, A., Slatter, R., Maher, E., Barton, D., Fryns, J.-P. and Davies, K. 1995. Two new cases of FMR1 deletions associated with mental impairment. Am. J. Hum. Genet., 56:67-74.

Hodges, P. E. and Beggs, J. D. 1994. RNA splicing. U2 fulfills a commitment. Curr. Biol., 4:264-267.

Hodgkin, J. 1980. More sex-determination mutants of *Caenorhabditis elegans*. Genetics, 96:649-664.

Hodgkin, J. A. and Brenner, S. 1977. Mutations causing transformation of sexual phenotype in the nematode *Caenorhabditis elegans*. Genetics, 86:275-287.

Hopper, A. K., Traglia, H. M. and Dunst, R. W. 1990. The yeast RNA1 gene product necessary for RNA processing is located in the cytosol and apparently excluded from the nucleus. J. Cell Biol., 111:309-321.

Huang, S., Deerinck, T. J., Ellisman, M. H. and Spector, D. L. 1997. The dynamic organization of the perinucleolar compartment in the cell nucleus. J. Cell Biol., 137:965-974.

Huang, S., Deerinck, T. J., Ellisman, M. H. and Spector, D. L. 1998. The perinucleolar compartment and transcription. J. Cell Biol., 143:35-47.

Huang, S. and Spector, D. L. 1997. Pre-mRNA splicing: nuclear organization of factors and substrates. Eukaryotic Messenger RNA Processing: Frontiers in Molecular Biology. A. R. Krainers. IRL Press, Oxford, UK. 37-67.

Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M., and Luhrmann, R. 1998. Snurportin 1, an m_3G-specific nuclear import receptor with a novel domain structure. EMBO J., 17:4114-4126.

Hunter, C. P. and Kenyon, C. 1996. Spatial and temporal controls target pal-1 blastomere-specification activity to a single blastomere lineage in C. elegans embryos. Cell, 87:217-226.

Imamoto, N., Shimamoto, T., Kose, S., Takao, T., Tachibana, T., Matsubae, M., Sekimoto, T., Shimonishi, Y. and Yoneda, Y. 1995. The nuclear pore-targeting complex binds to nuclear pores after association with a karyophile. FEBS Lett., 368:415-419.

Ishidate, T., Yoshihara, S., Kawasaki, Y., Roy, B. C., Toyoshima, K. and Akiyama, T. 1997. Identification of a novel nuclear localization signal in Sam68. FEBS Lett., 409:237-241.

Izaurralde, E., Stepinski, J., Darzynkiewicz, E. and Mattaj, I. 1992. A cap binding protein that may mediate nuclear export of RNA polymerase II-transcribed RNAs. J. Cell Biol., 118:1287-1295.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E. and Mattaj, I. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell, 78:657-668.

Izaurralde, E., Lewis, J., Gamberi, C., Jarmolowski, A., McGuigan, C. and Mattaj, I. 1995a. A cap-binding protein complex mediating U snRNA export. Nature, 376:709-712.

Izaurralde, E., McGuigan, C. and Mattaj, I. W. 1995b. Nuclear localization of a cap-binding protein complex. Cold Spring Harbor Symp. Quant. Biol., 60:669-675.

Izaurralde, E., Kutay, U., von Kobbe, C., Mattaj, I. W. and Gorlich, D. 1997a. The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. EMBO J., 16:6535-6547.

Izaurralde, E., Jarmolowski, A., Beisel, C., Mattaj, I. W. and Dreyfuss, G. 1997b. A role for the M9 transport signal of hnRNP A1 in mRNA export. J. Cell Biol., 137:27-35.

Izaurralde, E. and Adam, S. 1998. Transport of macromolecules between the nucleus and the cytoplasm. RNA, 4:351-364.

Izaurralde, E. and Mattaj, I. W. 1995. RNA export. Cell, 81:153-159.

Jackson, R. J. and Kaminski, A. 1995. Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond. RNA, 1:985-1000.

Jackson, R. J. and Standart, N. 1990. Do the poly(A) tail and 3' untranslated region control mRNA translation? Cell, 62:15-24.

Jamison, S. F., Crow, A. and Garcia-Blanco, M. A. 1992. The spliceosome assembly pathway in mammalian extracts. Mol. Cell. Biol., 12:4279-4287.

Jan, E., Motzny, C. K., Graves, L. E. and Goodwin, E. B. 1999. The STAR protein, GLD-1, is a translational regulator of sexual identity in Caenorhabditis elegans. EMBO J, 18:258-269.

Jarmolowski, A., Boelens, W. C., Izaurralde, E. and Mattaj, I. W. 1994. Nuclear export of different classes of RNA is mediated by specific factors. J. Cell Biol., 124:627-635.

Johannes, G. and Sarnow, P. 1998. Cap-independent polysomal association of natural mRNAs encoding c-myc, BiP and eIF4G conferred by internal ribosome entry sites. RNA, 4:1500-1513.

Jones, A. R., Francis, R. and Schedl, T. 1996. GLD-1, a cytoplasmic protein essential for oocyte differentiation, shows stage- and sex-specific expression during *Caenorhabditis elegans* germline development. Dev. Biol., 180:165-183.

Jones, A. R. and Schedl, T. 1995. Mutations in GLD-1, a female germ cell-specific tumor suppressor gene in *C.elegans*, affect a conserved domain also found in Sam68. Genes Dev., 9:1491-1504.

Jove, R. and Manley, J. L. 1984. In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites. J. Biol. Chem., 259:8513-8521.

Justice, M. J. and Bode, V. C. 1988. Three ENU-induced alleles of the murine quaking locus are recessive embryonic lethal mutations. Genet. Res. Camb., 51:95-102.

Justice, M. J. and Bode, V. C. 1986. Induction of new mutations in a mouse t-haplotype using ethylnitrosaurea mutagenesis. Genet. Res. Camb., 47:187-192.

Kaech, S., Covic, L., Wyss, A. and Ballmer-Hofer, X. 1991. Association of p60c-src with polyoma virus middle-T antigen abrogating mitosis-specific activation. Nature, 350:431-433.

Kaffman, A., Rank, N. M., O'Neill, E. M., Huang, L. S. and O'Shea, E. K. 1998. The receptor MsN5 exports the phosphorylated transcription factor Pho4 out of the nucleus. Nature, 396:482-486.

Kakizuka, A., Miller, Jr.,W. H., Umesono, K., Warrell, Jr., R. P., Frankel, S. R., Murty, V.

V. V. S., Dmitrovsky, E. and Evans, R. M. 1991. Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RARα with a novel putative transcription factor, PML. Cell, 66:663-674.

Kalderon, D., Roberts, B. L., Richardson, W. D. and Smith, A. E. 1984. A short amino acid sequence able to specify nuclear location. Cell, 39:499-509.

Kaminski, A., Hunt, S. L., Patton, J. G. and Jackson, R. J. 1995. Dorect evidence that polypyrimidine tract binding protein (PTB) is essential for internal initiation of translation of encephalomyocarditis virus RNA. RNA, 1:924-938.

Kataoka, N., Ohno, M., Kangawa, K., Tokoro, Y. and Shimura, Y. 1994. Cloning of a complementary DNA encoding an 80 kilodalton nuclear cap binding protein. Nucl. Acids Res., 22:3861-3865.

Kataoka, N., Ohno, M., Moda, I. and Shimura, Y. 1995. Identification of the factors that interact with NCBP, an 80 kilodalton cap binding protein. Nucl. Acids Res., 23:3638-3641.

Keller, W., Bienroth, S., Lang, K. M. and Christofori, G. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. EMBO J., 10:4241-4249.

Keller, W. 1995. No end yet to messenger RNA 3' processing. Cell, 81:829-832.

Khandjian, E. W., Bardoni, B., Corbin, F., Sittler, A., Giroux, S., Heitz, D., Tremblay, S., Pinset, C., Montarras, D., Rousseau, F. and Mandel, J. 1998. Novel isoforms of the fragile X related protein FXR1P are expressed during myogenesis. Hum. Mol. Genet., 7:2121-2128.

Khandjian, E. W., Corbin, F., Woerly, S. and Rousseau, F. 1996. The fragile X mental retardation protein is associated with ribosomes. Nat. Genet., 12:91-93.

Kiledjian, M., Wang, X. and Liebhaber, S. A. 1995. Identification of two KH domain proteins in the α-globin mRNP stability complex. EMBO J., 14:4357-4364.

Kim, E., Du, L., Bregman, D. B. and Warren, S. L. 1997. Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. J. Cell Biol.,

136:19-28.

Kim-Ha, J., Kerr, K. and Macdonald, P. M. 1995. Translational regulation of oskar mRNA by Bruno, an ovarian RNA-binding protein, is essential. Cell, 81:403-412.

Kim-Ha, J., Smith, J. L. and MacDonald, P. M. 1991. *oskar* mRNA is localized to the posterior pole of the *Drosophila* oocyte. Cell, 66:23-35.

Kjems, J., Frankel, A. D., and Sharp, P. A. 1991. Specific regulation of mRNA splicing in vitro by a peptide from HIV-1 Rev. Cell, 67:169-178.

Klausner, R. D., Rouault, T. A. and Harford, J. B. 1993. Regulating the fate of mRNA: The control of cellular iron metabolism. Cell, 72:19-28.

Koepp, D. M. and Silver, P. A. 1996. A GTPase controlling nuclear trafficking: Running or walking RANdomly. Cell, 87:1-4.

Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A. and Manley, J. L. 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature, 368:119-124.

Koken, M. H. M., Puvion-Dutilleul, F., Guillemin, M. C., Viron, A., Linares-Cruz, G., Stuurman, N., de Jong, L., Szostecki, C., Calvo, F., Chomienne, C., Degos, L., Puvion, E. and de Thé, H. 1994. The t(15;17) translocation alters a nuclear body in a retinoic acid-reversible fashion. EMBO J., 13:1073-1083.

Konarska, M., Padgett, R. and Sharp, P. 1984. Recognition of cap structure in splicing in vitro of mRNA precursors. Cell, 38:731-736.

Konarska, M. M., Grabowski, P. J., Padgett, R. A. and Sharp, P. A. 1985. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature, 313:552-557.

Konarska, M. M. and Sharp, P. A. 1987. Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. Cell, 49:763-774.

Kondo, T., Furuta, T., Mitsunaga, K., Ebersole, T. A., Shichiri, M., Wu, J., Artzt, K., Yamamura, K. and Abe, K. 1999. Geomic organization and expression analysis of the

mouse qkI locus. Mammalian Genome, 10:662-669.

Kramer, A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. Annu. Rev. Biochem., 65:367-409.

Kramer, A. 1992. Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a pre-splicing complex. Mol. Cell. Biol., 12:4545-4552.

Krawczun, M. S., Jenkins, E. C. and Brown, W. T. 1985. Analysis of the fragile-X chromosome: localization and detection of the fragile site in high resolution preparations. Hum. Genet., 69:209-211.

Kremer, E. J., Pritchard, M., Lunch, M., Yu, S., Holman, K., Baker, E., Warren, S. T., Schlessenger, D., Sutherland, G. R. and Richards, R. I. 1991. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science, 252:1711-1714.

Kutay, U., Izaurralde, E., Bischoff, F. R., Mattaj, I. W. and Gorlich, D. 1997a. Dominant-negative mutants of importin- block multiple pathways of import and export through the nuclear pore complex. EMBO J., 16:1153-1163.

Kutay, U., Bischoff, F. R., Kostka, S., Kraft, R. and Gorlich, D. 1997b. Export of importin α from the nucleus is mediated by a specific nuclear transport factor. Cell, 90:1061-1071.

Kutay, U., Lipowsky, G., Izaurralde, E., Bischoff, F. R., Schwanzmaler, P., Hartmann, E. and Gorlich, D. 1998. Identification of a tRNA-specific nuclear export receptor. Mol. Cell, 1:359-369.

Kuwabara, P. E., Okkema, P. G. and Kimble, J. 1992. *tra-2* encodes a membrane protein and may mediate cell communication in the *Caenorhabditis elegans* sex determination pathway. Mol. Biol. Cell, 3:461-473.

Laggerbauer, B., Achsel, T. and Luhrmann, R. 1998. The human 200kD DEXH-box U5snRNP protein unwinds U4/U6 RNA duplexes in vitro. Proc. Natl. Acad. Sci. USA, 95:4188-4192.

Laird, A. D., Taylor, S. J., Oberst, M. and Shalloway, D. 1995. Raf-1 is activated during mitosis. J. Biol. Chem., 270:26742-26745.

Lamond, A. I. and Earnshaw, W. C. 1998. Structure and function in the nucleus. Science, 280:547-553.

Lang, V., Mege, D., Semichon, M., Gary-Gouy, H. and Bismuth, G. 1997. A dual participation of ZAP-70 and src protein tyrosine kinases is required for TCR-induced tyrosine phosphorylation of Sam68 in Jurkat T cells. Eur. J. Immunol., 27:3360-3367.

Lawe, D. C., Hahn, C. and Wong, A. J. 1997. The Nck SH2/SH3 adaptor protein is present in the nucleus with the nuclear protein Sam68. Oncogene, 14:223-231.

Lee, B., Matera, A. G., Ward, D. C. and Craft, J. 1996a. Association of RNase mitochondrial RNA processing enzyme with ribonuclease P in higher ordered structures in the nucleolus: a possible coordinate role in ribosome biogenesis. Proc. Natl. Acad. Sci. USA, 93:11471-11476.

Lee, J.-S. and Burr, J. G. 1999. Salpα and Salpβ, growth-arresting homologs of Sam68. Gene, 240:133-147.

Lee, M. S., Henry, M. and Silver, P. A. 1996b. A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. Genes Dev., 10:1233-1246.

Lewis, H. A., Chen, H., Edo, C., Buckanovich, R. J., Yang, Y. Y., Musunuru, K., Zhong, R., Darnell, R. B. and Burley, S. K. 1999. Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. Structure, 7:191-203.

Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., and Burley, S. K. 2000. Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. Cell, 100:323-332.

Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C. and Mattaj, I. W. 1996. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap proximal 5' splice site. Genes Dev., 10:1683-1698.

Lewis, J. D. and Izaurralde, E. 1997. The role of the cap structure in RNA processing and nuclear export. Eur. J. Biochem., 247:461-469.

Lin, C. H. and Patton, J. G. 1995. Regulation of alternative 3' splice site selection by constitutive splicing factors. RNA, 1:234-245.

Lin, Q., Taylor, S. J. and Shalloway, D. 1997. Specificity and determinants of Sam68 RNA binding. J. Biol. Chem., 272:27274-27280.

Liu, K. and Hanna, M. M. 1995. NusA contacts nascent RNA in Escherichia coli transcription complexes. J. Mol. Biol., 247:547-558.

Liu, Q. and Dreyfuss, G. 1996. A novel nuclear structure containing the survival of motor neurons protein. EMBO J., 15:3555-3565.

Liu, Q., Fischer, U., Wang, F. and Dreyfuss, G. 1997. The spinal muscular atrophy gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. Cell, 90:1013-1021.

Lo, P. C. and Frasch, M. 1997. A novel KH-domain protein mediates cell adhesion processes in Drosophila. Dev. Biol. (Orlando), 190:241-256.

Lock, P., Fumagalli, S., Polakis, P., McCormick, F. and Courtneidge, S. A. 1996. The human p62 cDNA encodes Sam68 and not the ras-GAP-associated p62 protein. Cell, 84:23-24.

Logan, J., Falck-Pedersen, E., Darnell, J. E. and Shenk, T. 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse b^{maj}-globin gene. Proc. Natl. Acad. Sci. USA, 84:8306-8310.

Lopez, A. J. 1998. Alternative splicingof pre-mRNA: developmental consequences and mechanisms of regulation. Annu. Rev. Genet., 32:279-305.

Lou, H., Gagel, R. F. and Berget, S. M. 1996. An intron enhancer recognized by splicing factors activates polyadenylation. Genes Dev., 10:208-219.

Lubs, J. A., Jr. 1969. A marker X chromosome. Am. J. Hum. Genet., 21:231-244.

Lugenbeel, K. A., Peier, A. M., Carson, N. L., Chudley, A. E. and Nelson, D. L. 1995. Intragenic loss of function mutations demonstrate the primary role of *FMR1* in fragile X syndrome. Nat. Genet., 10:483-485.

Luhrmann, R., Kastner, B. and Bach, M. 1990. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. Biochim. Biophys. Acta Gene Struct. Express., 1087:265-292.

Luo, Y., Yu, H. and Peterlin, B. M. 1994. Cellular protein modulates effects of human immunodeficiency virus type 1 Rev. J. Virol., 68:3850-3856.

Maa, M.-C., Leu, T.-H., Trandel, B. J., Chang, J.-H. and Parsons, S. J. 1994. A protein that is related to GTPase activating protein-associated p62 complexes with phospholipase Cγ. Mol. Cell. Biol., 14:5466-5473.

MacDonald, C. C., Wilusz, J. and Shenk, T. 1994. The 64kD subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Mol. Cell. Biol., 14:6647-6654.

Mader, S., Lee, H., Pause, A. and Sonenberg, N. 1995. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4γ and the translational repressors 4E-binding proteins. Mol. Cell. Biol., 15:4990-4997.

Madhani, H. D. and Guthrie, C. 1994. Dynamic RNA-RNA interactions in the spliceosome. Annu. Rev. Genet., 28:1-26.

Mahajan, R., Delphin, C., Guan, T., Gerace, L. and Melchior, F. 1997. A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex RanBP2. Cell, 88:97-107.

Mahone, M., Saffman, E. E. and Lasko, P. F. 1995. Localised Bicaudal-C RNA encodes a protein containing a KH domain, the RNA-binding motif for FMR1. EMBO J., 14:2043-55.

Malim, M. H., Bohnlein, S., Hauber, J. and Cullen, B. R. 1989a. Functional dissection of the HIV-1 Rev trans-activator-derivation of a trans-dominant repressor of Rev protein. Cell, 58:205-214.

Malim, M. H., Hauber, J., Le, S.-Y., Maizel, J. V. and Cullen, B. R. 1989b. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature, 338:254-257.

Malim, M. H., Tiley, L. S., McCarn, D. F., Rusche, J. R., Hauber, J. and Cullen, B. R. 1990.

HIV-1 structural gene expression requires binding of the Rev trans-acivator to its RNA target sequence. Cell, 60:675-683.

Manley, J. L. 1995. A complex protein assembly catalyze polyadenylation of mRNA precursors. Curr. Opin. Genet. Dev., 5:222-228.

Matera, A. G., Frey, M. R., Margelot, K. and Wolin, S. L. 1995. A perinucleolar compartment contains several RNA polymerase III transcripts as well as the polypyrimidine tract-binding protein, hnRNP I. J. Cell Biol., 129:1181-1193.

Matera, A. G. 1999. Nuclear bodies: multifaceted subdomains of the interchromotin space. Trends Cell Biol., 9:302-309.

Matera, A. G. and Frey, M. R. 1998. Coiled bodies and gems: janus or gemini? Am. J. Hum. Genet., 63:317-321.

Mattaj, I. W. 1986. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. Cell, 46:905-911.

Mattaj, I. W. 1988. U snRNP assembly and transport. Structure and function of major small nuclear ribonucleoprotein particles. M. Birnstiels. Springer Verlag, New York. 100-114.

Mattaj, I. W. 1998. Ribonucleoprotein assembly: Clues from spinal muscular atrophy. Curr. Biol., 8:R93-R95.

Mattaj, I. W. and Englmeier, L. 1998. Nucleocytoplasmic transport: the soluble phase. Annu. Rev. Biochem., 67:265-306.

Mattaj, I. W. and Nagai, K. 1995. Recruiting proteins to the RNA world. Nat. Struct. Biol., 2:518-522.

Matunis, M. J., Wu, J. and Blobel, G. 1998. SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex. J. Cell Biol., 140:499-509.

Matunis, M. J., Coutavas, E. and Blobel, G. 1996. A novel ubiquitin-like modification

modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. J. Cell Biol., 135:1457-1470.

Matunis, M. J., Michael, W. M. and Dreyfuss, G. 1992. Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. Mol. Cell. Biol., 12:164-171.

McBride, A. E., Schlegel, A. and Kirkegaard, K. 1996. Human protein Sam68 relocalization and interaction with poliovirus RNA polymerase in infected cells. Proc. Natl, Acad. Sci. USA, 93:2296-2301.

McBride, A. E., Taylor, S. J., Shalloway, D. and Kirkegaad, K. 1998. KH domain integrity is required for wild-type localization of Sam68. Exp. Cell Res., 241:84-95.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Amgen EST Program, Shuman, S. and Bentley, D. L. 1997. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev., 11:3306-3318.

Meijer, H., de Graaff, E., Merckx, D. M. L., Jongbloed, R. J. E., de Die-Smulders, C. E. M., Engelen, J. J. M., Fryns, J.-P., Curfs, P. M. G. and Oostra, B. A. 1994. A deletion of 1.6 kb proximal to the CGG repeat of the FMR1 gene causes the clinical phenotype of the fragile X syndrome. Hum. Mol. Genet., 3:615-620.

Melchior, F., Paschal, B., Evans, J. and Gerace, L. 1993. Inhibition of nuclear protein import by nonhydrolyzable analogues of GTP and identification of the small GTPase Ran/TC4 as an essential transport factor. J. Cell Biol., 123:1649-1659.

Merendino, L., Guth, S., Bilbao, D., Martinez, C. and Valcarcel, J. 1999. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF^{35} and the 3' splice site AG. Nature, 402:838-841.

Meyer, B. J. 1997. Sex determination and X chromosome dosage compensation. In Riddle, D. L., Blumenthal, T., Meyer, B. J. and Priess, J. R. (eds), C. elegans II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y. p.209-240.

Mezquita, J., Pau, M. and Mezquita, C. 1998. Four isoforms of the signal-transduction and RNA-binding protein QKI expressed during chicken spermatogenesis. Mol. Reproduct.

Dev., 50:70-78.

Michael, W., Choi, M. and Dreyfuss, G. 1995a. A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. Cell, 83:415-422.

Michael, W. M., Siomi, H., Choi, M., Pinol-Roma, S., Nakielny, S., Liu, Q. and Dreyfuss, G. 1995b. Signal sequences that target nuclear import and nuclear export of pre-mRNA binding proteins. Cold Spring Harbor Symp. Quant. Biol., 60:663-668.

Michael, W. M., Eder, P. S. and Dreyfuss, G. 1997. The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein. EMBO J., 16:3587-3598.

Michaud, S. and Reed, R. 1991. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev., 5:2534-2546.

Michaud, S. and Reed, R. 1993. A functional association between the 5' and 3' splice sites is established in the earliest prespliceosome complex (E) in mammals. Genes Dev., 7:1008-1020.

Misteli, T., Caceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F. and Spector, D. L. 1998. Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. J. Cell Biol., 143:297-307.

Misteli, T. and Spector, D. L. 1996. Serine/threonine phosphatase I modulates the subnuclear distribution of pre-mRNA splicing factors. Mol. Biol. Cell, 7:1559-1572.

Misteli, T. and Spector, D. L. 1999. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. Mol. Cell, 3:697-705.

Mohler, J. and Wieschaus, E. 1986. Dominant maternal-effect mutations of Drosophila melanogaster causing the production of double-abdomen embryos. Genetics, 112:803-822.

Monneron, A. and Bernhard, W. 1969. Fine structural organization of the interphase nucleus in some mammalian cells. J. Ultrastruct. Res., 27:266-288.

Moore, M. J., Query, C. C. and Sharp, P. A. 1993. Splicing of precursors to mRNAs by the

spliceosome. The RNA world. R. F. Gesteland and J. F. Atkinss. Cold Spring Harbor Laboratory Press, Cold Spring Harbor. 303-357.

Moore, M. J. and Sharp, P. A. 1993. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature, 365:364-368.

Moore, M. S. 1996. Protein translocation: nuclear export--out of the dark. Curr. Biol., 6:137-140.

Moore, M. S. and Blobel, G. 1993. The GTP-binding protein Ran/TC4 is required for protein import into the nucleus. Nature, 365:661-663.

Morioianu, J., Blobel, G. and Radu, A. 1995. Previously identified protein of uncertain function is karyopherin α and together with karyopherin β docks import substrate at nuclear pore complex. Proc. Natl. Acad. Sci. USA, 92:2008-2011.

Mortillaro, M. J., Blencowe, B. J., Wei, X., Nakayasu, H., Du, L., Warren, S. L., Sharp, P. A. and Berezney, R. 1996. A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. Proc. Natl. Acad. Sci. USA, 93:8253-8257.

Murphy, R. and Wente, S. R. 1996. An RNA-export mediator with an essential nuclear export signal. Nature, 383:357-360.

Murthy, K. G. and Manley, J. L. 1992. Characterization of the multisubunit cleavage polyadenylation specificity factor from calf thymus. J. Biol. Chem., 267:14804-14811.

Musco, G., Kharrat, A., Stier, G., Fraternali, F, Gibson, T. J., Nilges, M. and Pastore, A. 1997. The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. Nat. Struct. Biol., 4:712-716.

Musco, G., Stier, G., Joseph, c., Morelli, M. A. and Pastore, A. 1996. Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. Cell, 85:237-245.

Myer, B. E. and Malim, M. H. 1994. The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. Denes Dev., 8:1538-1547.

Nabel-Rosen, H., Dorevitch, N., Reuveny, A. and Volk, T. 1999. The balance between two isoforms of the Drosophila RNA-binding protein How controls tendon cell differentiation. Mol. Cell, 4:573-584.

Nakielny, S., Siomi, M. C., Siomi, H., Michael, W. M., Pollard, V. and Dreyfuss, G. 1996. Transportin: nuclear transport receptor of a novel nuclear protein import pathway. Exp. Cell Res., 229:261-266.

Nandabalan, K., Price, L., and Roeder, G. S. 1993. Mutations in U1 snRNA bypass the requirement for a cell type-specific splicing factor. Cell, 73:407-415.

Neugebauer, K. M. and Roth, M. B. 1997. Transcription units as RNA processing units. Genes Dev., 11:3279-3285.

Neuman de Vegvar, H. E. and Dahlberg, J. E. 1990. Nucleocytoplasmic transport and processing of small nuclear RNA precursors. Mol. Cell. Biol., 10:3365-3375.

Newman, A. and Norman, C. 1991. Mutations in yeast U5 snRNA alter the specificity of 5' splice-site cleavage. Cell, 65:115-123.

Newman, A. and Norman, C. 1992. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. Cell, 68:743-754.

Newman, A. J., TeigelKamp, S. and Beggs, J. D. 1995. snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. RNA, 1:968-980.

Nigg, E. A. 1997. Nucleocytoplasmic transport: signals, mechanisms and regulation. Nature, 386:779-787.

Nishi, K., Yoshida, M., Fujiwara, D., Nishikawa, M., Horinouchi, S. and Beppu, T. 1994. Leptomycin B targets a regulatory cascade of crm1, a fission yeast nuclear protein, involved in control of higher order chromosome structure and gene expression. J. Biol. Chem., 269:6320-6324.

Nussbaum, R. L. and Ledbetter, D. H. 1995. The fragile X syndrome. Metabolic basis of inherited disease. C. R. Scriver, A. Beaudet, W. S. Sly and D. Valles. MacGraw-Hill, New

york. 795-810.

Nussbaum, R. L. and Ledbetter, D. H. 1986. Fragile-X syndrome: a unique mutation in man. Annu. Rev. Genet., 20:109.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M. F. and Mandel, J.-L. 1991. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science, 252:1097-1102.

O'Keefe, R. T., Norman, C. and Newman, A. J. 1996. The invariant U5 snRNA loop I is dispensable for the first catalytic step of pre-mRNA splicing in yeast. Cell, 86:679-689.

Ohno, M., Fornerod, M. and Mattaj, I. W. 1998. Nucleocytoplasmic transport: the last 200 nanometers. Cell, 92:327-336.

Ohtsubo, M., Okazaki, H. and Nishimoto, T. 1989. The RCC1 protein, a regulator for the onset of chromosome condensation locates in the nucleus and binds to DNA. J. Cell Biol., 109:1389-1397.

Ossareh-Nazari, B., Bachelerie, F. and Dargemont, C. 1997. Evidence for a role of CRM1 in signal-mediated nuclear protein export. Science, 278:141-144.

Palacios, I., Hetzer, M., Adams, S. A. and Mattaj, I. W. 1997. Nuclear import of U snRNPs requires importin β. EMBO J., 16:6783-6792.

Pante, N. and Aebi, U. 1995. Exploring nuclear pore complex structure and function in molecular detail. J. Cell Sci., (suppl.) 19:1-11.

Parker, R., Siliciano, P. G. and Guthrie, C. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell, 49:229-239.

Parry, H. D., Scherly, D. and Mattaj, I. W. 1989. 'Snurpogenesis': The transcription and assembly of U snRNP components. Trends Biochem. Sci., 14:15-19.

Patton, J. G., Porro, E. B., Galceran, J., Tempst, P. and Nadal-Ginard, B. 1993. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. Genes Dev., 7:393-406.

Pellizzoni, L., Kataoka, N., Charroux, B. and Dreyfuss, G. 1998. A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. Cell, 95:615-624.

Pemberton, L. F., Blobel, G. and Rosenblum, J. S. 1998. Transport routes through the nuclear complex. Curr. Opin. Cell Biol., 10:392-399.

Pembrey, M. E., Winter, R. M. and Davies, K. E. 1986. Frafile X mental retardation: current controversies. Trends Neurosci., 9:58-62.

Perez, I., Lin, C. H., McAfee, J. G. and Patton, J. G. 1997. Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. RNA, 3:764-778.

Pikielny, C. W., Rymond, B. C. and Rosbash, M. 1986. Electrophoresis of ribonucleoproteins reveals an ordered pathway of yeast splicing complexes. Nature, 324:341-345.

Pinol-Roma, S. and Dreyfuss, G. 1992. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. Nature, 355:730-732.

Pinol-Roma, S. and Dreyfuss, G. 1993. hnRNP proteins: localization and transport between the nucleus and the cytoplasm. Trends Cell Biol., 3:151-155.

Pollard, V. W., Michael, W. M., Nakieiny, S., Siomi, M. C., Wang, F. and Dreyfuss, G. 1996. A novel receptor-mediated nuclear protein import pathway. Cell, 86:985-994.

Powell, D. M., Amaral, C. M., Wu, J. Y., Maniatis, T. and Greene, W. C. 1997. HIV Rev-dependent binding of SF2/ASF to the Rev response element: Possible role in Rev-mediated inhibition of HIV RNA splicing. Proc. Natl. Acad. Sci. USA, 94:973-978.

Proudfoot, N. J. 1996. Ending the message is not so simple. Cell, 87:779-781.

Proudfoot, N. J. 1989. How RNA polymerase II terminates transcription in higher eukaryotes. Trends Biochem. Sci., 14:105-110.

Query, C. C., Moore, M. J. and Sharp, P. A. 1994. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev., 8:587-597.

Quignon, F., De Bels, F., Koken, M., Feunteun, J., Ameisen, J. C. and de The, H. 1998. PML induces a novel caspase-independent death process. Nature Genet., 20:259-265.

Radu, A., Blobel, G. and Moore, M. S. 1995. Identification of a protein complex that is required for nuclear import and mediates docking of import substrates to distinct nucleoporins. Proc. Natl. Acad. Sci. USA, 92:1769-1773.

Raghunathan, P. L. and Guthrie, C. 1998. RNA unwinding in U4-U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr. Biol., 8:847-855.

Rain, J. C., Rafi, Z., Rhani, Z., Legrain, P. and Kramer, A. 1998. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. RNA, 4:551-565.

Rasmussen, E. B. and Lis, J. T. 1993. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc. Natl. Acad. Sci. USA, 90:7923-7927.

Reddy, R. and Busch, H. 1988. RNA sequences, structure, and modifications. Structure and function of major and minor small nuclear ribonucleoprotein particles. M. Birnstiels. Springer Verlag, New York. 1-37.

Reddy, T. R., Xu, W., Mau, J. K. L., Goodwin, C. D., Suhasini, M., Tang, H., Frimpong, K., Rose, D. W., and Wong-Staal, F. 1999. Inhibition of HIV replication by dominant negative mutants of Sam68, a functional homolog of HIV-1 Rev. Nat. Med., 5:635-642.

Reed, R. 1996. Initial splice-site recognition and pairing during pre-mRNA splicing. Curr. Opin. Genet. Dev., 6:215-220.

Rein, A. 1994. Retroviral RNA packaging: A review. Arch. Virol. Suppl., 9:513-522.

Resnick, R. J., Taylor, S. J., Lin, Q. and Shalloway, D. 1997. Phosphorylation of the Src substrate Sam68 by Cdc2 during mitosis. Oncogene, 15:1247-1253.

Rexach, M. and Blobel, G. 1995. Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. Cell, 83:683-692.

Richard, S., Yu, D., Blumer, K. J., Hausladen, D., Olszowy, M. W., Connelly, P. A. and

Shaw, A. S. 1995. Association of p62, a multi-functional SH2- and SH3-binding protein, with src-family tyrosine kinases, Grb2, and phospholipase Cγ-1. Mol. Cell. Biol., 15:186-197.

Richards, R. I. and Sutherland, G. R. 1992. Dynamic mutations: a new class of mutations causing human disease. Cell, 70:709-712

Robbins, J., Dilworth, S. M., Laskey, R. A. and Dingwall, C. 1991. Two interdependent basic domain in nucleoplasmin targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell, 64:615-623.

Robert, R., Timchenko, N. A., Miller, J. W., Reddy, S., Caskey, C. T., Swanson, M. S. and Timchenko, L. T. 1997. Altered phosphorylation and intracellular distribution of a (CUG)n triplet repeat RNA-binding protein in patients with myotonic dystrophy and in myotonin protein kinase knockout mice. Proc. Natl. Acad. Sci. USA, 94:13221-13226.

Roche, S., Fumagalli, S. and Courtneidge, S. A. 1995. Requirement for Src family protein tyrosine kinase in G2 for fibroblast cell division. Science, 269:1567-1569.

Rongo, C., Gavis, E. R. and Lehmann, R. 1995. Localization of oskar RNA regulates oskar translation and requires Oskar protein. Development, 121:2737-2746.

Roscigno, R. F. and Garcia-Blanco, M. A. 1995. SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. RNA, 1:692-706.

Rousseau, F., Heitz, D., Biancalana, V., Blumenfeld, S., Kretz, C., Boue, J., Tommerup, N., Van Der Hage, C., DeLozier-Blanchet, C. and Croquette, M. F. 1991. Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. N. Engl. J. Med., 325:1673-1681.

Rousseau, F., Rouillard, P., Morel, M. L., Khandjian, E W., and Morgan, K. 1995. Prevalence of carriers of premutation-size alleles of the FMR1 gene and implications for the population genetics of the fragile X syndrome. Am. J. Hum. Genet., 57:1006-1018.

Ruhl, M., Himmelspach, M., Bahr, G. M., Hammerschmid, F., Jaksche, H., Wolff, B., Aschauer, H., Farrington, G. K., Probst, H. and Bevec, D. 1993. Eukaryotic initiation factor 5A is a cellular target of the human immunodeficiency virus type 1 Rev activation domain

mediating transactivation. J. Cell Biol., 123:1309-1320.

Ruskin, B., Krainer, A. R., Maniatis, T. and Green, M. R. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell, 38:317-331.

Ruskin, B., Zamore, P. D. and Green, M. R. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell, 52:207-219.

Russell, I. D. and Tollervey, D. 1995. Yeast Nop3p has structural and functional similarities to mammalian pre-mRNA binding proteins. Eur. J. Cell Biol., 66:293-301.

Ryner, L. C., Takagaki, Y. and Manley, J. L. 1989. Sequence downstream of AAUAAA signals affect pre-mRNA cleavage and polyadenylation in vitro both directly and indirectly. Mol. Cell. Biol., 9:1759-1771.

Saccomanno, L., Loushin C., Jan, E., Punkay, E., Artzt, K. and Goodwin, E. B. 1999. The STAR protein Qk1-6 is a translational repressor. Proc. Natl. Acad. Sci., 96:12605-12610.

Sachs, A. B. 1993. Messenger RNA degradation in eukaryotes. Cell, 74:413-421.

Sachs, A. B., Sarnow, P. and Hentze, M. W. 1997. Starting at the beginning, middle, and the end: translation initiation in eukaryotes. Cell, 89:831-838.

Saffman, E. E., Styhler, S., Rother, K., Li, W., Richard, S. and Lasko, P. 1998. Premature translation of oskar in oocytes lacking the RNA binding protein Bicaudal-C. Mol. Cell. Biol., 18:4855-4862.

Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S. and Darnell, J. 1980. The addition of the 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. Cell, 19:69-78.

Samorajski, T., Friede, R. L. and Reimer, P. R. 1970. Hypomyelination in the quaking mice. J. Neuropath. Exper. Neur., 29:507-523.

Sanchez-Margalet, V. and Najib, S. 1999. p68 Sam is a substrate of the insulin receptor and associates with the SH2 domains of p85 PI3K. FEBS Lett., 455:307-310.

Schul, W., Groenhout, B., Koberna, K., Takagaki, Y., Jenny, A., Manders, E. M., Raska, I.,

van Driel, R. and de Jong, L. 1996. The RNA 3' cleavage factors CstF64 kDa and CPSF100 kDa are concentrated in nuclear domains closely associated with coiled bodies and newly synthesized RNA. EMBO J., 15:2883-2892.

Schupbach, T. and Wieschaus, E. 1991. Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. Genetics, 129:1119-1136.

Schwer, B. and Gross, C. 1998. Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J., 17:2086-2094.

Shapiro, P. S., Vaisberg, E., Hunt, A. J., Tolwinski, N. S., Whalen, A. M., McIntosh, R. and Ahn, N. G. 1998. Activation of MKK/ERK pathway during somatic cell mitosis: direct interaction of active ERK with kinetochores and regulation of the mitotic 3F3/2 phosphoantigen. J. Cell. Biol., 142:1533-1545.

Shatkin, A. J. 1985. mRNA cap binding proteins: essential factors for initiation of translation. Cell, 40:223-224.

Shatkin, A. J. 1976. Capping of eukaryotic mRNAs. Cell, 9:645-653.

Shedlovsky, A., King, T. R. and Dove, W. F. 1988. Saturation germ line mutagenesis of the murine t region including a lethal allele of the quaking locus. Proc. Natl. Acad. Sci., 85:180-184.

Shuman, S. 1995. Capping enzyme in eukaryotic mRNA synthesis. Prog. Nucleic Acid Res. Mol. Biol., 50:101-129.

Sidman, R. L., Dickie, M. M. and Appel, S. H. 1964. Mutant mice (quaking and jimpy) with deficient myelination in the central nervous system. Science, 144:309-311.

Siebel, C. W., Admon, A., and Rio, D. C. 1995. Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. Genes Dev., 9:269-283.

Singh, R., Valcarcel, J. and Green, M. R. 1995. Distinct binding specificities and function of higher eukaryotic polypyrimidine tract-binding proteins. Science, 268:1173-1176.

Singh, R. and Reddy, R. 1989. γ-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. Proc. Natl. Acad. Sci. USA, 86:8280-8283.

Singleton, D. R., Chen, S., Hitomi, M., Kumagai, C. and Tartakoff, A. M. 1995. A yeast protein that bidirectionally affects nucleocytoplasmic transport. J. Cell Sci., 108:265-272.

Siomi, H. and Dreyfuss, G. 1995. A nuclear localization domain in the hnRNP A1 protein. J. Cell Biol., 129:551-560.

Siomi, H., Choi, M., Siomi, M. C., Nussbaum, R. L. and Dreyfuss, G. 1994. Essential role for KH domain in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. Cell, 77:33-39.

Siomi, H., Matunis, M. J., Michael, W. M. and Dreyfuss, G. 1993a. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. Nucl. Acids. Res., 21:1193-1198.

Siomi, H., Siomi, M. C., Nussbaum, R. L. and Dreyfuss, G. 1993b. The protein product of the fragile X gene, FMR1 has characteristics of an RNA binding protein. Cell, 74:291-298.

Siomi, H. and Dreyfuss, G. 1997. RNA-binding proteins as regulators of gene expression. Curr. Opin. Genet. Dev., 7:345-353.

Siomi, M. C., Eder, P. S., Kataoka, N., Wan, L., Liu, Q. and Dreyfuss, G. 1997. Transportin-mediated nuclear import of hetrogeneous nuclear RNP proteins. J. Cell Biol., 138:1181-1192.

Sittler, A., Devys, D., Weber, C. and Mandel, J.-L. 1996. Alternative splicing of exon 14 determines nuclear or cytoplasmic localization of *fmr1* protein isoforms. Hum. Mol. Genet., 5:95-102.

Smibert, C. A., Wilson, J. E., Kerr, K. and Macdonald, P. M. 1996. smaug protein represses translation of unlocalized nanos mRNA in the Drosophila embryo. Genes Dev., 10:2600-2609.

Sonenberg, N. 1996. mRNA 5' cap binding protein eIF4E and control of cell growth. Translational Control. Hershey, J. W. B., Mathews, M. B. and Sonenberg, N. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sonenberg, N. and Gingras, A.-C. 1998. The mRNA 5' cap-binding protein eIF4E and control of cell growth. Curr. Opin. Cell Biol., 10:268-275.

Sontheimer, E. J. and Steitz, J. A. 1993. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science, 262:1989-1996.

Spector, D. L. 1993. Macromolecular domains within the cell nucleus. Annu. Rev. Cell Biol., 9:265-315.

Spector, D. L., Lark, G. and Huang, S. 1992. Differences in snRNP localization between transformed and nontransformed cells. Mol. Biol. Cell., 3:555-569.

Stade, K., Ford, C. S., Guthrie, C. and Weiss, K. 1997. Exportin 1 (Crm 1p) is an essential nuclear export factor. Cell, 90:1041-1050.

Staknis, D. and Reed, R. 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol. Cell. Biol., 14:7670-7682.

Staley, J. P. and Guthrie, C. 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. Cell, 92:315-326.

Steinmetz, E. J. 1997. Pre-mRNA processing and the CTD of RNA polymerase II: The tail that wags the dog? Cell, 89:491-494.

Stutz, F. and Rosbash, M. 1998. Nuclear RNA export. Genes Dev., 12:3309-3319.

Sutcliffe, J. S., Nelson, D. L., Zhang, F., Pieretti, M., Caskey, C. T., Saxe, D. and Warren, S. T. 1992. DNA methylation represses *FMR-1* transcription in fragile X syndrome. Hum. Mol. Genet., 1:397-400.

Swanson, M. S. and Dreyfuss, G. 1988. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. Mol. Cell. Biol., 8:2237-2241.

Takagaki, Y., Ryner, L. C. and Manley, J. L. 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. Genes Dev., 3:1711-1724.

Takagaki, Y., Manley, J. L., MacDonald, C., Wilusz, J. and Shenk, T. 1990. A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. Genes Dev., 4:2112-2120.

Takagaki, Y., MacDonald, C. C., Shenk, T. and Manley, J. L. 1991. The 64kD polyadenylation factor contains an RNP-type RNA binding domain and unusual auxilliary motifs. Proc. Natl. Acad. Sci. USA, 89:1403-1407.

Tamanini, F., Meijer, N., Verheij, C., Willems, P. J., Galjaard, H., Oostra, B. A. and Hoogeveen, A. T. 1996. FMRP is associated to the ribosomes via RNA. Hum. Mol. Genet., 5:809-813.

Tanaka, H., Abe, K. and Kim, C. H. 1997. Cloning and expression of the quaking gene in the Zebrafish embryo. Mech. Develop., 69:209-213.

Tange, T. O., Jensen, T. H. and Kjems, J. 1996. In vitro interaction between human immunodeficiency virus type 1 Rev protein and and splicing factor ASF/SF2-associated protein, p32. J. Biol. Chem., 271:10066-10072.

Taylor, S. J., Anafi, M., Pawson, T. and Shalloway, D. 1995. Functional interaction between c-src and its mitotic target, Sam68. J. Biol. Chem., 270:10120-10124.

Taylor, S. J. and Shalloway, D. 1994. An RNA-binding protein associated with src through its SH2 and SH3 domains in mitosis. Nature, 368:867-871.

Teigelkamp, S., Newman, A. J. and Beggs, J. D. 1995. Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. EMBO J., 14:2602-2612.

Timchenko, L. T., Miller, J. W., Timchenko, N. A., DeVore, D. R., Datar, K. V., Lin, L., Roberts, R., Caskey, C. T. and Swanson, M. S. 1996. Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. Nucleic Acids Res., 24:4407-4414.

Toda, T., Iida, A., Miwa, T., Nakamura, Y. and Imai, T. 1994. Isolation and characterization of a novel gene encoding nuclear protein at a locus (D11S636) tightly linked to multiple endocrine neoplasia type 1 (MEN1). Hum. Mol. Genet., 3:465-470.

Truant, R. Cullen, B. R. 1999. The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin β-dependent nuclear localization signals. Mol. Cell. Biol., 19:1210-1217.

Trub, T., Frantz, J. D., Miyazaki, M., Band, H. and Shoelson, S. E. 1997. The role of a lymphoid-restricted, Grb2-like SH3-SH2-SH3 protein in T cell receptor signaling. J. Biol. Chem., 272:894-902.

Turner, G., Daniel, A. and Frost, M. 1980a. X-linked mental retardation macro-orchidism and the X-q27 fragile site. J. Pediatr., 96:837-841.

Turner, G. and Opitz, J. M. 1980b. X-linked mental retardation. Am. J. Med. Genet., 7:407-415.

Ullmann, K. S., Powers, M. A. and Forbes, D. J. 1997. Nuclear export receptors: from importin to exportin. Cell, 90:967-970.

Urlaub, H., Kruft, V., Bischof, O., Mueller, E. C. and Wittmann-Liebold, B. 1995. Protein-rRNA binding features and their structural and functional implications in ribosomes ad determined by croo-linking studies. EMBO J., 14:4578-88.

Utans, U., Behrens, S. E., Luhrmann, R. and Kramer, A. 1992. A splicing factor that is inactivated during in vivo heat shock is functionally equivalent to the [U4/U6.U5] triple snRNP-specific proteins. Genes Dev., 6:631-641.

Vakan, P., McGuigan, C. and Mattaj, I. W. 1990. Domains of U4 and U6 snRNAs required for snRNP assembly and splicing complementation in Xenopus oocytes. EMBO J., 9:3397-3404.

Valcarcel, J. and Gebauer, F. 1997. Post-transcriptional regulation: the dawn of PTB. Curr. Biol., 7:R705-R708.

Venables, J. P., Vernet, C., Chew, S. L., Elliott, D. J., Cowmeadow, R. B., Wu, J., Cooke, H. J., Artzt, K. and Eperon, I. C. 1999. T-STAR/ETOILE: a novel relative of Sam68 that interacts with an RNA-binding protein implicated in spermatogenesis. Hum. Mol. Genet., 8:959-969.

Verkerk, A. J. M. H., Pieretti, M., Sutcliffe, J. S., Fu, Y.-H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F., Eussen, B. E., van Ommen, G. J. B., Blonden, L. A. J., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A. and Warren, S. T. 1991. Identification of a gene (FMR1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell, 65:905-14.

Vernet, C. and Artzt, K. 1997. STAR, a gene family involved in signal transduction and activation of RNA. Trends in Genet., 13:479-484.

Vincent, M., Lauriault, P., Dubois, M.-F., Lavoie, S., Bensaude, O. and Chabot, B. 1996. The nuclear matrix protein p225 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. Nucleic Acids Res., 24:4649-4652.

Visa, N., Alzhanova-Ericsson, A. T., Sun, X., Kiseleva, E., Bjorkroth, B., Wurtz, T. and Daneholdt, B. 1996. A pre-mRNA binding protein accompanies the RNA from the gene through the nuclear pores and into polysomes. Cell, 84:253-264.

Volk, T. 1999. Singling out Drosophila tendon cells - a dialogue beween two distinct cell types. Trends Genet., 15:448-453.

Wagner, J. D. O., Jankowsky, E., Company, M., Pyle, A. M. and Abelson, J. N. 1998. The DEAH-box protein Prp22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. EMBO J., 17:2926-2937.

Wahle, E. 1995. 3'-end cleavage and polyadenylation of mRNA precursors. Biochim. Biophy. Acta, 1261:183-194.

Wahle, E. and Keller, W. 1992. The biochemistry of 3' end cleavage and polyadenylation of mRNA precursors. Annu. Rev. Biochem., 61:419-440.

Wang, L. L., Richard, S. and Shaw, A. S. 1995. p62 association with RNA is regulated by tyrosine phosphorylation. J. Biol. Chem, 270:2010-2013.

Wang, Y., Wagner, J. D. O. and Guthrie, C. 1998. The DEAH-box splicing factor Prp16 unwinds RNA dulpexes in vitro. Curr. Biol., 8:441-451.

Wang, Z. G., Delva, L., Mirella, G., Rivi, R., Giorgio, M., Cordon-Cardo, C., Grosveld, F.

and Pandolfi, P. P. 1998a. Role of PML in cell growth and the retinoic acid pathway. Science, 279:1547-1551.

Wang, Z. G., Ruggero, D., Ronchetti, S., Zhong, S., Gaboli, M., Rivi, R. and Pandolfi, P. P. 1998b. Pml is essential for multiple apoptotic pathways. Nat. Genet., 20:266-272.

Waring, D. A. and Kenyon, C. 1990. Selective silencing of cell communication influences anteroposterior pattern formation in C. elegans. Cell, 60:123-131.

Waring, D. A. and Kenyon, C. 1991. Regulation of cellular responsiveness to inductive signals in the developing C. elegans nervous system. Nature, 350:712-715.

Warren, S. T. and Nelson, D. L. 1994. Advances in molecular analysis of fragile X syndrome. J. Am. Med. Assoc., 271:536-542.

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. Identification of Src, Fyn, and Lyn SH3-binding proteins: Implications for a function of SH3 domains. Mol. Cell. Biol., 14:4509-4521.

Weighardt, F., Biamonti, G. and Riva, S. 1995. Nucleo-cytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1. J. Cell Sci., 108:545-555.

Weis, K., Mattaj, I. W. and Lamond, A. I. 1995. Identification of hSRP1α as a functional receptor for nuclear localization sequences. Science, 268:1049-1053.

Weis, K., Ryder, U. and Lamond, A. I. 1996. The conserved amino-terminal domain of hSRP1α is essential for nuclear import. EMBO J., 15:1818-1825.

Weis, K., Rambaud, S., Lavau, C., Jansen, J., Carvalho, T., Carmo-Fonseca, M., Lamond, A. and Dejean, A. 1994. Retinoic acid regulates aberrant nuclear localization of PML-RARa in acute promyelocytic leukemia cells. Cell, 76:345-356.

Wen, W., Meinkoth, J. L., Tsien, R. Y. and Taylor, S. S. 1995. Identification of a signal for rapid export of proteins from the nucleus. Cell, 82:463-473.

Wharton, R. P. and Struhl, G. 1991. RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen nanos. Cell, 67:955-967.

Whitelaw, E. and Proudfoot, N. 1986. a-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human α globin gene. EMBO J., 5:2915-2922.

Will, C. L. and Luhrmann, R. 1997. Protein functions in pre-mRNA splicing. Curr. Opin. Cell Biol., 9:320-328.

Wilusz, J., Shenk, T., Takagaki, Y. and Manley, J. L. 1990. A multicomponent complex is required for the AAUAAA-dependent crosslinking of a 64kD protein to polyadenylation substrates. Mol. Cell. Biol., 10:1244-1248.

Wilusz, J. and Shenk, T. 1988. A 64kD nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. Cell, 52:221-228.

Witherell, G. W., Schultz-Witherell, C. S. and Wimmer, E. 1995. Cis-acting elements of the encephalomyocarditis virus internal ribosomal entry site. Virology, 214:660-663.

Wohrle, D., Kotzot, D., Hirst, M. C., Manca, A., Korn, B., Schmidt, A., Barbi, G., Rott, H. D., Poustka, A. and Davies, K. E. 1992. A microdeletion of less than 250 kb, including the proximal part of the FMR-1 gene and the fragile-X site, in a male with the clinical phenotype of fragile-X syndrome. Am. J. Hum. Genet., 51:299-306.

Wolff, B., Sanglier, J. J. and Wang, Y. 1997. Leptomycin B is an inhibitor of nuclear export: inhibition of nucleo-cytoplasmic translocation of the human immunodeficiency virus type 1 (HIV-1) Rev protein and Rev-dependent mRNA. Chem. Biol., 4:139-147.

Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M. F., Polakis, P. and McCormick, F. 1992. Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. Cell, 69:551-558.

Wrehlke, C., Schmitt-Wrede, H., Qiao, Z. and Wunderlich, F. 1997. Enhanced expression in spleen macrophages of the mouse homolog to the human putative tumor suppressor gene ZFM1. DNA Cell Biol., 16:761-767.

Wu, J., Zhou, L., Tonissen, K., Tee, R. and Artzt, K. 1999. The quaking I-5 protein (QKI-5)

has a novel nuclear localization signal and shuttles between the nucleus and the cytoplasm. J. Biol. Chem., 274:29202-29210.

Wu, J. and Manley, J. L. 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev., 3:1553-1561.

Wu, J. Y. and Maniatis, T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell, 75:1061-1070.

Wu, S., Romfo, C. M., Nilsen, T. W. and Green, M. R. 1999. Functional recognition of the 3' splice site AG by the splicing factor U2AF^{35}. Nature, 402:832-835.

Xu, Y., Reddy, T. R., Fischer, W. and Wong-Staal, F. 1996. A novel hnRNP specifically interacts with HIV-1 RRE RNA. J. Biomed. Sci., 3:82-91.

Yamanashi, Y. and Baltimore, D. 1997. Identification of the Abl- and rasGAP-asssociated 62 kDa protein as a docking protein, dok. Cell, 88:205-211.

Yu, S., Pritchard, M., Kremer, E., Lunch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J. C., Warren, S. T., Schlessenger, D., Sutherland, G. R. and Richards, R. I. 1991. Fragile-X geneotype characterized by an unstable region of DNA. Science, 252:1179-1181.

Zaffran, S., Astier, M., Gratecos, D. and Semeriva, M. 1997. The held out wings (how) Drosophila gene encodes a putative RNA binding protein involved in the control of muscular and cardiac activity. Development, 124:2087-2098.

Zamore, P. D., Patton, J. G. and Green, M. R. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. Nature, 355:609-614.

Zamore, P. D. and Green, M. R. 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc. Natl. Acad. Sci. USA, 86:9243-9247.

Zamore, P. D. and Green, M. R. 1991. Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. EMBO J., 10:207-214.

Zapp, M. L. and Green, M. R. 1989. Sequence-specific RNA binding by the HIV-1 Rev protein. Nature, 342:714-716.

Zecevic, M., Catling, A. D., Eblem, S. T., Renzi, L., Hittle, J. C., Yen, T. J., Gorbsky, G. J. and Weber, M. J. 1998. Active MAP kinase in mitosis: localization at kinetochores and association with the motor protein CENP-E. J. Cell. Biol., 142:1547-1558.

Zhang, Y., O'Connor, J. P., Siomi, M. C., Srinivasan, S., Dutra, A., Nussbaum, R. L. and Dreyfuss, G. 1995. The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. EMBO J., 14:5358-5366.

Zhao, J., Hyman, L. and Moore, C. 1999. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol. Mol. Biol. Rev., 63:405-445.

Zhuang, Y. and Weiner, A. M. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell, 46:827-835.

Zhuang, Y. and Weiner, A. M. 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev., 3:1545-1552.

Zieve, G. W. and Sauterer, R. A. 1990. Cell biology of the snRNP particles. Crit. Rev. Biochem. Mol. Biol., 25:1-46.

Zorio, D. A. R. and Blumenthal, T. 1999. Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. Nature, 402:835-838.

Zorn, A. M., Grow, M., Patterson, K. D., Ebersole, T. A., Chen, Q., Artzt, K. and Krieg, P. A. 1997. Remarkable sequence conservation of transcripts encoding amphibian and mammalian homologues of *quaking* a KH domain RNA-binding protein. Gene, 188:199-206.

Zorn, A. M. and Krieg, P. A. 1997. The KH domain protein encoded by quaking functions as a dimer and is essential for notochord development in Xenopus embryos. Genes & Development, 11:2176-2190.

CONTRIBUTIONS TO ORIGINAL KNOWLEDGE

The present work has focused on the biochemical properties and functions of the GSG domain and GSG domain-containing proteins Sam68 and Qk1. These studies have been published in peer-reviewed journals. The major contributions to original knowledge are summarized below:

1. The finding that the GSG domain mediates RNA binding and self-association and directs protein localization is unique.

2. The identification of Sam68/SLM nuclear bodies is a novel discovery.

3. The finding that failure to dimerize is the molecular defect of a *quaking* lethal point mutation is original.