Mutations in \textit{gld-1}, a female germ cell-specific tumor suppressor gene in \textit{Caenorhabditis elegans}, affect a conserved domain also found in Src-associated protein Sam68

Allan R. Jones and Tim Schedl

Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110 USA

The \textit{gld-1} gene of \textit{Caenorhabditis elegans} is a germ-line-specific tumor suppressor gene that is essential for oogenesis. We have cloned the \textit{gld-1} gene and find that it encodes two proteins that differ by 3 amino acids. The predicted proteins contain a \textasciitilde{}170-amino-acid region that we term the GSG domain (GRP33/Sam68/GLD-1), on the basis of significant similarity between GLD-1, GRP33 from shrimp, and the Src-associated protein Sam68 from mouse (also described as GAPap62 from humans). A conserved structural motif called the KH domain is found within the larger GSG domain, suggesting a biochemical function for GLD-1 protein in binding RNA. The importance of the GSG domain to the function of \textit{gld-1} in vivo is revealed by mutations that affect 5 different conserved GSG domain residues. These include missense mutations in an absolutely conserved residue of the KH domain that eliminate the tumor suppressor function of \textit{gld-1}.

[Key Words: \textit{C. elegans}; germ-line development; tumor suppressor; KH domain; GSG domain]

Received March 21, 1995; revised version accepted April 28, 1995.

An important contribution to our understanding of cancer biology has been the identification of tumor suppressor genes in which loss of gene function is a factor in tumor formation. A number of tumor suppressor genes are involved in the cell cycle [for recent review, see Hartwell and Kastan 1994]. These include cell cycle checkpoint genes [e.g., p53] and DNA repair genes [e.g., MSH2]. Other tumor suppressor genes, possibly including WT1 [Kreidberg et al. 1993], appear to be tied to development, acting in cell type specification or developmental aspects of cell cycle control.

Two recent reports describe genetic studies of \textit{gld-1}, a tumor suppressor gene that is required for normal oocyte development in the nematode \textit{Caenorhabditis elegans} [Francis et al. 1995a,b]. In wild-type \textit{C. elegans}, diploid animals with a single \textit{X} chromosome develop as males that make only sperm, whereas diploids with two \textit{X} chromosomes are hermaphrodites that produce some sperm before switching to the production of oocytes (Fig. 1). \textit{gld-1} has no essential function in the male germ line or soma but is absolutely required for oogenesis in XX hermaphrodites [Francis et al. 1995a]. In \textit{gld-1(null)} hermaphrodites, germ cells that should develop as oocytes enter the meiotic pathway normally and progress as far as the pachytene stage of meiotic prophase. However, these cells then exit meiotic prophase, re-enter a mitotic cell cycle, and proliferate ectopically to produce a germ-line tumor. Neither the somatic sexual environment nor the chromosomal sex appear important for generation of the hermaphrodite tumor. Instead, tumor formation requires only that the germ-line sex determination pathway be set in the female mode that normally leads to oogenesis [Francis et al. 1995b]. On the basis of these properties, \textit{gld-1} can be regarded as a tumor suppressor locus that acts to direct oocyte differentiation and meiotic prophase progression (Fig. 1, function 1).

Previous studies have indicated that in addition to its essential role in oogenesis, \textit{gld-1} is likely to participate in at least two other aspects of germ-line development. \textit{gld-1} appears to act to regulate mitotic proliferation negatively among premeiotic germ-line stem cells (Fig. 1, function 2), as a \textit{gld-1(null)} allele can partially suppress the premeiotic proliferation defects conferred by certain alleles of the \textit{glp-1} gene [Francis et al. 1995b]. In contrast to its role in oogenesis, the activity of \textit{gld-1} in premeiotic germ cells is nonessential, possibly because of genetic redundancy, and functions in both the male and hermaphrodite germ lines. The finding that \textit{gld-1} acts to regulate premeiotic proliferation negatively suggests...
Figure 1.  
gld-1 (+) functions in three aspects of C. elegans germ-line development. Diagrammed is one gonad arm from a wild-type adult hermaphrodite. Germ cell nuclei with characteristic chromosomal morphology are shown. Male germ-line development is complete by the adult and progression through meiotic prophase (distal mitotic); entry (transition zone) and progression through meiotic prophase (pachytene through diakinesis), and oocyte development are ongoing throughout adulthood (see Clifford et al. 1994). The three gld-1 (+) functions (see text) include (1) an essential function in directing oocyte development and progression through pachytene for germ cells that are developing in the female mode; (2) a nonessential and sex nonspecific function in negatively regulating proliferation in premeiotic germ cells; and (3) a nonessential function in male sex determination in the hermaphrodite germ line. The diagram is not meant to imply spatial or temporal modes of gld-1 function.

that the role of gld-1 in directing oocyte development may also involve negative regulation of mitotic cell cycle factors.

The third function of gld-1 is promotion of the male germ cell fate in the hermaphrodite germ line (Fig. 1, function 3). Because a reduced level of spermatogenesis is observed in XX hermaphrodites with only one functional gld-1 allele, the role of the gene in promoting spermatogenesis appears to be haploinsufficient. gld-1 is not required for sperm formation in either sex, however, and its role in promoting hermaphrodite spermatogenesis may be limited to assisting the set of genes (fog-1, fog-3, and the fem genes) that act to specify the male fate (Francis et al. 1995b). Several classes of gld-1 gain-of-function [gf] alleles have been characterized that cause transformations in germ-line sexual fate by poising the germ-line sex determination process. Two classes of gf alleles feminize the XX (hermaphroditism) germ line dominantly (causing all germ cells to develop as oocytes); a third class of gf mutations feminizes both the XX and XO (male) germ lines; and a fourth class masculinizes the XX germ line (causing germ cells to develop inappropriately as sperm instead of oocytes). Thus, gld-1 can be mutated in different ways to cause gf disruptions in the process that specifies germ-line sexual fates.

The genetic properties of gld-1, including its multiple germ-line functions and classes of mutant alleles, suggest that the gene's regulation and function are likely to be complex. This report presents a molecular study of gld-1 that includes sequence analysis of the wild-type gene and 32 mutants.

Results

Physical mapping of the gld-1 locus

gld-1 maps genetically to chromosome I between unc-13 and lin-10 (Francis et al. 1995a). This region is represented entirely by an ordered set of yeast artificial chromosome (YAC) and cosmid clones (Coulson et al. 1988). Through PCR deletion mapping (Barstead et al. 1991; see Materials and methods), we delimited the physical location of gld-1 to a five-cosmid interval (flanked by F29D11 and ZK1014; Fig. 2B).

Genomic Southern blot analyses revealed that two gld-1 alleles are correlated with restriction fragment length polymorphisms (RFLPs). We found an insertion of 2.3 kb associated with gld-1(q343), and a small deletion associated with the psoralen-induced allele, gld-1(q485). Both RFLPs are detected by a 6.8-kb subclone (pAJ37) of cosmid C29D7 (Fig. 2C).

Gene structure

We sequenced most of pAJ37 [data not shown]. The portion of this genomic sequence corresponding to gld-1 cDNAs [see below] is shown in Figure 3. All 32 gld-1 mutations [Fig. 6, below] affect the gene as defined in Figure 3, providing strong evidence that this is the gld-1 gene.

To determine the intron/exon structure of gld-1, we isolated cDNA clones from a mixed-stage cDNA library (see Materials and methods). One clone, judged to be nearly full length [pLAJ1], was sequenced in its entirety (Fig. 3). The intron/exon structure of gld-1 is summarized in Figure 2D. The cDNA contains a single open reading frame of 1392 nucleotides. A 5'-untranslated region (UTR) of 16 nts is trans-spliced to SL1 (see below). The 3' UTR is 812 nts and ends with a stretch of poly(A) 10 nts downstream of the consensus polyadenylation signal AATAAA. To determine the precise 5' end of the gld-1 transcript, we isolated additional cDNA clones by 5' rapid amplification of cDNA ends (RACE)-PCR (Frohman et al. 1988). Sequence of these clones [see Materials and methods] verified that the gld-1 transcript is trans-spliced to SL1 (Krause and Hirsh 1987).

Northern blot analysis detected a single band of ~2.5 kb in both mixed-stage poly(A)-selected RNA and total RNA from various developmental stages and germ-line mutants [described below; A. Jones and T. Schedl, unpubl.]. This transcript size is consistent with that of the cDNA pLAJ1 (2.3 kb). However, when sequencing 5' RACE–PCR products, we discovered evidence for an al-
Figure 2. A summary of the mapping of the gld-1 locus. (A) A genetic map showing the position of gld-1 in relation to other genetic markers on chromosome I (adapted from Francis et al. 1995a). Deficiencies (ozDf5, nDf24, and nDf25) are shown with approximate end points. (B) A contiguous physical map showing the cosmids from the physical mapping project (Coulson et al. 1988) that fall within the two PCR assays (opposing arrows) used to delimit the region containing gld-1 (see Materials and methods). (C) A restriction map showing the position of restriction sites as well as the two RFLPs associated with gld-1 alleles, as determined from Southern blotting and subcloning. (X)XhoI; (H)HindIII; (E)EcoRI; (P)PstI. q485 is a small deletion; q343 is a 2.3-kb insertion. The double bar indicates the subclone from cosmid C29D7 (pAJ37) that was sequenced. (D) A composite summary of the gene structure as determined from cDNA sequencing. Coding regions are shown as rectangles. The gene structure represents the two transcripts as described in the text, with the alternate splice on exon 6 occurring as shown. trans-splicing occurs at the 5' end of the transcript to the 22-nucleotide leader SLI (see text). Putative initiator methionine codon (ATG), stop codon (TAA), polyadenylation signal (AATAAA), and 3'-untranslated region (3' UTR) occur as shown. Regions of homology as described in the text are shown as shaded regions superimposed on the gene structure. The GSG domain (see text) is hatched and includes the single KH motif in exon 5. The CGA region (light shading; see Fig. 4) is in exon 7 as shown.

ternatively spliced transcript. One cDNA clone contained 9 additional nucleotides at the 5' end of exon 6. This alternative splice inserts 3 amino acids into the predicted protein (Fig. 3). We hypothesized that two transcripts are produced from the gld-1 locus that are colinear except for 9 nucleotides. Consistent with this hypothesis, reverse transcriptase–PCR (RT–PCR) experiments using primer sets made to amplify specifically only the alternative splice product give single-band products of the predicted sizes [data not shown].

Conserved domains in the GLD-1 predicted protein

Conceptual translation of the cDNA pLAJ1 yields a 463-amino-acid protein. BLAST data base searches (Altschul et al. 1990) revealed significant sequence identity with several proteins. These proteins are shown in alignment in Figure 4A, and diagrammatically in Figure 4B. The most extensive similarity is found with GTPase-activating protein (GAP)-associated p62 [GAPp62] from humans (Wong et al. 1992), and glycine-rich protein [GRP33] from brine shrimp [Cruz-Alvarez and Pellicer 1987]. The sequence reported for GAPp62 is highly related to a 68-kD protein from mouse that is associated with Src during mitosis, and we refer to this protein as Sam68 [for Src-associated in mitosis 68 kD; see Courtneidge and Fumagalli 1994]. The similarity between Sam68/GAPp62 and GRP33 has been reported previously (Wong et al. 1992). However, this original alignment is extended by >50 amino acids in the amino-terminal direction when GLD-1, Sam68, and GRP33 are aligned [Fig. 4A]. We term the ~170-amino-acid region of similarity shared by these three proteins a GSG domain [for GRP33, Sam68, and GLD-1]. Overall, the GSG domains of GLD-1 and Sam68 are 34% identical and 61% similar.

Within the GSG domain is a smaller motif proposed to be involved in RNA binding, termed the KH domain [Siomi et al. 1993a]. Both GRP33 and Sam68 have been identified as containing a KH motif [Gibson et al. 1993; Siomi et al. 1993b]. Other genes containing this motif include FMR-1, the gene responsible for fragile X syn-
drome in humans [Verkerk et al. 1991], and hnRNP K, a heteronuclear ribonucleoprotein from humans [Siomi et al. 1993a]. From the analysis of >40 potential KH motifs, Gibson et al. [1993] generated a structural prediction for the KH motif that consists of three β-strands and two α-helices in the order β-α-β. Shown in alignment in Figure 5 is the central α-β prediction of the KH motifs of several proteins. The importance of the KH motif to the in vivo function of GLD-1 is revealed by missense mutations in gld-1 that affect an absolutely conserved glycine residue [see below, circled in Fig. 5].

Two other proteins also show significant similarity to GLD-1, Sam68, and GRP33 (see Fig. 4): A predicted protein from C. elegans, BO280.11 [Wilson et al. 1994], and ZFM1 from humans [Toda et al. 1994]. We confirmed that a gene is transcribed from the BO280.11 region by isolating and sequencing cDNA clones [see Materials and methods]. Because this gene has a different intron/exon structure than predicted for BO280.11, we refer to the protein encoded by our cDNAs as BO280.11b. Similarity between BO280.11b and ZFM1 with the other GSG proteins does not extend as far amino-terminally, but both of these proteins contain a small region of similarity to GLD-1 that is just carboxy-terminal to the GSG domain (see Fig. 4). We call this short sequence the CGA [for Carboxy-terminal GSG domain-associated] region.

To identify conserved portions of the gld-1 transcripts and coding region, we cloned and sequenced cDNAs that include part of the coding region of the gld-1 gene from the male/female nematode species Caenorhabditis remanei [data not shown; see Materials and methods]. The gld-1 transcripts from both C. elegans and C. remanei contain identical trans-spliced SL1 leaders, and the same length [although not identical] 5′ UTRs. The partial coding region [corresponding to amino acids 1–370 in C. elegans] showed 88% identity, with 99% similarity within the GSG domain and CGA region. All of the amino acids mutated in the gld-1 missense alleles described below are conserved in C. remanei.

Molecular analysis of gld-1 mutations that affect oogenesis

The gld-1 gene performs an essential function in directing oogenesis and a nonessential role in promoting spermatogenesis in the hermaphrodite [Francis et al. 1995a]. Thirty-two gld-1 mutations, which abolish the major function of gld-1, have been placed into six classes (A–F) on the basis of the nature of germ-line phenotypes observed in homozygotes [Francis et al. 1995a]. We determined the molecular lesions in each gld-1 allele [Table 1; Fig. 6; see Materials and methods].

Class A gld-1 alleles are null or strong loss-of-function mutations, which abolish the major function of gld-1. In XX animals homozygous for class A alleles, presumptively female germ cells never undergo oogenesis but, instead, form a tumor.
emphasize that it occurs outside of the region of homology with GLD-1. The region of the Sam68 protein responsible for binding to SH2 and SH3 domains (Weng et al. 1994, Richard et al. 1995) is shown to contain lesions that are likely to eliminate gene function. In addition to another small deletion (oz127), other class A alleles contain premature stop codons (q365, q268, q495, and q93oz49), a splice site mutation (q93oz53), and missense mutations (oz89, q93oz55, q361, and oz170oz47). Of particular interest is that three of the class A missense mutations (oz89, q93oz55, and q361) affect a single residue [Gly-227 Ser or Asp] in the gld-1 KH motif. This glycine is absolutely conserved in the alignment of KH motifs (see Fig. 5).

Figure 4. An alignment of five proteins. (A) A hybrid alignment of GLD-1 in either three-, four-, or five-way comparisons performed by the CLUSTAL V alignment program (Higgins et al. 1992). Aligned with GLD-1 are five proteins, GAP-associated p62 from humans (Sam68/GAP-p62, see text), glycine-rich protein from Artemia salina [GRP33, see text], ZFM1 from humans (Toda et al. 1994), and BO280.1lb from C. elegans (see text). Amino acid numbering is given at left, and nonaligned portions are not shown for clarity, the five-way alignment begins at residue 138 for ZFM1 and 125 for BO280.1lb. For GLD-1, amino acid numbering refers to the product predicted by the cDNA pLA1. The position of the alternative splice addition of 3 amino acids (see text) is shown inserted between amino acids 257 and 258 of the GLD-1 protein. For all alignments, amino acids that are present in a majority are shaded. Residues that are conserved in all proteins are listed below each line in uppercase letters, with conservative replacement [based on a score ≥0 on a BLOSUM 62 matrix (Henikoff and Henikoff 1993)] indicated by a colon (:). In the five-way alignment, lowercase letters are used for residues found in three or four of the aligned proteins. Amino acids that are missense for gld-1 mutations (Table 1) are circled. The KH motif and the GSG region as defined in the text are delimited as shown. BLAST searches reveal that several cDNAs identified as expressed sequence tags (ESTs) in corn, Arabidopsis, mouse, and human are also predicted to contain portions of the GSG domain and GCA region (not shown). (B) Diagram illustrating the positions of the various regions of similarity aligned in A within each of the five proteins. The GSG domain is delimited as shown [hatched region]. Also shown is the position of the single KH motif [solid region] that lies within this larger region of homology (see text) and the GCA region [shaded region]. The portion of the Sam68 protein responsible for binding to SH2 and SH3 domains (Weng et al. 1994, Richard et al. 1995) is shown to emphasize that it occurs outside of the region of homology with GLD-1.

contain lesions that are likely to eliminate gene function. q485, the canonical genetic null mutation [Francis et al. 1995a], contains an 83-bp deletion that shifts the reading frame near the amino terminus of the predicted protein. In addition to another small deletion (oz127), other class A alleles contain premature stop codons (q365, q268, q495, and q93oz49), a splice site mutation (q93oz53), and missense mutations (oz89, q93oz55, q361, and oz170oz47). Of particular interest is that three of the class A missense mutations (oz89, q93oz55, and q361) affect a single residue [Gly-227 Ser or Asp] in the gld-1 KH motif. This glycine is absolutely conserved in the alignment of KH motifs (see Fig. 5).

The class B, E, and F gld-1 alleles are partial reduction-of-function mutations that block oogenesis in either early or late stages. Class B alleles show a recessive phenotype in which presumptive female germ cells arrest in pachytene of meiosis I. Most class B alleles are intragenic revertants of the class C allele q93 [see below] and contain both the q93 mutation and an additional change that is either missense [q93oz56, q93oz12, and q93oz45] or affects a splice site [q93oz52]. oz116 is the only class B allele that was not obtained as an intragenic revertant. This mutation changes the 3' splice acceptor site in the last intron, from cagAGC to caaAGC. RT-PCR analysis revealed that the adjacent AG is used as a splice acceptor at an appreciable level in oz116 RNA (see Materials and methods), resulting in a reading frameshift predicted to produce a carboxy-truncated protein.

One class B allele [q93oz50] has a nonsense change in codon 13 of the predicted protein. Although this might be expected to be a molecular null mutation, q93oz50 does retain residual gene activity [Francis et al. 1995a]. Two hypotheses to explain the residual q93oz50 gene activity are exon skipping during splicing [Fisher et al. 1993] and translational initiation at a downstream AUG.

The phenotypic classes E and F are each represented by a single allele. Both alleles produce abnormal oocytes and behave as partial loss-of-function mutations with respect to oogenesis [Francis et al. 1995a]. The class E
**Figure 5.** An amino acid alignment of the KH domain. The amino acid sequences of several KH domains are shown. The alignment is based on several descriptions of the content of the KH domain that have placed >10 different proteins ranging from bacteria to higher eukaryotes into the KH domain family (cf. Gibson et al. 1993). The α-α-β structural predictions are delimited by lines above the alignment and are from Gibson et al. (1993). Shown in alignment are the KH domains from the human proteins hnRNP K [(1, 2, and 3); Siomi et al. 1993a], and FMR-1 [(1 and 2); Verkerk et al. 1991], a yeast protein HX [(1, 2, 3, and 4); Delahodde et al. 1986], GRP33 (Cruz-Alvarez and Pellicer 1987), Sam68/GAPap62 [Wong et al. 1992], ZFM1 (Toda et al. 1994), BO280.11b (see Materials and methods), and GLD-1 (Fig. 3). Conserved hydrophobic residues are boxed, with the absolutely conserved glycine residues in black. The glycine residue circled in the GLD-1 protein is missense in three gld-1 tumorous mutations (see text).

**Figure 6.** Molecular lesions in 32 alleles of gld-1. A summary of the position of each mutation in relation to the structure of the gene (see legend to Fig. 2D) by phenotypic class (see Table 1 and text). Blocked lines (for q267, q485, q266, and q126) indicate deletions, opposite pointing arrows indicate an insertion (for q343) not drawn to scale. For the intragenic revertants (e.g., q93oz50), only the new mutation is shown.

Molecular analysis of gld-1 mutations that affect hermaphrodite spermatogenesis

Aside from its essential role in directing oogenesis, gld-1 has a nonessential role in germ-line sex determination, promoting spermatogenesis in hermaphrodites. Class D gld-1 alleles form normal oocytes. However, these two alleles eliminate hermaphrodite spermatogenesis [a feminization of the germ-line (Fog) phenotype] as the result of a gf defect. In addition, class D alleles show a novel gld-1 defect, they feminize the XO male germ line, caus-
Molecular analysis of gld-1

Table 1. Molecular lesions of gld-1 alleles

| Allele   | Phenotypic class | Nucleotide change | Predicted result of molecular lesion |
|----------|------------------|-------------------|--------------------------------------|
| q485     | A1 {tum}         | deletion of 481–563. | shift in reading frame, multiple stop codons |
| oz127    | A2 {tum}         | deletion of 261–270. | shift in reading frame, multiple stop codons |
| q93oz53  | A2 {tum}         | g 654 a           | 3' splice acceptor changed, exon 3 |
| q268, q495| A2 {tum}        | g 1202 a          | Gly-227 Ser |
| q266     | A2 {tum}         | g 1032 a          | Gly-227 Asp |
| q365     | A2 {tum}         | c 1065 t          | Gln-238 Stop {uaq} |
| q93oz12  | A2 {tum}         | g 1007 a          | Trp-267 Stop {uaq} |
| q93oz45  | A2 {tum}         | c 1275 t          | Gln-292 Stop {uaq} |
| oz17oz47 | A2 {tum}         | g 1374 a          | Asp-310 Asn |
| q93oz50  | B {undiff.}      | c 206 t           | Gln-14 Stop {uaa} |
| q93oz56  | B {undiff.}      | g 907 a           | Gly-200 Glu |
| q93oz12  | B {undiff.}      | c 958 t           | Pro-217 Leu |
| q93oz52  | B {undiff.}      | g 1007 a          | Ala-294 Thr |
| q93oz45  | B {undiff.}      | g 1281 a          | 3' splice acceptor changed, exon 5 |
| oz116    | B {undiff.}      | g 1617 a          | Ala-294 Thr |
| q93, oz17, oz34, oz35, q62 | C1 {Mog} | g 1095 a          | Gly-248 Arg |
| oz30, oz16, oz29, oz33, oz70 | C2 {Mog} | g 1101 a          | Gly-250 Arg |
| oz10     | C3 {Mog}         | 1. c 1449 t       | 1. Pro-335 Ser |
|          |                  | 2. deletion of    | 2. deletion within the 3’ UTR |
|          |                  | 1911–2426         | |
| q126, oz142 | D {Fog}    | g 1369 a          | Gly-308 Glu |
| q266     | E {Fog/abn. oocyte} | deletion of 1412–1438 | amino acids 322–331 deleted, Ile-323 Met |
| q343     | F {abn. oocyte}  | insertion after 2164 | insertion of Tc2 in the 3' UTR |

Molecular lesions were determined as described in the text and Materials and methods. The numbering of the altered nucleotides is given in Fig. 3. Amino acids affected are numbered according to the predicted protein from the major transcript, represented by the cDNA pLAJ1 (see text). Phenotypic class designations are defined in Francis et al. (1995a). Both class A1 and A2 alleles have a recessive tumorous germ-line phenotype {tum}, where presumptive female germ cells exit pachytene and return to mitotic proliferation. Class A1 is null for all aspects of gld-1 function, whereas A2 alleles display gf poisoning activity that results in feminization of the germ-line phenotype. Class A2 alleles show a recessive phenotype in which presumptive female germ cells do not differentiate {undiff.} and are arrested in pachytene. Class C alleles all result in a recessive masculinization of the germ-line (Mog) phenotype. Homozygous germ- line defects result in masculinization of hermaphrodites and males, germ cells that would normally develop as sperm instead develop as oocytes. The class E allele shows a recessive abnormal oocyte (abn. oocyte) phenotype, where female germ cells fail to complete oocyte differentiation and undergo meiotic prophase development normally. In addition, the class E allele has a masculinization-specific Mog phenotype. For both classes D and E, the Mog phenotypes are the result of gf defects. The class F allele also displays an abnormal oocyte phenotype but does not have a Mog phenotype. Some class A2 and B alleles are intragenic revertants of the class C1 Mog mutations (e.g., q93oz50); both the original and a new mutation were found, but only the new mutation is presented. Missense alleles of gld-1 that affect residues conserved among Sam68, GRP33, ZFM1, and GLD-1 (see Fig. 4), are shown with the wild-type amino acid in boldface type.

ing males to make oocytes [Francis et al. 1995a]. The two class D mutations (q126 and oz142) contain identical nucleotide changes resulting in Gly-308 Glu. This mutation alters one of the few amino acids not conserved between GLD-1 and 2B280.11b in the CGA region (see Fig. 4A). The class E allele q266 also alters amino acids in the CGA region (see above) and, like many of the gld-1 alleles, has a gf effect on hemaphrodite spermatogenesis that results in a Mog phenotype. Unlike the class D alleles, however, q266 does not affect XO males overtly.

In contrast, class C alleles are rare gf mutations of gld-1 that increase hemaphrodite sperm production [a masculinization of the germ-line (Mog) phenotype]. On the basis of distinct genetic and phenotypic properties [Francis et al. 1995a] the 11 gld-1[Mog] alleles have been divided into three subclasses (C1, C2, and C3). Homozygous C1 and C2 mutants make excess sperm and fail to switch to oogenesis. Genetic tests indicate that these alleles, while having a gf defect for male sex determination (in the hemaphrodite), are partial loss of function for the major role of gld-1 in directing oogenesis. All class C1 mutations (q93, q62, oz17, oz34, and oz35) result in Gly-248 Arg, and all class C2 mutations (oz30, oz16, oz29, oz33, and oz70) result in Gly-250 Arg. Thus, missense mutations in two absolutely conserved glycine residues in the GSG domain (see Fig. 4A) are responsible for the fully penetrant Mog phenotype of all the Mog alleles.
Figure 7. gld-1 RNA expression in the germ line. A Northern blot probed with the cDNA clone pLAJ1 is shown. The first four lanes contain ~5 μg of total RNA from predominantly young adult-stage animals, with the animals in the first three lanes grown at 25°C (restrictive temperature for these mutations). gld-1 RNA is easily detectable in total RNA on overnight exposure and is presumably a relatively abundant RNA. The last lane contains ~0.5 μg of poly(A)+-selected RNA from a mixed developmental stage population of wild-type animals (N2). An act-1-specific probe was used as a control for loading (a gift from L. Schriefer, Washington University, St. Louis, MO).

frequency of 1×10^-5 (Francis et al. 1995a). Animals homozygous for the sole C3 allele [oz1O] produce a vast excess of sperm but sometimes switch to oogenesis and become self-fertile. The oz10 allele contains two mutations: a missense mutation [Pro-335 Ser] and a 515-bp deletion within the 3' UTR of the gene. This 3' UTR mutation has an obvious parallel to gf alleles of fern-3, which also contain 3' UTR mutations that produce a Mog phenotype (Ahringer and Kimble 1991); however, interpretation of the oz10 phenotype is confounded by the presence of the missense mutation.

Tissue specificity of gld-1 RNA

We analyzed gld-1 RNA levels in wild type (Fig. 7). A Northern blot of mixed-stage poly(A)+ RNA from wild type (N2) shows a single band of ~2.5 kb when probed with the cDNA clone pLAJ1; no additional bands are seen on extended exposures. In Northern blots of RNA from developmentally staged populations, a 2.5-kb band is also visible in the L3 larval stage and increases in abundance up to the adult stage (A. Jones, unpubl.). This single band presumably represents the two gld-1 transcripts identified by cDNA sequence and RT-PCR analysis.

To examine the tissue specificity of gld-1 RNA, total RNA was analyzed from adult populations of several mutants that affect the germ line. glp-4(bn2) is a conditional mutant that makes very few germ cells at the restrictive temperature (Beanan and Strome 1992). The absence of gld-1 transcript in total RNA from glp-4(bn2) adults (Fig. 7) suggests that the production of gld-1 RNA is dependent on, and perhaps limited to, the germ line. The 2.5-kb gld-1 RNA band is detected in total RNA from animals with germ lines undergoing both male [fem-3(q20gf); Barton et al. 1987] and female development [fem-1(hc17); Hodgkin 1986], consistent with the sex-nonspecific role of gld-1 in premeiotic proliferation (Francis et al. 1995b).

Analysis of the two gld-1 transcripts

Alternative splicing of gld-1 RNA produces transcripts whose predicted protein products differ by 3 amino acids. The alternative splice product of gld-1 is predicted to add Leu-Leu-Lys within the GSG domain (see Fig. 4A). Because these changes could add functionality or change the specificity of RNA binding (e.g., Gorlach et al. 1994) or possibly eliminate function (e.g., Liang et al. 1991), we wanted to address whether the alternative splice was regulated developmentally or sex specific. To do this we undertook radioactive RT-PCR using RNA samples in which gld-1 RNA is present at detectable levels on Northern blots. In all developmental stages tested, as well as in RNA samples from adult hermaphrodite animals that were making only sperm [fem-3(q20gf)] or only oocytes [fem-1(hc17)], the smaller RT-PCR product is present at roughly two to three times the amount of the larger product (Fig. 8). Thus, we found no evidence that
Figure 9. A comparison of the two splice sites for exon 6 in C. elegans and C. remanei. Genomic sequence for C. elegans corresponds to nucleotides 1155–1175 (see Fig. 3). Bars match identical nucleotides, and arrows indicate the two splice sites that are used. Translations of the coding regions are shown above and below the respective sequences, and both the nucleotides and amino acids are italicized for the longer transcript.

selection of either splice site is developmentally regulated. None of the gld-1 mutations described above affect the region of alternative splicing.

Interestingly, we found evidence that both of these splice sites are conserved and used in splicing in the related species C. remanei [see Materials and methods]. As shown in Figure 9, the nucleotides around each splice site are conserved but the three additional amino acids encoded by the second splice site are not conserved. Because both of the spliced forms are present in C. remanei [a male/female species], we think it is unlikely that alternative splicing is used to control the gld-1 activity that promotes hermaphrodite spermatogenesis. We do not have information about the relative abundance of each of these transcripts in C. remanei.

Discussion

gld-1 has a tumor suppressor function in vivo

The gld-1 gene functions as a female germ cell-specific tumor suppressor. XX animals homozygous for the genetic null allele gld-1(q485) exhibit a phenotype in which germ cells fail to undergo oogenesis and proliferate ectopically, forming a germ-line tumor (Francis et al. 1995a). The results of the molecular analysis of gld-1(q485) are consistent with the genetic interpretation of this allele as null for the locus: q485 contains an amino-terminal frameshifting deletion and fails to make a gene product [A. Jones and T. Schedl, unpubl.].

GLD-1 contains evolutionarily conserved domains

The predicted GLD-1 protein has extensive similarity over a third of its length with two proteins: Sam68 and GRP33. We call this region, comprising ~170 amino acids, a GSG domain [see Fig. 4]. GRP33 is a putative hnRNP from brine shrimp [Cruz-Alvarez and Pellicer 1987] but has not been studied extensively since its identification. The mouse phosphoprotein Sam68 associates with Src in a cell cycle-dependent manner [Fumagalli et al. 1994; Taylor and Shalloway 1994] and is thought to be nearly identical to the sequence reported for GAP-associated p62 from humans [see Courtneidge and Fumagalli (1994) for clarification of nomenclature]. Sam68 contains several regions with the potential to interact with the Src homology regions SH2 and SH3 [Weng et al. 1994; Richard et al. 1995]; however, these regions of Sam68 lie outside the GSG domain and are not found in GLD-1 or GRP33.

The importance of the GSG domain to the in vivo function of GLD-1 is revealed by 16 independent gld-1 missense mutations that occur throughout this region of the protein, many in absolutely conserved residues [see Table 1; Fig. 4]. These mutations affect gld-1 function in various ways, resulting in one of three phenotypic classes of gld-1 mutations: Missense mutations in Gly-227 abolish the essential function of gld-1 in directing oogenesis and suppressing tumor formation [phenotypic class A2]; mutations in either Gly-248 or Gly-250 result in a gf defect that leads to the masculinization of the hermaphrodite germ line [Mog phenotype: classes C1 and C2]; and mutation in any one of four residues in the GSG domain can suppress class C1 mutations intragenically to produce a loss-of-function phenotype [either mutant class A2 or B].

Another C. elegans gene, BO280.11b, encodes part of the GSG domain [see Fig. 4B]. Within this partial GSG domain, GLD-1 is more similar to Sam68 [41 amino acids identical] than it is to BO280.11b [35 amino acids identical]. However, a second region of GLD-1 and BO280.11b is highly conserved [68% identical over 22 amino acids]. This small region, which we refer to as the CGA region, is just carboxy-terminal to the GSG domain and apparently not found in Sam68 and GRP33. This region, as well as a partial GSG domain, is also found in ZFM1, a gene that is tightly linked to multiple endocrine neoplasia type 1 [MEN1] in humans [Toda et al. 1994]. Like the GSG domain, the CGA region is important for GLD-I function, as several gld-1 mutations affect it [refer to Table 1; and Fig. 4]: An intragenic mutation in Glu-310 suppresses the class C1 mutant defect, with the resulting double mutant having a tumorous phenotype [phenotypic class A2]; mutation of Gly-308 results in a gf defect leading to a feminization of the germ-line [Fog] phenotype in both the hermaphrodite and male but does not disrupt oocyte differentiation [class D]; finally, deletion of the last 4 amino acids of the CGA region results in a defect in oogenesis as well as a gf defect that results in a hermaphrodite-specific Fog phenotype [class E].

The gld-1 GSG domain may be involved in RNA binding

A potential biochemical function for the GSG domain is suggested by a general in vitro RNA-binding activity for Sam68 [Wong et al. 1992; Taylor and Shalloway 1994; Wang et al. 1995]. This ability to bind RNA may be attributable in part to the presence of a single KH motif within the larger region of similarity [see Fig. 4B]. A number of other proteins that contain the KH motif have been implicated in RNA metabolism [Gibson et al. 1993]. In most cases, the KH motif is present in multiple copies [Gibson et al. 1993; Siomi et al. 1993b]. GSG domain proteins, however, contain a single KH motif. Furthermore, the single KH motifs of GSG domain proteins
are more similar to each other than to any other KH motif [see Fig. 4]. The high degree of similarity between GSG domain proteins, as well as a comparison of GSG domain proteins to other KH motif-containing proteins in overall structure, provides additional support for the assertion that the GSG domain proteins represent a specialized family of KH motif-containing proteins (Gibson et al. 1993). Conserved regions of the GSG domain outside of the KH motif may themselves be critical for modulating an RNA-binding activity of the KH motif [i.e., affecting the binding affinity or specificity]. Alternatively, these regions may be involved in contacts with other proteins.

Potential RNA targets for binding by KH motif-containing proteins have not been well characterized. Specific RNA targets for Sam68 binding have not been reported. FMR-1 has been demonstrated to bind its own transcript but has also been reported to bind <4% of all mRNA from human brain, implying a general nonspecific RNA-binding ability [Ashley et al. 1993]. However, a number of hnRNPs, originally thought to be nonspecific RNA-binding proteins, have been found to bind to specific targets with high affinity [e.g., Dreyfuss et al. 1993; Burd and Dreyfuss 1994; Gorlach et al. 1994]. In addition, in vitro experiments with hnRNP K suggest that the KH motif may also mediate DNA binding [Takimoto et al. 1993]. GLD-1 protein appears to be absent from the nucleus as determined by antibody staining [A. Jones and T. Schedl, in prep.]. Therefore, if GLD-1 is an RNA-binding protein, it must likely functions in the cytoplasm to control translation or mRNA stability.

Mutations in the KH motif of GLD-1 eliminate the major in vivo function of GLD-1 in directing oogenesis. Two different substitutions of an absolutely conserved glycine in the KH motif of gld-1 [see Fig. 5] result in the formation of germ-line tumors that are similar to those found in the genetic null allele q485 [Francis et al. 1995a]. Thus, missense mutations in the KH motif of GLD-1 result in loss of the GLD-1 tumor suppressor function in vivo.

gld-1 mutant protein may poison other gene products

Many gld-1 alleles behave in a dominant manner with respect to the nonessential role of gld-1 in promoting hermaphrodite spermatogenesis. The presence or absence of sperm in a hermaphrodite is easily scored, and the effect of various gld-1 alleles on this function was quantified [Francis et al. 1995a]. The analysis revealed a weak haploinsufficient Fog phenotype for large deletions of the entire locus as well as the class A1 genetic null allele q485.

Significantly higher penetrance of the Fog phenotype was observed for all other class A alleles [designated class A2], indicating that although they are nonfunctional for the major role of gld-1 in directing oogenesis, they exhibit gf activity that results in a poisoning or a dominant-negative effect on germ-line sex determination [Francis et al. 1995a]. Interestingly, the class A2 alleles that possess the strongest gf activity are the three alleles with missense mutations in the KH motif [q289, q930z55, and q361]. We believe that the simplest explanation for the in vivo behavior of the gld-1 KH mutations is that stable protein is made, but the portion of GLD-1 that carries out the major functions of the protein is catalytically or structurally inactive. The gf character of the KH motif mutants might result from stable mutant protein either titrating out a limited supply of a factor necessary for male sex determination or poisoning a GLD-1 multimer. We believe this effect involves the poisoning of other gene products because, unlike the gld-1 [null] mutation, KH motif mutations can cause tumor formation in XO male germ lines [R. Francis and T. Schedl, unpubl.].

Other class A2 alleles that are predicted to produce little or no stable gld-1 product show a less penetrant Fog phenotype than the KH motif mutations. These mutations include q2127, a small out-of-frame deletion, as well as four nonsense mutations [q365, q268, q495, and q930z49]. These alleles may produce a small amount of gld-1 product that, like the product produced by the KH mutation alleles, might poison hermaphrodite spermatogenesis. Concordant with this idea, there is a correlation between the level of gf poisoning and nonsense mutation position: The most carboxy-terminal nonsense mutation, q365, has a more penetrant phenotype than does q268, which in turn has a more penetrant phenotype than q495, the amino-terminal mutation. The poisoning interaction seen with the class A2 mutations may be attributable to protein–protein rather than protein–RNA interactions, as the q365 mutations should generate a GLD-1 protein that does not contain a complete KH motif. Candidates for interacting proteins include the products of any of the fem genes [Hodgkin 1986], fog-1 [Barton and Kimble 1990], or fog-3 [Ellis and Kimble 1995], all of which promote male germ-line sex determination. The gf class D and E mutations could produce a Fog phenotype by a similar mechanism.

Interestingly, KH motif missense mutations in FMR-1 from humans [De Boulle et al. 1993] and the Bic-C gene in Drosophila [Mahone et al. 1995] also lead to a phenotype that is more severe than the null phenotype. These mutant proteins could poison other molecules in a manner analogous to that of gld-1 KH motif mutations.

The gf class C mutations increase spermatogenesis in hermaphrodites, possibly by interfering with the gene products required for the switch from spermatogenesis to oogenesis [Francis et al. 1995a]. In addition, class C1 and C2 Mog mutants behave as partial loss-of-function mutations with respect to oogenesis [Francis et al. 1995b]. The molecular lesions in the class C1 and C2 Mog mutants occur in amino acids absolutely conserved among the GSG domain proteins [see Fig. 4A], adjacent to the KH motif. Given the position of these mutations in relation to the KH motif, the class C1 and C2 Mog mutations might act by failing to regulate the activity of the KH motif properly. Consistent with this hypothesis is the finding that the allele q930z55, which corresponds to the cis double mutant of the Mog allele q93 [class C1] and the tumorous allele oz89 [class A2], is phenotypi-
GLD-1 as a regulator of germ-line development

GLD-1 might function by regulating mRNA stability or translation. By functioning in this manner, GLD-1 could control aspects of germ-line development by acting as a specific regulator of a few key RNAs or a more general regulator of many RNAs. Germ-line translational control in *C. elegans* has been shown to be an important mode of regulating proliferation ([*gdp-1*]; Evans et al. 1994), male development ([*fem-3*]; Ahringer and Kimble 1991), and female development ([*tra-2*]; Goodwin et al. 1993). Although these genes could be regulated ostensibly by GLD-1, none stand out as likely targets: GLD-1 tumors still form in the absence of *gdp-1* product ([Francis et al. 1995b]), indicating that tumor formation is not caused by a failure to regulate *gdp-1* activity; epistasis analysis revealed that *gdp-1* acts downstream of *tra-2* ([Francis et al. 1995b]), and masculinization in the *gld-1* class C1 Mog allele may not result from increased *fem-3* activity ([Ellis and Kimble 1995]).

If GLD-1 functions as a translational regulator, how might it act as a female germ-line tumor suppressor? The *mos* proto-oncogene provides an example of a cell cycle control factor that is regulated translationally during meiotic prophase. In *Xenopus* oocytes, the *mos* serine-threonine protein kinase is needed for activation of maturation promoting factor in meiosis I and for meiosis II arrest (Yew et al. 1993). Although *mos* mRNA synthesis begins early in oogenesis, it is only translated in mature oocytes after the stimulus for meiotic maturation (Sagata et al. 1988). Furthermore, inappropriate expression of *mos* can result in tumor formation (Yew et al. 1993). In *gld-1* (null) mutants, mRNAs encoding factors that are synthesized for use late in oocyte development and are capable of promoting mitosis (such as *mos*) might be translated inappropriately. Alternatively, loss of *gld-1* activity might result in the inappropriate translation of mitotic factors packaged in the oocyte for embryonic cell divisions. Such ectopic mitotic activity early in oocyte development might then result in an exit from meiotic prophase, a return to mitotic proliferation, and subsequent tumor formation.

GLD-1, Sam68, and mitosis

*gld-1* has a demonstrated role in vivo for regulating mitotic activity negatively during pachytenic of meiotic prophase ([Francis et al. 1995a] and a nonessential, probably redundant role in regulating proliferation negatively in premeiotic germ cells ([Francis et al. 1995b]). Biochemical experiments show that the mammalian protein Sam68 interacts with Src in a cell cycle-specific manner ([Fumagalli et al. 1994; Taylor and Shalloway 1994] and is phosphorylated on tyrosine during mitosis ([Fumagalli et al. 1994]). Given that both Sam68 and GLD-1 contain GSG domains, it is tempting to speculate that both molecules might be involved in the regulation of mitosis in similar ways. However, the region of Sam68 that binds to Src shows no similarity to GLD-1. Interestingly, tyrosine phosphorylation of the carboxy terminus of Sam68 may regulate its RNA-binding activity ([Wang et al. 1995]). Regions outside of the GSG domain could regulate GLD-1 activity in an analogous manner. The similarity between GLD-1 and Sam68 within the GSG domain suggests that if these two proteins do have a regulatory role in mitosis in common, targets for interaction with the GSG domain (either proteins or RNAs) might be conserved across phyla as well.

Materials and methods

Physical mapping of the *gld-1* gene

Genetic mapping placed the *gld-1* locus in relation to three multilocus deletions, *nd24*, *nd25*, and *ozDf5* ([Fig. 2A]; Francis et al. 1995a). Using sequence information from various clones spanning this region, PCR assays were developed and used on individual homozygotes from each deletion ([Barstead et al. 1991]). PCR primer sets were derived from sequence of the cosmids *MO2B2* and *ZK1014* ([primers MO2B2 (a and b) and ZK1014 (a and b)]; Y. Kozono, pers. comm.) and sequence of the LRF gene, which resides on cosmid *F29D11* (primers *F29D11-a* and *b*; Yochem and Greenwald 1993). These assays placed *gld-1* between the cosmid clones *F29D11* and *ZK1014* on the physical map, and individual cosmids in this interval were used as probes on Southern blots to identify *gld-1* RFLPs ([Fig. 2B]; Coulson et al. 1988).

Transgenic experiments provide results that are consistent with identification of the *gld-1* gene. Six independent extra-chromosomal array-containing lines were generated by microinjecting *pA37* (wild-type *gld-1* at 10 μg/ml) and the dominant rolling marker plasmid *pRF4* (100 μg/ml; Mello et al. 1991). All stable lines produced variable dominant phenotypes similar to *gld-1* loss-of-function (ectopic proliferation, pachytenic arrest, abnormal oocytes, and feminization of the germ line) in a wild-type background, and all failed to rescue the tumorous phenotype after crossing into *gld-1(q485)* ([A. Jones unpubl.]). Similarly, wild-type transgenes of another germ-line locus, *gdp-1*, can produce a variable dominant *gld-1* loss-of-function germ-line phenotype in a wild-type background ([A. Fire, pers. comm.]). The basis of these results is not understood presently.

Nucleic acid preparation and analysis

Standard methods for manipulation of DNA and RNA were used ([Ausubel et al. 1991]). Sequence was determined for the majority of the 6.8-kb subclone *pA37* by methods similar to those described in Wilson et al. ([1994]). RNA was prepared from developmentally staged nematode populations as described ([Meyer and Casson 1986; Goetinck and Waterston 1994]).

Isolation and analysis of cDNA clones

Clones (10^9) from a *ZAP* cDNA library (a gift from R. Barstead, Oklahoma Medical Research Foundation, Oklahoma City) were screened using a 1.0-kb EcoRI–HindIII fragment as a probe ([see Fig. 2C]). Five clones were characterized by restriction
analysis and sequencing. Three appear to be identical and near full length, with the 5's end containing the last five nucleotides of the *C. elegans* trans-spliced leader SL1 (Krause and Hirsh 1987). One of these clones, pLA1, was sequenced entirely.

Analysis of the 5's end of the *gld-1* transcript was performed using the 5'RACE system (GIBCO/BRL) according to the manufacturer's instructions. Starting material was 1 μg of mixed-stage total RNA, with primer AlgR29a used to prime reverse transcriptions. In amplification reactions with the supplied anchor primer, both primers AlgR28 and AlgR23 gave single-band products as assayed on agarose gels. Amplification products were cloned, and the 5's ends of 28 independent cDNA clones were sequenced: 23 contained the full-length leader SL1, and 5 terminated at different points within the coding region. To look for additional differentially spliced products (other than the alternate splice described in Results), equivalent amounts of RNA from each developmental stage were combined and reverse transcribed with random primers as described below for analysis of the two known spliced forms. PCR analysis of this sample using numerous combinations of 5's specific primers did not provide evidence for additional alternative splicing of the *gld-1* transcript (data not shown).

**Characterization of** BO280.11b

A BLAST search with *gld-1* identified similarity to the product of the predicted gene BO280.11 on the *C. elegans* cosmid BO280 (GenBank accession no. U10438). This sequence was used to make primers for 5'RACE analysis as described above for *gld-1*. Reverse transcription was primed with AlgR032, and the cDNA was amplified with AlgR030 and anchor primer. A predominant band of ~600 bp was cloned. Four independent clones were sequenced, none of which contained a trans-spliced leader sequence. However, two clones with inserts of ~600 bp had identical 5's ends. The sequence of these clones revealed two important differences from the predicted splicing pattern. First, the 5'RACE clones do not contain the first five predicted exons of BO280.11, which code for a potential tyrosine phosphatase. Second, some of the remaining five exons use different splice sites than predicted. The resulting gene, which we refer to as BO280.11b, has the following structure (numbered according to cosmid BO280): exon 1 (33408-33398); exon 2 (33396-33851); exon 3 (33798-33474); exon 4 (33425-33328); and exon 5 (33281-33049). We believe the BO280.11b cDNAs represent an independent translational unit, as exon 1 contains stop codons in each reading frame before the putative initiator methionine codon starting at 34027. We have no information about the 3's end of BO280.11b.

**Isolation of gld-1 from *C. remanei***

To obtain sequence from the *C. remanei* *gld-1* gene, we used the primer pair AlgR16 and AlgR20 in low annealing temperature PCR reactions on *C. remanei* genomic DNA. Using the sequence of the amplification product, we designed the *C. remanei*-specific primers AlgR15, AlgR16, and AlgR15. 5'RACE cloning and sequencing, as described above, was performed on *C. remanei* RNA using AlgR16 and AlgR15. Several clones that contained a trans-spliced leader identical to the *C. elegans* trans-spliced leader SL1 (Krause and Hirsh 1987) were sequenced completely. Additional sequence was obtained from the RT-PCR product of primers AlgR16 and AlgR11.

The genomic sequence between exons 5 and 6 in *C. remanei* was determined from the amplification product of primers AlgR16 and AlgR8 on genomic DNA (sequence around the splice sites is shown in Fig. 9). On the basis of this sequence, we constructed the primer Alremaltf, which contains the last 12 nucleotides of exon 5 and the first 6 nucleotides of the alternate splice. Alremaltf, in combination with Alcr8, AlRcR9, or Acr11 gave single-band products in PCR on reverse-transcribed *C. remanei* RNA (data not shown), providing evidence that this splice site is used in *C. remanei*.

**Analysis of gld-1(oz116) splicing**

For the RT-PCR analysis of *gld-1(oz116)*, RNA was isolated from the strain nDp4; unc-13(e51) gld-1(oz116). Primer Acr20 was used to prime reverse transcription of 1 μg of total RNA. PCR was performed using primers Acr19 and Acr22, and the products were cloned and sequenced. Twenty-six independent clones were sequenced; 18 had the predicted wild-type sequence through the splice site junction at the final exon, and 8 had a 2-bp deletion relative to wild-type cDNAs, suggesting that these clones represent aberrant splice products that use an AG acceptor 2 nucleotides downstream of the normal acceptor. The wild-type clones may either represent RNA from the wild-type copy of *gld-1* on nDp4 and/or correct splicing of mutant oz116 RNA similar to that seen by Aroian et al. (1993).

**RT-PCR analysis of gld-1 RNA**

Approximately 5 μg of each RNA sample was reverse transcribed with 200 units of Superscript RT (GIBCO/BRL) at 42°C for 30 min with 15 μg of random hexamer primers (GIBCO/BRL). One-twentieth of each reaction was PCR amplified directly using primers AlgR20 ([3P end-labeled]) and AlRcR9. The amplification cycle profile was 94°C for 30 sec, 64°C for 30 sec, and 72°C for 45 sec, ending with 7 min at 72°C after the final cycle. Separate reactions were done at different numbers of cycles to determine the linear range of amplification for each sample. Amplification products were separated on denaturing polyacrylamide gels (5%, 7 m urea). Quantitation was performed using a PhosphorImager (Molecular Dynamics). To show that both spliced products would amplify at the same efficiency, experiments were first done on cDNA plasmid clones of known concentration (Fig. 8, data not shown). Because of the sensitivity of the assay, a no-DNA control was always included and all RNA samples were tested for contamination by performing the reaction on RNA that had not been reverse transcribed.

**Allele sequencing**

Five adult animals homozygous for each allele of *gld-1* were processed essentially as described in Williams et al. (1992). PCR was performed on these samples using one of three primer pairs to generate DNA for sequencing (AlgR9/AlgR23, AlgR21/AlgR32, or AlgR27/AlgR37). Of each product, 25–50 ng was sequenced using [3P-end-labeled] primers in a cycle sequencing reaction (Craxton 1991). The genomic DNA containing the *gld-1* coding region was sequenced completely for all alleles with the exception of Mog alleles oz30, oz33, q62, oz34, oz16, and oz70, most of which were only sequenced with AlgR21 and confirmed on the opposite strand with AlgR18. In addition, six alleles (q93, q126, q268, q93oz50, oz10, and oz35) were sequenced through the 810-bp 3' UTR. For q343, ~200 bp of sequence from each end of the 2.3-kb insertion was obtained using the primers AlgR28 and AlgR32. A TA target site duplication occurred at the site of insertion, as has been described previously for Tc2 (Ruvolo et al. 1992). All mutations were confirmed on the opposite DNA strand.
Oligonucleotide primers used

The following oligonucleotides were used for this work, with the corresponding genomic sequence (numbered according to the sequence in Fig. 3) in parentheses; descending numbers indicate that the primer is on the opposite strand: AJgf9 (7-27); AJgf10 (178-201); AJgf14 (521-540); AJgr14 (550-539); AJgf16 (678-695); AJgr17 (813-796); AJgf18 (939-958); AJgr20 (1089-1068); AJgf20 (1068-1088); AJgr21 (1182-1164); AJcr8 (1209-1193); AJgf21 (1206-1226); AJrt2 (1301-1316, 1362-1364); AJ RCCcr9 (1377-1362, 1316-1314); AJgr23 (1442-1424); AJgf24 (678-695); AJgrl7 (813-796); AJgf18 (939-958); AJgr20 (1089-1168); AJgf27 (1814-1831); AJgr27 (1831-1814); AJgf28 (1852-1871); AJgf33 (2375-2394); AJcrl9 (2478-2459); AJcr20 (2530-2516); AJgr37 (2798-2781}. Additional oligonucleotides used were MO2B2a GCTCAGACCGCAGTTTCAG, MO2B2b GAGCACGTGCC, F29D1lb CAGAATGAGCAAGACGATCAGTC, AJremaltf GACAAGTCAAAGTCATTC, AJRGl5 GCGTCC.

Note added in proof

The GSG domain and the CGA region of the GLD-1 protein are quite similar (58% identical, 82% similar over 192 amino acids) to part of the mouse quaking gene product (T.A. Ebersole, Q. Chen, M.J. Justice, and K. Artzt, pers. comm.). All eight of the amino acids in which gld-1 missense mutations are found are conserved between GLD-1 and the predicted quaking protein.

Acknowledgments

We are most grateful to Bob Clifford and Ross Francis for continuous help in all aspects of the construction of the manuscript. We are also appreciative of comments on the manuscript from Patty Kuwabara and Bethany Westlund. We thank Yueping Zhang for technical assistance. Protocols advice, vectors, reagents, and discussion were freely provided by Laura Wilson Berry, Jim McCarter, Andrey Shaw, Kim Wieties-Clary, Phil Green, Steve L’Herault, Bert Schrank, Pam Hoppe, LaDeana Hillier, Rick Wilcut, and Bob Waterston and are greatly appreciated. We especially thank John Lye for help with fluorescent sequencing and its encoded protein, twitchin, of Caenorhabditis elegans. Genetics 125: 29-39.

Barton M.K. and J. Kimble. 1990. fung-1, a regulatory gene required for specification of spermatogenesis in the germ line of Caenorhabditis elegans. Genetics 125: 29-39.

Benian G.M., S.W. L’Hernault, and M.E. Morris. 1993. Additional sequence complexity in the muscle gene, unc-22, and its encoded protein, twitchin, of Caenorhabditis elegans. Genetics 134: 1097-1104.

Burd, C.G. and G. Dreyfuss. 1994. RNA binding specificity of hnRNP A1: Significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. EMBO J. 13: 1197-1204.

Clifford, R., R. Francis, and T. Schedl. 1994. Somatic control of germ cell development in Caenorhabditis elegans. Semin. Dev. Biol. 5: 21-30.

Coulson, A., R. Waterston, J. Kiff, J. Sulston, and Y. Kohara. 1988. Genome linking with yeast artificial chromosomes. Nature 335: 184-186.

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

Delahodde, A., A.M. Becam, L. Perea, and C. Jacq. 1986. A yeast gene and regulation of its encoded protein, twitchin, of Caenorhabditis elegans. Genetics 125: 29-39.

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

Ellis, R.E. and J. Kimble. 1995. The fog-3 gene and regulation of cell fate in the germ line of Caenorhabditis elegans. Genetics 139: 561-577.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Aroian, R.V., A.D. Levy, M. Koga, Y. Ohshima, J.M. Kramer, and P.W. Sternberg. 1993. Splicing in Caenorhabditis elegans does not require an AG at the 3′ splice acceptor site. Mol. Cell. Biol. 13: 626-637.

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

Ausbel, F.M., R. Brent, R.E. Kingston, J.G. Seidman, J.A. Smith, and K. Struhl, eds. 1991. Current protocols in molecular biology. Greene/John Wiley, New York.

Barstead, R.J., L. Kleiman, and R.H. Waterston. 1991. Cloning, sequencing, and mapping of an alpha-actinin gene from the nematode Caenorhabditis elegans. Cell Motil. Cyto. 20: 69-78.

References

Ahringer J. and J. Kimble. 1991. Control of the sperm-oocyte switch in Caenorhabditis elegans hermaphrodites by the fem-3 3′ untranslated region. Nature 349: 346-348.
Fumagalli, S., N.F. Totty, J.J. Hsuan, and S.A. Courmeidge. Higgins, D.G., A.J. Bleasby, and R. Fuchs. 1992. CLUSTAL V: Hartwell, L.H. and M.B. Kastan. 1994. Cell cycle control and Gorlach, M., C.G. Burd, and G. Dreyfuss. 1994. The determi- Henikoff, S. and J.G. Henikoff. 1993. Performance evaluation of

Francis, G.R., E. Maine, and T. Scheld. 1995b. Analysis of the

Hodgkin J. 1986. Sex determination in the nematode

Liang, Y., T.L. Jetton, E.C. Zimmerman, H. Najafi, F.M. Matschinsky, and M.A. Magnuson. 1991. Effects of alternate RNA splicing on glucokinase isoform activities in the pancreatic islet, liver, and pituitary. J. Biol. Chem. 266: 6999–7007.

Mahone, M., E.E. Saffman, and P. Lasko. 1995. Localized Bicaudal-C RNA encodes a protein containing a KH-domain, the RNA-binding motif of FMR-1. EMBO J. [in press].

Mello C.C., J.M. Kramer, D. Stinchcomb, and V. Ambros. 1991. Efficient gene transfer in C. elegans: Extrachromosomal maintenance and integration of transforming sequences. EMBO J. 10: 3959–3970.

Meyer, B.J. and L.P. Casson. 1986. Caenorhabditis elegans compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. Cell 47: 871–881.

Richard, S., D. Yu, K. Blumer, D. Hausladen, M.W. Olszowy, P.A. Connelly, and A.S. Shaw. 1995. Association of p62, a multifunctional SH2- and SH3-domain-binding protein, with src family tyrosine kinases, grb2 and phospholipase Cy-1. Mol. Cell. Biol. 15: 186–197.

Ruvolo, V., J.E. Hill, and A. Levitt. 1992. The Tc2 transposon of Caenorhabditis elegans has the structure of a self-regulated element. DNA Cell Biol. 11: 111–122.

Sagata, N., M. Oskarsson, T. Copeland, J. Brumbaugh, and G.F. Vande Woude. 1988. Function of c-mos proto-oncogene product in mesotic maturation in Xenopus oocytes. Nature 335: 519–525.

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

Siomi, H., M.C. Siomi, R.L. Nussbaum, and G. Dreyfuss. 1993b. The protein product of the fragile X gene FMR1, has characteristics of an RNA-binding protein. Cell 74: 291–298.

Takimoto, M., T. Tomonaga, M. Matunis, M. Avigan, H. Kruitzsch, G. Dreyfuss, and D. Levens. 1993. Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-myc promoter, in vitro. J. Biol. Chem. 268: 18249–18258.

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

Toda, T., A. Iida, T. Miwa, Y. Nakamura, and T. Imai. 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.

Verkerk, A.J.M.H., M. Pieretti, J.S. Sutcliffe, Y.-H. Fu, D.P.A. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M.F. Victoria, F. Zhang, et al. 1991. Identification of a gene [FMRI-1] containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65: 905–914.

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

Weng, Z., S.M. Thomas, R.J. Rickles, J.A. Taylor, A.W. Brauer, C. Seidel-Dugan W.M. Michael, G. Dreyfuss, and J.S. Brugge. 1996. Identification of Src, Fyn, and Lyn SH3-binding proteins: Implications for a function of SH3 domains. Mol. Cell. Biol. 14: 4509–4521.

Williams, B.D., B. Schrank, C. Huynh, R. Showkenne, and R.H. Waterston. 1992. A genetic mapping system in Caenorhabditis elegans based on polymorphic sequence-tagged sites. Genetics 131: 609–624.

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

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

Yew, N., M. Strobel, and G.F. Vande Woude. 1993. Mos and the cell cycle: The molecular basis of the transformed phenotype. Curt. Opin. Genet. Dev. 3: 19–25.

Yoshem, J. and I. Greenwald. 1993. A gene for a low density lipoprotein receptor-related protein in the nematode Caenorhabditis elegans. Proc. Natl. Acad. Sci. 90: 4572–4576.
Mutations in gld-1, a female germ cell-specific tumor suppressor gene in Caenorhabditis elegans, affect a conserved domain also found in Src-associated protein Sam68.

A R Jones and T Schedl

*Genes Dev.* 1995, 9:
Access the most recent version at doi:10.1101/gad.9.12.1491