nature structural & molecular biology

RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1

Sean P Ryder[1], Leah A Frater[2], Dana L Abramovitz[1,3], Elizabeth B Goodwin[2] & James R Williamson[1]

The post-transcriptional regulation of gene expression underlies several critical developmental phenomena. In metazoa, gene products that are expressed, silenced and packaged during oogenesis govern early developmental processes prior to nascent transcription activation. Furthermore, tissue-specific alternative splicing of several transcription factors controls pattern formation and organ development. A highly conserved family of proteins containing a STAR/GSG RNA-binding domain is essential to both processes. Here, we identify the consensus STAR-binding element (SBE) required for specific mRNA recognition by GLD-1, a key regulator of *Caenorhabditis elegans* germline development. We have identified and verified new GLD-1 repression targets containing this sequence. The results suggest additional functions of GLD-1 in X-chromosome silencing and early embryogenesis. The SBE is present in Quaking and How mRNA targets, suggesting that STAR protein specificity is highly conserved. Similarities between the SBE and the branch-site signal indicate a possible competition mechanism for STAR/GSG regulation of splicing variants.

Post-transcriptional control is a conserved mechanism of gene regulation that is fundamental to the development of multicellular organisms[1–3]. Transcription and subsequent silencing of mRNA enable the preprogramming of developmental events during transcriptionally inactive cellular states. Post-transcriptional regulation has a major role in the formation of morphogen gradients required for pattern formation and asymmetric cell division. Modes of post-transcriptional regulation include translation inhibition, mRNA destabilization, alternative splicing and sequestration of RNA into discrete subcellular bodies.

The highly conserved STAR/GSG family of RNA-binding proteins has a central role in several processes that govern metazoan development (**Fig. 1a**). GLD-1 and Qk1 regulate the translation of their mRNA targets[4,5]. Sam68, How and Kep-1 couple signal transduction pathways to the formation and export of alternatively spliced mRNA[6–8]. SF-1 is the mammalian intronic branch-site RNA-binding protein (mBBP)[9].

STAR/GSG proteins are composed of a single KH domain flanked by two regions homologous to the murine quaking gene, Qua1 and Qua2 (ref. 10). Several experiments have demonstrated that STAR domain proteins form functional homodimers in cells via the Qua1 domain[11,12]. Genetic, biochemical and NMR analysis of the KH and Qua2 regions has established that this domain forms an extended RNA interaction surface[13,14].

The *C. elegans* germline developmental regulator GLD-1 is a STAR/GSG protein[15] (**Fig. 1b**). Mutant *gld-1* alleles have variable germline defects that include ectopic proliferation of mitotic stem cells (tumor or null phenotype) and the inability to develop mature oocytes (MOG) or spermatocytes (FOG)[16,17]. The pleiotropic nature of the mutants implicates GLD-1 in the regulation of each major transition in spatial and temporal development of the hermaphrodite gonad.

A well characterized role of GLD-1 is in the spermatogenesis-to-oogenesis switch[18]. GLD-1 binds to the 3′ untranslated region (UTR) of *tra-2* mRNA and recruits a complex that silences translation[4,19]. Repression of *tra-2*, concurrent with expression of *fem-3*, leads to the production of spermatocytes prior to the L4-adult junction. Subsequently, *gld-1* and *fem-3* expression is repressed by the assembly of a distinct protein complex, leading to the activation of *tra-2* and production of oocytes[20,21].

Jan and co-workers have identified the GLD-1 response element (GRE) in the 3′ UTR of *tra-2* and shown that it is composed of two 28-nucleotide repeats (termed TGE) separated by a 4-nucleotide spacer[4,22] (**Fig. 2a**). We present here the affinity, stoichiometry and sequence specificity of GLD-1's interaction with the TGE. The results define a consensus STAR protein–binding element (SBE) necessary for high-affinity interaction. This consensus is found in the 3′ UTR of every known GLD-1 target and is present in a series of additional mRNAs transcribed in the germline. We have verified GLD-1 interaction with a subset of these candidate targets. There is 100% conservation of the RNA-binding surface between GLD-1, Qk1 and How, suggesting that these STAR/GSG proteins recognize identical RNA sequences. Therefore, identification of the SBE will probably facilitate mapping of the repression pathways that govern the development of multiple species.

[1]Department of Molecular Biology, Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, Mail Stop MB-33, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [2]Department of Genetics, University of Wisconsin, Madison, 402 Genetics Building, 445 Henry Mall, Madison, Wisconsin 53706, USA. [3]Present address: Ingenuity Systems, 1565 Charleston Road, Mountain View, California 94043, USA. Correspondence should be addressed to J.R.W. (jrwill@scripps.edu).

Published online 29 December 2003; doi:10.1038/nsmb706

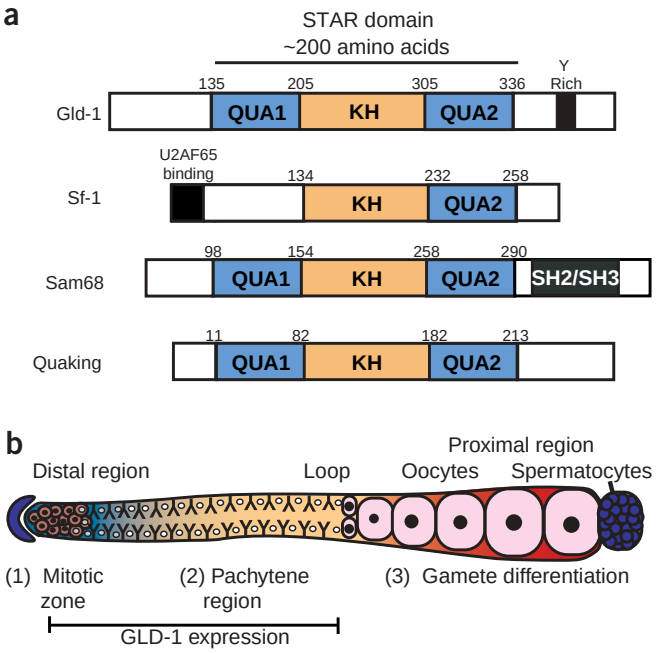


Figure 1 A member of the STAR family of RNA-binding proteins regulates germline development in *C. elegans*. (**a**) Domain structure of representative members of the STAR/GSG protein family. The Qua1, KH and Qua2 domains are shaded. Where known, other functional domains are annotated. (**b**) Cartoon model of a *C. elegans* hermaphrodite gonad. The limit of GLD-1 expression is denoted by a bar. The major regions of the germline are labeled.

RESULTS

GLD-1 binds to a single TGE repeat

The affinity and binding stoichiometry of GLD-1 for GRE has not been determined. To address this question, the equilibrium dissociation constant (K_d) of the interaction between purified recombinant GLD-1-STAR domain with GRE and TGE RNA was determined by quantitative gel mobility shift analysis. The affinity of GLD-1 for TGE RNA (K_d = 11.4 nM, n = 1.1; **Fig. 2b**) is comparable to that for the intact GRE (data not shown), demonstrating that a single TGE element is sufficient for high-affinity binding. In addition, only one shifted species is observed with TGE RNA whereas at least two separate shifted bands are observed with GRE RNA. These data suggest that the GRE contains more than one GLD-1 binding site. For this reason, we chose TGE RNA for more detailed analysis.

The stoichiometry of the interaction of GLD-1 with TGE was determined by quantitative gel mobility shift analysis under conditions in which the RNA concentration was not limiting (**Fig. 2c**). The data were compared to theoretical saturation curves for protein/RNA ratios of 1:1, 2:1 and 4:1 to determine the apparent stoichiometry. Two equivalents of GLD-1 are required to saturate the binding of TGE RNA.

Oligomerization analysis of GLD-1 STAR/GSG domain

Several lines of evidence suggest that STAR domain proteins form functional homodimers or homotrimers in cells[11,12]. We used analytical gel filtration chromatography to probe the oligomeric state of recombinant GLD-1-STAR. By comparing the retention time of GLD-1-STAR with that of five protein standards, we determined its apparent molecular mass (**Fig. 3a**) to be 135 kDa, approximately twice the calculated molecular mass of the protein (67.4 kDa). MALDI-TOF MS confirmed the molecular mass of GLD-1-STAR (67.432 kDa). Together, these data demonstrate that purified protein is dimeric in solution, in the absence of RNA or other cellular factors.

To determine the K_d of the GLD-1-STAR homodimer, we repeated the analytical gel filtration chromatography with varying protein concentrations (1–100 μM; 1 μM represents the limit of detection of this protein by UV absorption at 220 nm). At all concentrations tested, the protein migrated through the column as a dimer. These data set the upper limit for the K_d of dimerization to 1 μM. However, the absence of cooperativity in the GLD-1–TGE interaction (n = 1.1) implies that GLD-1 binds to the RNA as a preformed dimer. If so, then the dimerization K_d must be substantially tighter than <10 nM.

Analysis of STAR subdomain function

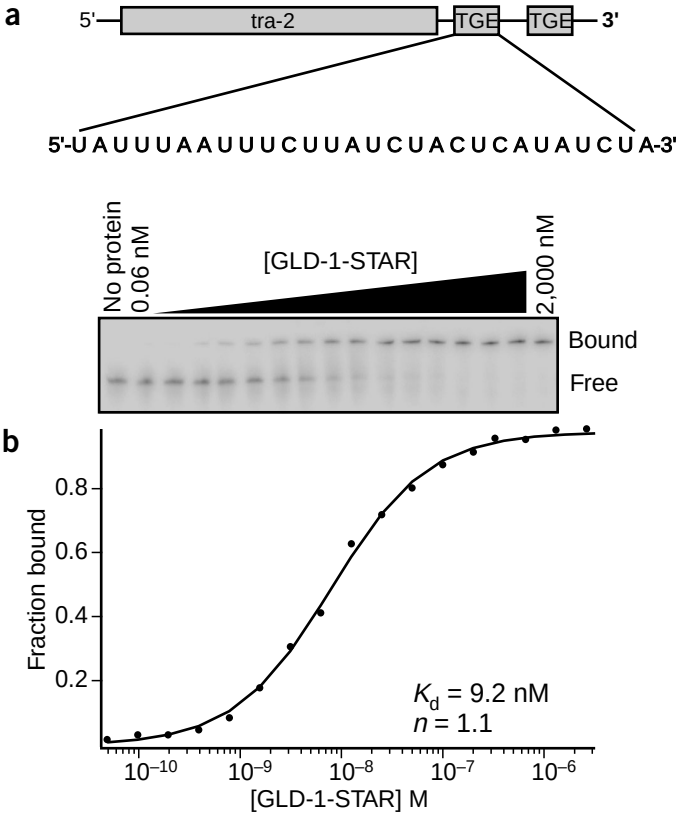
To probe the domain structure of GLD-1, truncation mutants lacking the Qua1, Qua2, or KH and Qua2 regions were prepared (**Fig. 3a**). The oligomeric state and the RNA-binding activity were measured for each variant by analytical gel filtration chromatography and electrophoretic gel mobility shift assay, respectively (**Fig. 3**). All mutants containing the Qua1 region migrate through the analytical gel filtration column as a dimer. In contrast, the Qua1 deletion mutant, GLD-1-KH-Q2, has an apparent molecular mass consistent with a monomer (**Fig. 3a**). Together, these experiments show that the Qua1 region of GLD-1 is both necessary and sufficient for dimerization. This is consistent with previous experiments that identified a point mutation in the Qua1 domain that disrupted oligomerization *in vivo*[12].

In contrast, the variants without the Qua2 region lack TGE-binding activity, indicating that this region is required for high-affinity binding (**Fig. 3b**). Surprisingly, deleting the Qua1 domain also destabilizes the TGE interaction by approximately one order of magnitude. These data show that the Qua1 domain contributes ~1.5 kcal mol^{-1} to the interaction. It is unclear whether this reduction in affinity is due to loss of direct contacts between Qua1 residues and the RNA or if it is due to indirect effects caused by destabilization of the dimer. Notably, the Hill coefficient of the GLD-1-KH-Q2 interaction with TGE RNA is 2.3 (**Fig. 3c**). The high cooperativity of the binding suggests that at least two copies of this monomeric GLD-1 variant bind to a single TGE repeat. The cooperativity may reflect weak residual dimerization mediated by the presence of RNA. This is consistent with weak dimerization (~1–3 mM) of a KH domain observed by NMR analysis[23].

A six-nucleotide element defines the specificity of GLD-1 binding

The gel electrophoretic mobility shift assay was used to probe the nucleotide sequence specificity of the interaction. We prepared 19 point mutations of the TGE by chemical synthesis. We determined the K_d for each mutant by direct titration of GLD-1-STAR (**Fig. 4a**). Of the 19 mutations, 8 showed a substantial increase in the K_d relative to wild-type TGE RNA. These eight mutations cluster into two regions, a dinucleotide (U5-A6) in the 5′ end of the TGE and a contiguous six-nucleotide sequence (U17-ACUC-A22) near the 3′ end of the sequence.

The hexanucleotide element is markedly similar to the specificity determinant for the branchpoint site recognition of introns mediated by SF-1 (UACUAA)[24,25]. The NMR structure of SF-1 KH-Q2–branch site RNA complex shows that SF-1 specifically recognizes five of the six nucleotides that form the branch-site sequence[14]. An alignment of GLD-1 with SF-1 reveals that 17 out of 25 RNA-binding residues identified in the SF-1 structure are identical in GLD-1. This correlation strongly suggests that GLD-1 interacts with this six-nucleotide element in a manner analogous to SF-1. However, the affinity of GLD-1 for TGE RNA (~10 nM) is much tighter than that of SF-1 for its RNA target



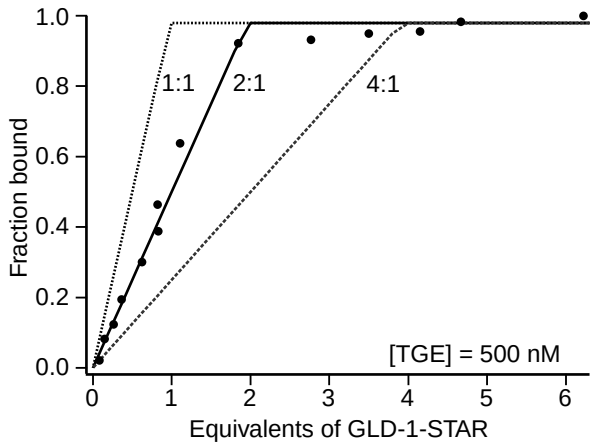


Figure 2 Affinity and stoichiometry of GLD-1-STAR for the TGE. (**a**) Sequence of the TGE repeat from the 3′ UTR of *tra-2* mRNA. Binding experiments were carried out with the GLD-1-STAR domain, composed of residues 135–336 fused to the C terminus of MBP, and TGE RNA. (**b**) Sample electrophoretic mobility shift experiment. The concentration of GLD-1-STAR increases by a factor of 2 from 0.06 nM to 2 μM across the titration, represented by a filled triangle. Bound and free TGE RNA are labeled. A plot of the fraction of bound TGE RNA as a function of protein concentration is shown, including a fit to the Hill equation. The K_d and Hill coefficient for this experiment are given. (**c**) Plot of the fraction of labeled TGE RNA bound vs. molar equivalent of GLD-1-STAR for a stoichiometry mobility shift experiment. Theoretical saturation curves for 1:1, 2:1 and 4:1 protein/RNA stoichiometry are shown.

(~1–6 μM). This discrepancy in affinity is possibly due to the sequence variation between GLD-1 and SF-1, but it is more likely due to additional RNA recognition attained by dimerization of GLD-1. As SF-1 monomers recognize a similar hexanucleotide element, we reasoned that one protomer of the GLD-1 dimer recognizes the U17-ACUC-A22 sequence whereas the other binds to the upstream U5-A6 dinucleotide.

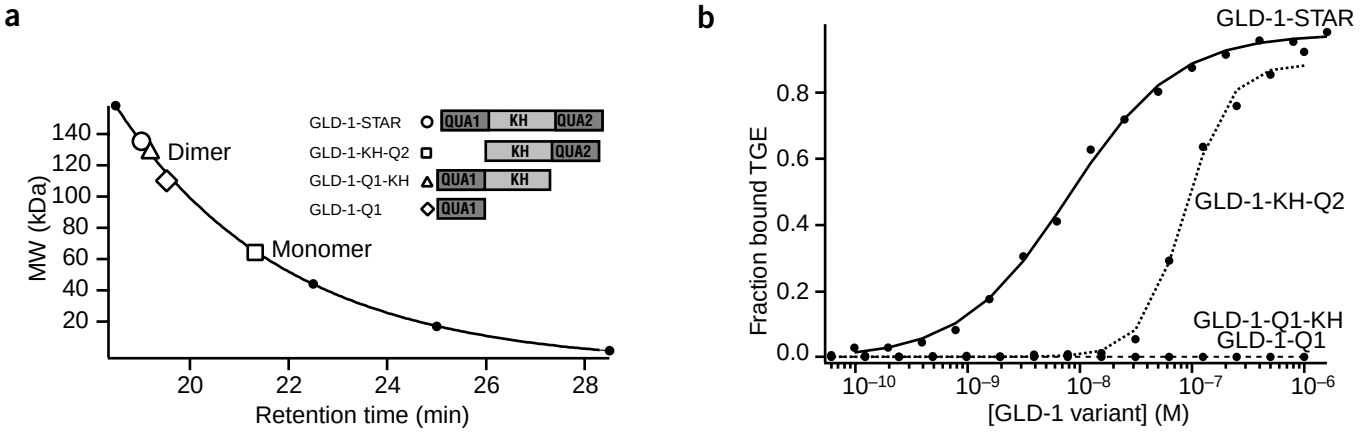
If both GLD-1 protomers are active and in equivalent conformations, then GLD-1 dimers should be able bind to a shorter RNA containing the hexanucleotide element with an apparent 1:1 protein/RNA stoichiometry. We tested this using end-mapping analysis, direct and competition titration experiments and isothermal titration calorimetry (ITC).

End mapping takes advantage of partial alkaline hydrolysis of the target RNA to delineate the smallest RNA fragment capable of binding to the protein (**Fig. 4b**). RNA fragments are equilibrated with a limiting concentration of protein, so that only fragments with the highest affinity are able to bind. Bound fragments are physically separated from unbound RNA, then visualized by denaturing PAGE. By repeating the experiment with 5′- and 3′-end labels, it is possible to delineate the minimal RNA element needed to bind to the protein. In this experiment, TGE RNA was used as a probe. Fragments of TGE RNA that bound to GLD-1-STAR were separated from unbound RNA by amylose resin chromatography. The hexanucleotide consensus element (UACUCA) is the smallest RNA fragment capable of binding to GLD-1.

To define the affinity and stoichiometry of GLD-1 for the minimal fragment, we prepared an RNA construct by chemical synthesis composed of residues A14–U25. This RNA contains the hexanucleotide element flanked by three nucleotides on either end. We observed no mobility shift under any condition tested by direct titration with GLD-1-STAR (data not shown). There are at least two possible explanations for the lack of shift. First, the shorter RNA may bind with substantially weaker affinity, although this hypothesis is not consistent with the end-mapping results. Second, the RNA may interact with the protein with dissociation kinetics unfavorable for gel mobility shift analysis.

Fortunately, the relative efficiency of protein binding to RNA can be determined by competition. In this approach, the well defined shift of GLD-1-STAR with TGE RNA is used to provide a signal, and the ability of unlabeled RNA constructs to compete GLD-1 binding to TGE RNA is monitored. The efficiency of protein binding to an unlabeled competitor RNA is determined by titrating increasing concentrations of competitor RNA (**Fig. 4c,d**) into a bound GLD-1–TGE complex. By fitting the fraction of TGE RNA bound as a function of competitor concentration, the half-maximal inhibitory concentration (IC_{50}) can be derived. By this approach, the IC_{50} of TGE RNA (self-competition) is 120 ± 33 nM, whereas the IC_{50} for the 12-mer RNA is 230 ± 55 nM (**Fig. 4c**). This represents a modest two-fold reduction in competition efficiency for the shorter RNA, suggesting that 12-mer RNA binds almost as well as the full-length TGE.

However, competition experiments do not address the stoichiometry of binding. To determine the stoichiometry and to directly measure the K_d for the 12-mer RNA, we titrated GLD-1-STAR into TGE or 12-mer RNA in an isothermal titration calorimeter (**Fig. 5**). This instrument measures the heat evolved upon interaction of a macromolecule with a ligand, which is used to determine the change in enthalpy, the K_d and the binding stoichiometry. The K_d of GLD-1-STAR for TGE RNA determined by ITC is ~50-fold weaker than that determined by gel shift assay (K_d = 540 nM; ΔH_{obs} = –13.1 ± 0.3 kcal mol^{-1}). However, the measured protein/RNA stoichiometry of binding is identical in both assays, 2:1. The discrepancy in the dissociation constant is not unusual for RNA-binding proteins, and can be partially explained by differences in the equilibration temperature, binding buffer and protein concentration required for the calorimetry experi-



c

	K_d (nM)	Hill	Oligomeric state
STAR	11.4 ± 2.3	1.1 ± 0.2	Dimer
KH-Q2	104 ± 24	2.3 ± 0.1	Monomer
Q1-KH	No shift	N/A	Dimer
Q1	No shift	N/A	Dimer

Figure 3 Functional analysis of STAR subdomains. (**a**) Subdomain structure of GLD-1 variant constructs. The circle, square, triangle and diamond represent GLD-1-STAR (residues 135–336),GLD-1-KH-Q2 (residues 205–336), GLD-1-Q1-KH (residues 135–305) and GLD-1-Q1 (residues 135–205), respectively. A standardization curve for a SEC-250 analytical gel filtration column is shown with Bio-Rad molecular mass standards (●). The retention time of each GLD-1 variant is denoted by its assigned symbol. (**b**) Plot of the fraction of TGE RNA bound as a function GLD-1 variant concentration. The plots and their fits are labeled. (**c**) Table of binding parameters and oligomerization state for each GLD-1 variant.

ments[26]. The K_d of the 12-mer–GLD-1-STAR interaction determined by ITC is 620 nM (ΔH_{obs} = –10.0 ± 0.2 kcal mol^{-1}). Notably, the protein/RNA stoichiometry of this interaction is 1:1. This confirms that both protomers of the GLD-1 dimer are capable of independently recognizing the hexanucleotide element.

Sequence requirements of the six-nucleotide element

To define a consensus for the sequence requirements in the hexanucleotide fragment, we repeated the competition gel shift experiment with a collection of mutant 12-mer RNA oligonucleotides (**Fig. 4d**,**e**). This set of mutants contains every possible point substitution at each position of the hexanucleotide element. We determined the relative binding affinities by comparing the IC_{50} of the mutant with that of wild-type 12-mer RNA. Only one mutation, C21A, competed as well as the wild-type 12-mer. Two mutants, U17G and C19A, have a competition efficiency within five-fold of that of the wild type. Three additional mutants are reduced by ten-fold (U17A, U17C and C21U). From this data, a conservative UACU(C/A)A and relaxed (U>G>C/A)A(C>A)U(C/A>U)A consensus for GLD-1 recognition can be derived.

GLD-1 interaction with TGE elements from other targets

In addition to *tra-2*, two other genes are targets of GLD-1 repression. Xu and co-workers have shown that the expression pattern of MES-3 expands in *gld-1* mutant worms and demonstrated a direct interaction between GLD-1 and the 3′ UTR of *mes-3* (ref. 27). MES-3 protein is expressed in the germline and forms a specific complex with MES-2 and MES-6 that is involved in germline X-chromosome silencing[28]. Based on sequence comparison between the *tra-2* TGE repeat and the 3′ UTR of *mes-3*, the authors identified three putative TGE-like sequences in *mes-3* RNA (**Fig. 4c**, *mes-3* TGE1-3).

Similarly, Lee and Schedl have demonstrated that GLD-1 represses translation of *rme-2* RNA[29]. RME-2 is a yolk receptor that is required for oocyte development[30]. The authors used biotin-RNA pulldowns to localize the binding of GLD-1 to two distinct regions of the mRNA. These regions consist of a 50-nucleotide stretch in the 5′-coding region and an 84-nucleotide sequence in the 3′ UTR. Based on sequence similarity in these two regions, the authors suggested a CU(A/U)UUUAUU consensus potentially involved in the binding interaction. We prepared RNA constructs composed of 28 nucleotides from each region that both align well with the TGE repeat from *tra-2* RNA and contain the suggested CU(A/U)UUUAUU consensus element (**Fig. 4c**, *rme-2* TGE1-2).

To address the specificity of the hypothetical target sequences, we prepared RNA for each TGE-like sequence and determined the affinity for GLD-1-STAR. Of the *mes-3* TGE-like sequences, only *mes-3* TGE1 binds with measurable affinity (K_d = 520 nM ± 196). This is substantially weaker than *tra-2* TGE (K_d = 11.4 nM ± 2.4). In contrast, *rme-2* TGE1 binds to GLD-1-STAR with high affinity (K_d = 5.1 nM ± 1.6), whereas *rme-2* TGE2 does not bind at all. These data lead to two distinct conclusions. First, the CU(A/U)UUUAUU element present in both *rme-2* TGEs is not sufficient to achieve high-affinity binding to GLD-1. Second, both *tra-2* TGE and *rme-2* TGE1 contain all of the sequence elements that lead to high-affinity binding, and therefore sequence conserved between these two elements is likely to represent the determinants of GLD-1 binding. Notably, *rme-2* TGE1 RNA contains a hexanucleotide element (UACUAA) that falls within the conservative consensus criteria derived by the mutagenesis experiments outlined above.

To identify high-affinity GLD-1 binding sites in the *mes-3* and *rme-2* 3′ UTRs, we searched for alternative TGE-like sequences that contain the consensus hexanucleotide element. We identified one sequence in each UTR, termed *mes-3* TGE4 and *rme-2* TGE3 (**Fig. 4**). RNA for these elements was prepared and the affinity for GLD-1-STAR determined. Though sequence outside of the hexanucleotide consensus is not conserved, both putative TGEs bind as well as *tra-2* TGE. These

data demonstrate that the hexanucleotide consensus is necessary for high-affinity binding.

Identification of novel GLD-1 targets

The identification of a consensus for GLD-1 recognition of RNA provides an opportunity to search for additional repression targets of this protein. Because GLD-1 recognizes a UACU(C/A)A hexanucleotide, we reasoned that targets of GLD-1 should contain this sequence in their mRNA, presumably in the 3′ UTR. Furthermore, as GLD-1 is expressed in the hermaphrodite germline, the target genes should be expressed in this tissue. As a preliminary search for novel targets, we scanned the 3′ UTRs of 53 genes known to be expressed in the

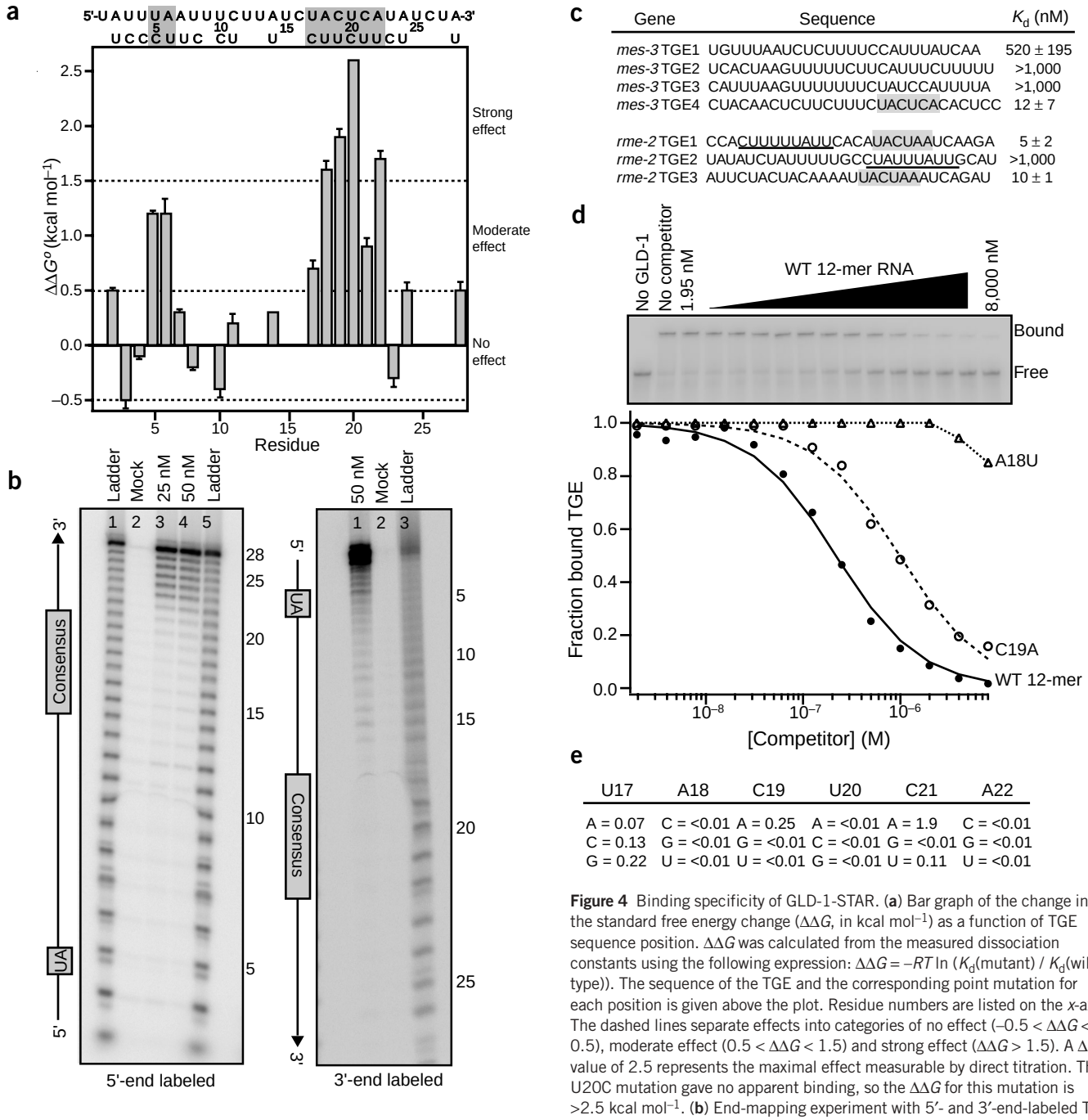


Figure 4 Binding specificity of GLD-1-STAR. (**a**) Bar graph of the change in the standard free energy change ($\Delta\Delta G$, in kcal mol^{-1}) as a function of TGE sequence position. $\Delta\Delta G$ was calculated from the measured dissociation constants using the following expression: $\Delta\Delta G = -RT \ln (K_d(\text{mutant}) / K_d(\text{wild type}))$. The sequence of the TGE and the corresponding point mutation for each position is given above the plot. Residue numbers are listed on the *x*-axis. The dashed lines separate effects into categories of no effect ($-0.5 < \Delta\Delta G < 0.5$), moderate effect ($0.5 < \Delta\Delta G < 1.5$) and strong effect ($\Delta\Delta G > 1.5$). A $\Delta\Delta G$ value of 2.5 represents the maximal effect measurable by direct titration. The U20C mutation gave no apparent binding, so the $\Delta\Delta G$ for this mutation is >2.5 kcal mol^{-1}. (**b**) End-mapping experiment with 5′- and 3′-end-labeled TGE RNA. The orientation of the sequence is shown to the left of each gel (15% (w/v) denaturing polyacrylamide). The position of the UA dinucleotide and the hexanucleotide consensus is denoted with a gray box. Lane 2 on each gel contains a no-GLD-1 control. (**c**) Proposed TGE-like sequences for *mes-3* and *rme-2* mRNA. The previously proposed CU(U/A)UUUAUU element is underlined. The hexanucleotide specificity determinant is boxed in gray. The K_d for each RNA is given. (**d**) Native gel of a typical competition experiment. Free and bound labeled TGE RNA are denoted to the right of the gel. The concentration of competitor RNA for each lane is given above the gel and decreases by a factor of two from a maximum of 8 μM. The concentration of GLD-1-STAR in each lane is 100 nM, with the exception of the lane labeled no GLD-1. Below is a plot of the normalized fraction of bound TGE as a function of competitor RNA concentration for three representative competitor 12-mer RNAs, along with the fit of each data set to the sigmoidal dose-response function. WT, wild type. (**e**) Relative IC_{50} value for each mutant 12-mer compared with the wild-type 12-mer sequence. The relative IC_{50} is listed as <0.01 if the apparent IC_{50} is >10 μM.

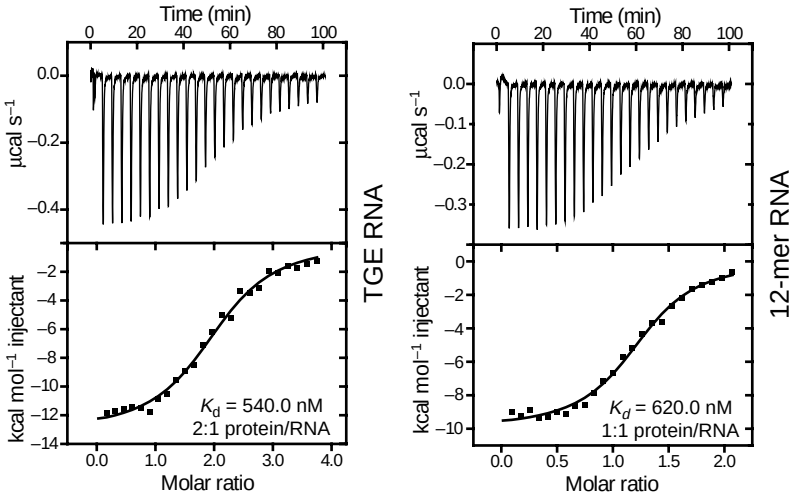


Figure 5 Determination and analysis of the minimal RNA fragment that binds to GLD-1-STAR. Results of an isothermal titration calorimetry experiment. Left, measurements for the TGE RNA; right, 12-mer RNA.

germline that have defined 3′-untranslated regions. Seven genes contain the conservative consensus hexanucleotide in their 3′ UTRs. (**Fig. 6a**, top). Three of these are the well-established targets of GLD-1: *tra-2*, *mes-3* and *rme-2*. The other genes are *mes-4*, *peb-1*, *gld-1* and *cdc-25.1*. Each of these genes are involved in developmental regulation[28,31,32]. Fifteen additional genes were identified that contain at least one copy of the relaxed consensus element (U>G>C/A)A(C>A)U(C/A>U)A (**Fig. 6a**, bottom). These include genes of diverse function, such as signal transduction (*apx-1*, *glp-1* and *wee-1.3*), RNA binding (*nos-3*, *pie-1* and *puf-8*) and transcription regulation (*nhr-23*, *tra-1* and *unc-62*).

Lee and Schedl identified several potential GLD-1 targets via a subtractive hybridization approach[29] . Seven of these genes have annotated untranslated regions (*cej-1*, *exo-3*, *lin-45*, *puf-5*, *puf-6*, *puf-7* and *rme-2*). As a positive control, we searched this set for the conservative binding hexanucleotide. All seven contain a conservative consensus-binding site (UACU(C/A)A). The strong correlation between the subtractive hybridization results and the presence of a perfect binding site suggests that the hexanucleotide element is mediating GLD-1's interaction with these mRNAs and bolsters the model that these genes are targets of GLD-1.

To test whether GLD-1 recognizes the consensus sequence *in vivo*, we carried out coimmunoprecipitation experiments on a subset of predicted GLD-1 targets. Specifically, we immunoprecipitated GLD-1 and used RT-PCR to assay for the presence of specific mRNAs in the precipitate. Three candidate targets were tested: *mes-4*, *pie-1* and *tra-1*. All three targets coimmunoprecipitate with GLD-1 (**Fig. 6b**). This interaction is GLD-1 dependent as none of these targets are coimmunoprecipitated by a mutant GLD-1 protein that contains a missense mutation in the KH domain that disrupts RNA binding.

DISCUSSION

Homodimers of GLD-1 bind to a single TGE repeat

Several lines of evidence demonstrate that STAR/GSG domain proteins such as GLD-1 are homodimeric proteins[11,12]. Here we have shown that recombinant, purified GLD-1-STAR is dimeric in the absence of cellular factors and RNA. Deletion of the Qua1 domain (residues 135–205) eliminated dimerization activity. Conversely, expression of the Qua1 domain as a C-terminal fusion to maltose binding protein produced a dimeric protein. Together, these data show that the Qua1 domain of GLD-1 is both necessary and sufficient to confer dimerization activity.

The interaction of GLD-1-STAR with a single TGE repeat was characterized, demonstrating that GLD-1 binds to the TGE with high affinity (~10 nM) with an apparent 2:1 protein/RNA stoichiometry. There is no cooperativity in the interaction. Together, these data suggest that GLD-1 homodimers bind to TGE RNA as a preformed unit (**Fig. 7a**). Two additional observations are consistent with this model. First, GLD-1-STAR forms dimers in the absence of RNA. Second, monomeric GLD-1-KH-Q2 binds to GLD-1 with high cooperativity, suggesting that two copies of this GLD-1 variant interact but only in an RNA-dependent manner.

We have demonstrated that a hexanucleotide consensus element is required for high-affinity GLD-1 binding. Closer inspection of the two TGE regions of *tra-2* mRNA identifies three consensus binding sites. The *in vitro* binding experiments presented here show that GLD-1 dimers can bind independently to each TGE. However, both protomers of the GLD-1 dimer are capable of individually recognizing the hexanucleotide element. Therefore, a single GLD-1 dimer may recognize any two of the three consensus sites in the UTR leading to a looped-out RNA structure. The presence of additional factors may limit the conformation of *tra-2* mRNA within the GLD-1 dimer.

GLD-1 target specificity

The determination of a hexanucleotide specificity determinant facilitates identification of novel GLD-1 binding targets by database search. Though the presence of the consensus is not sufficient to unambiguously define a repression target, we show here that it is required to mediate GLD-1 binding. As an initial screen for candidate targets, we searched the untranslated regions of 61 genes for GLD-1 consensus binding sites. Fourteen contain a perfect consensus in their 5′ or 3′ UTR, including the well established targets of GLD-1: *tra-2*, *mes-3* and *rme-2* (refs. 4,27,29). Five are novel putative targets (*cdc-25.1*, *gld-1*, *mes-4*, *peb-1* and *unc-37*). Furthermore, 18 additional candidates were identified that contain a relaxed consensus binding site.

GLD-1 interacts with *mes-4* mRNA *in vivo* by coimmunoprecipitation. Four MES genes (maternal effect sterile: *mes-2*, *mes-3*, *mes-4* and *mes-6*) are required to silence X-chromosome expression in the germline[28,33]. MES-2, MES-3 and MES-6 form a multiprotein repression complex that colocalizes with X chromosomes. MES-4 specifically recognizes autosomes and occludes the MES-2–MES-3–MES-6 repression complex. GLD-1 represses translation of *mes-3* mRNA[27]. The data presented here suggest that *mes-4* is also a target of GLD-1, indicating that it is a key regulator of chromatin state in the germline.

Similarly, the results show that GLD-1 interacts with *pie-1* mRNA. PIE-1 is a CCCH zinc finger protein that specifies germline blastomere identity in early embryo development[34,35]. PIE-1 is preferentially localized to germline daughter cells during early division. Blastomere-specific repression of PIE-1 production by maternal GLD-1 may lead to cell-specific localization of PIE-1 (ref. 34). However, Seydoux and co-workers have clearly shown that localization of PIE-1 is achieved primarily at the protein level[36]. If *pie-1* expression is repressed by GLD-1, then GLD-1-mediated repression of nascent *pie-1* translation may be necessary in conjunction with PIE-1 degradation to efficiently

a

	Gene	Function	Sites
Conservative consensus UACU(C/A)A	*cdc-25.1*	Phosphatase	1
	cej-1	Cell junction protein	1
	gld-1	Translation regulator	1
	lin-45	raf homolog	1
	mes-3	X-chromosome silencing	2
	mes-4	Autosome binding	1
	peb-1	Developmental regulation	2
	puf-5	RNA-binding protein	1
	puf-6	RNA-binding protein	1
	puf-7	RNA-binding protein	1
	puf-8	RNA-binding protein	1
	rme-2	Yolk receptor	1
	tra-2	Sex determination	4
Relaxed consensus (U>A/C/G)A(C>A)U(C/A>U)A	*apx-1*	glp-1/lin-12/notch ligand	3
	cdc-25.1	Phosphatase	1
	ced-7	ABC tramsporter protein	1
	cej-1	Cell junction protein	3
	cgh-1	RNA helicase	1
	cpb-1	Poly(A) element binding	1
	daz-1	RNA-binding protein	1
	emo-1	sec-61 homolog	2
	fbf-1/2	Translation regulator	1, 2
	gld-1	Translation regulator	3
	glp-1	Notch homolog	1
	lad-1	Fibronectin family	3
	lin-36	Unknown	1
	lin-45	raf homolog	1
	mes-3	X-chromosome silencing	2
	mes-4	Autosome binding	2
	nhr-23	Nuclear hormone receptor	3
	nos-3	nanos homolog	1
	peb-1	Enhancer-binding protein	2
	pie-1	Germline regulatory protein	1
	puf-5	RNA-binding protein	2
	puf-6	RNA-binding protein	1
	puf-7	RNA-binding protein	1
	puf-8	RNA-binding protein	2
	rme-2	Yolk receptor	1
	tra-1	Sex determination	3
	unc-37	groucho homolog	1
	unc-62	hox transcription factor	2
	wee-1.3	Serine/threonine kinase	1

b

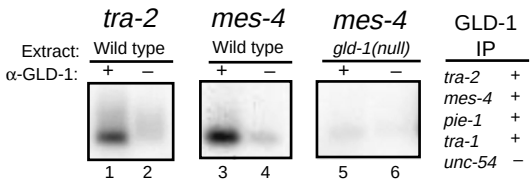


Figure 6 Novel targets of GLD-1. (**a**) Table of germline-expressed genes with perfect or relaxed consensus binding sites in their 5′ or 3′ UTR. Left frame contains genes with a conservative consensus binding site, whereas the bottom frame contains genes with a relaxed binding site. (**b**) GLD-1 binds several predicted targets *in vivo*. Protein A Sephacryl beads were incubated with extract either in the presence (lanes 1, 3 and 5) or absence (lanes 2, 4 and 6) of antibodies to GLD-1. After the coimmunoprecipitation procedure, RNA was isolated from the immunoprecipitated material and amplified using RT-PCR. Amplified product was detected in the presence of GLD-1 antibody for (lane 1) *tra-2* and (lane 3) *mes-4* mRNA but not in the absence of the antibody (lanes 2 and 4, respectively). No amplified product was detected for *mes-4* (lanes 5 and 6) when the extract was made from animals homozygous for a mutation in the GLD-1 KH domain that disrupts binding. Right, summary of coimmunoprecipitation experiments: *tra-2*, *mes-4*, *pie-1* and *tra-1* coimmunoprecipitated with GLD-1 whereas *unc-54* did not.

localize this protein to the germline daughter cell after division. Alternatively, GLD-1 may repress *pie-1* expression in the adult hermaphrodite germline.

Marin and Evans recently demonstrated that GLD-1 translationally represses *glp-1* expression in early embryogenesis[37]. Truncation analysis of the *glp-1* 3′ UTR in a reporter gene construct allowed the authors to identify a 32-nucleotide element required for repression. Subsequent mutation analysis of this region defined a sequence necessary for GLD-1 binding and repression activity. This element has the sequence GACUCAU and includes a relaxed consensus site (underlined). This element is the only potential GLD-1-binding site within the 3′ UTR of *glp-1* mRNA that matches the consensus criteria. This strengthens the conclusion that GLD-1 binding is mediated by the hexanucleotide consensus. In addition, the fact that a single relaxed consensus in the *glp-1* 3′ UTR is sufficient to yield translation repression suggests that many additional genes may be targeted by GLD-1.

Implications for the specificity of other STAR/GSG proteins

We have defined here the nucleotide sequence specificity of the STAR/GSG protein GLD-1. The hexanucleotide element that mediates GLD-1 binding to tra-2 mRNA (UACUCA) is almost identical to the specificity determinant of branchpoint site RNA (UACUAA) recognized by SF-1 (ref. 9). An alignment of GLD-1 and SF-1 reveals that the RNA-binding residues apparent from the SF-1 structure are 70% identical between these proteins[14] (**Fig. 7b**). Notably, mutations in the conserved RNA-binding residues are the most severe molecular lesions identified for several *gld-1* mutant alleles[15].

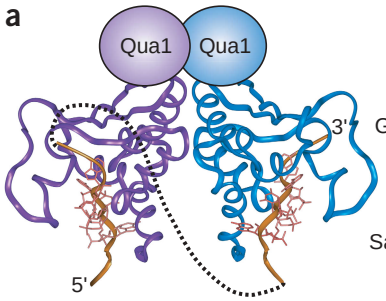
To address the binding specificity of other STAR/GSG proteins, the homology analysis was extended to include *Mus musculus* Qk1, *Drosophila melanogaster* How and *Homo sapiens* Sam68. Qk1 and How regulate the translation of at least one mRNA[6,38]. Qk1 regulates the alternative splicing pattern of myelin basic protein mRNA and is required for proper neural development[38–40]. Similarly, How regulates the expression of stripe mRNA, which is critical for muscle and tendon development in wing formation[6,41,42]. Sam68 has been implicated in the regulation of alternative splicing in response to ERK-mediated signaling and may have a role in the export of HIV RNA as a functional homolog of Rev[7,43].

GLD-1, How and Qk1 have 100% sequence identity in their RNA-binding residues. This strongly suggests that How and Qk1 bind to RNA with the same sequence specificity as GLD-1. This hypothesis is reinforced by the observation that an isoform of Qk1, specifically QKI-6, can functionally substitute for GLD-1 in the repression of a reporter gene fused to the 3′ UTR of *tra-2* in *C. elegans*[5].

In contrast, Sam68 is only 53% conserved in the RNA-binding site. Several of the differences are significant, including two charged amino acid substitutions. This suggests that Sam68 recognizes RNA with different sequence specificity. This is consistent with *in vitro* selection experiments showing that Sam68 binds with the highest affinity to RNA molecules that contain an AUUAAAA sequence, which is quite different from the GLD-1 binding consensus[44].

If Qk1 and How recognize RNA in a manner analogous to SF-1 and GLD-1, then consensus binding sites should be present in their target mRNAs (myelin basic protein and stripe mRNA, respectively). A search of the mRNA sequences of each of these genes identified two perfect and five relaxed consensus binding sites in the 3′ UTR of stripe mRNA and four relaxed consensus binding sites in the 3′ UTR of myelin basic protein mRNA. The presence of GLD-1-binding sites in the 3′ UTR of these RNAs is consistent with the model that a common sequence determinant is recognized by a broad subset of STAR family proteins. Therefore, the GLD-1 consensus defines a more general SBE.

The identity of the SBE consensus suggests a simple model for the mechanism of STAR/GSG-mediated regulation of alternative splicing. SF-1 is required in conjunction with U2AF65 to recognize and select the branchpoint signal site in mammalian introns, usually formed by a YNCURAY consensus[25]. The SBE contains a subset of possible branchpoint signals, suggesting that STAR/GSG proteins may regulate



b

	A		C	U	C/A		A
SF1	A248	R255	L244 L247 N151	G154 L155 G158 P159 R160	E149	V153 K184	I157 L164 I175 V183
GLD-1	A321	R328	L317 L320 N220	G223 R224 G227 P228 R229	D218	V222 R251	L226 A233 I245 M252
Qk1	A198	R205	L194 L197 N97	G101 R102 G105 P106 R107	D95	V99 R130	L103 A110 I122 M130
How	A243	R250	L239 L242 N152	G155 R156 G159 P160 R161	D150	V154 R185	L158 A165 I177 M185
Sam68	S272	E279	F268 L271 N171	G174 K175 G178 P179 Q180	K169	V173 R204	L177 A184 I196 M204

Figure 7 Homology analysis of GLD-1-STAR structure. (**a**) Model of the association of GLD-1 dimers with *tra-2* mRNA. The KH and Qua2 domains are represented as ribbons in purple and blue and are based on the SF-1 NMR structure[14]. The orientation of the dimer interface was modeled based on crystal contacts and NMR data observed for the Nova KH domains[23]. Orange ribbons represent RNA, with the bases for the hexanucleotide element in red. The Qua1 domain is a sphere. The connectivity of RNA between protomers is represented by a dashed line. The 5′ and 3′ ends are labeled. (**b**) Alignment of RNA-binding residues based on homology modeling to the structure of SF-1 bound to RNA. The first uridine of the consensus is not recognized in the SF-1 structure, therefore it is not possible to predict which residues interact with this base in GLD-1 or other STAR/GSG proteins. Proteins with 100% sequence conservation in their RNA-binding residues are highlighted in gray. Charge inversions in Sam68 are highlighted by colored rectangles. Positively charged amino acids are blue; negatively charged amino acids are red.

alternate splicing simply by competing with SF-1 for specific branch-site signals in the nucleus.

In summary, the results presented here define the specificity determinant required for GLD-1 binding. The primary determinant is a hexanucleotide element remarkably similar to branchpoint site RNA recognized by SF-1. Identification of this element facilitated a search for novel targets of GLD-1. Several candidate targets were selected and their interaction with GLD-1 verified by coimmunoprecipitation. The novel targets include *mes-4* and *pie-1*, suggesting that GLD-1 has a key regulatory role in X-chromosome dosage in the hermaphrodite germline and may function in the regulation blastomere identity in the early embryo. In addition, homology modeling of RNA-binding residues in SF-1, GLD-1, Qk1 and How suggests that these proteins all recognize the same hexanucleotide element (SBE). Because this signal overlaps with a subset of potential branchpoint site RNAs, we propose an SF-1 competition mechanism for STAR/GSG protein regulation of alternative splicing.

METHODS

RNA preparation and purification. The majority of RNA constructs used in this report, including TGE RNA and variants, were prepared by chemical synthesis (Dharmacon). RNA was deprotected, lyophilized and stored according to the manufacturer's protocol. GRE RNA (GGG-TGE-CUCA-TGE) was prepared by runoff *in vitro* transcription. 5′-end-labeled RNA was prepared by incubating 20 pmol of purified RNA with 25 pmol of [γ-^{32}P]ATP and 10 U of T4 polynucleotide kinase (PNK, NEB) for 1 h at 37 °C. 3′-end-labeled RNA was prepared by incubating 20 pmol of purified RNA with 25 pmol of 5′-[^{32}P]pCp and 20 U of T4 RNA ligase for 20 h at 15 °C.

Protein constructs and purification. All GLD-1 variants were cloned into pHMTc, a derivative of pMal-c2x (NEB) that includes an N-terminal His_6-tag and replaces the Factor Xa site with a TEV protease site. GLD-1 variants were prepared by PCR-amplifying the appropriate fragment from pGEX-14N and cloning the product in frame into the *Bam*H1 and *Hin*dIII sites of the pHMTc polylinker[4].

Each protein variant was expressed and purified as a MBP fusion. Plasmid DNA was transformed into BL-21 Gold *Escherichia coli* cells (Invitrogen), grown in liquid culture to an A_{600} of 0.6 and induced by the addition of 1 mM IPTG for 4 h. Cell pellets were lysed by sonication, clarified by centrifugation and purified over an amylose column (NEB). The fractions containing fusion protein were further purified by ion-exchange chromatography with a Bio-CAD SPRINT (Perceptive Biosystems) using POROS-HQ followed by POROS-HS resin. The resultant protein was >98% pure as determined by Coomassie blue–stained SDS-PAGE and MALDI-TOF mass spectrometry.

Electrophoretic mobility shift assays. The complex between GLD-1 variants and RNA was visualized by electrophoretic gel mobility shift assay. A constant concentration of radiolabeled RNA (100 pM) was equilibrated with varying concentrations of protein in equilibration buffer (EB: 10 mM Tris, pH 8.0, 25 mM NaCl, 0.1 mM EDTA, 0.1 mg ml^{-1} tRNA, 5 μg ml^{-1} heparin and 0.01% (w/v) IGEPAL CA630) for at least 3 h. Before loading, 4 μl of loading dye (30% (v/v) glycerol, 0.05% (w/v) xylene cyanol) was added to each sample. A 5 μl portion of each reaction was loaded onto a prerun native polyacrylamide gel (6% (w/v) 29:1 acrylamide/Bis-acrylamide, 0.5× TBE). Gels were resolved at 600 V for 30 min before they were dried and exposed to a PhosphorImager screen overnight (Molecular Dynamics). The fraction of bound RNA was determined and fit as described[26].

Stoichiometry binding experiments between GLD-1-STAR and TGE RNA were carried out as described for the titrations above, except that 500 nM unlabeled RNA was included in each equilibration. Data were compared to theoretical fractional saturation as described[45].

Competition experiments were carried out using the same assay, except that a constant concentration (100 nM) of GLD-1-STAR was included in each equilibration. Unlabeled competitor RNA was titrated into each reaction at varying concentrations and allowed to re-equilibrate for 3 h. The competitor concentration required to displace half of the bound TGE RNA (IC_{50}) was determined by plotting the fraction of bound radiolabeled TGE RNA as a function of competitor RNA concentration and fitting to a sigmoidal dose-response function in IGOR (Wavemetrics).

Analytical gel filtration chromatography. The apparent molecular mass of each MBP-GLD-1 variant was determined by analytical gel filtration chromatography using a Bio-Sil SEC-250 column (300 mm × 7.8 mm, Bio-Rad) attached to a Bio-Sil 250 guard (80 mm × 7.8 mm, Bio-Rad). Before each use, the column was standardized with Bio-Rad gel filtration standard to define the void volume and to delineate a calibration curve of molecular mass versus retention time. The column was equilibrated in filtration buffer (50 mM Tris, pH 8.0, 200 mM NaCl) for 2 h at a constant flow rate of 0.5 ml min^{-1} using a Beckman HPLC. Gel filtration standard was prepared as described by Bio-Rad. Retention time for each sample was obtained by analyzing an absorption trace for each sample at 280 nm. The retention time for each GLD-1 variant was compared to the standard curve to estimate the apparent molecular mass.

End-mapping analysis. Partially hydrolyzed TGE RNA fragments were prepared by incubating 20 pmol of 5′- or 3′-end-labeled RNA in 5 mM $NaCO_3$, pH 9.0 buffer for 20 min at 90 °C. Equilibration reactions (200 μl) were performed with 20 nM hydrolyzed RNA and 0, 25 or 50 nM GLD-1-STAR in equilibration buffer for 3 h. Bound fragments were isolated using amylose resin–packed spin columns (resin from NEB, empty columns from Amersham). To prevent nonspecific binding, the resin was preblocked with 10 mg ml^{-1} BSA in EB.

Equilibrated reactions were applied to the column, washed extensively with EB and eluted with 10 mM maltose in EB. The resultant pool was separated by 15% (w/v) denaturing PAGE and visualized with a PhosphorImager.

Isothermal titration calorimetry. GLD-1-STAR (110 μM) was titrated into either 4.9 μM 12-mer RNA or 10.5 μM TGE RNA using a MicroCal MCS isothermal titration calorimeter at 30 °C. Before injection, protein and RNA were extensively dialyzed into buffer containing 20 mM Tris, pH 8.0, 25 mM NaCl and 1 mM DTT. Otherwise, the experimental setup and data analysis was carried out as described[26].

GLD-1–RNA coimmunoprecipitations. Crude extract was prepared from a population of worms by lysing the animals using a French press at 1,600 psi in PBS with 1 μl ml^{-1} RNAguard (Amersham) and protease inhibitors (Roche). *gld-1*(*q361*-null) animals were handpicked from a population of *gld-1*(*q361*-null)/*hT2* animals. hT2 is a balancer that suppresses recombination over the left arm of LG I (linkage group 1). The extracts were spun at 1,000 rpm (82*g*) before being stored at –80 °C and again at 13,000 rpm (14,000*g*) immediately before use. Trisacryl immobilized protein-A beads (10 μl, Pierce) equilibrated in PBS were incubated overnight at 4 °C with 175 μl of worm extract at 2.46 μg μl^{-1} and 15 μl GLD-1 antibody. Control beads were incubated with extract in the absence of GLD-1 antibody. After incubation, beads were washed 3× with 500 μl PBS. After immunoprecipitation, RNA was isolated using Trizol (Invitrogen) and RT-PCR was carried out using oligo-dT primers and either Superscript II RT or Superscript III RT (Invitrogen). *tra-2*, *tra-1*, *mes-4* or *pie-1* cDNA was amplified from the reverse transcription reaction using gene-specific primers.

ACKNOWLEDGMENTS

We thank K. Lehmann, M. Recht and M. Trevathan for helpful discussions and critical comments concerning this manuscript, A. Manuell and Y. Pavlova for assistance with construct preparation and S. Bergqvist for assistance with ITC experiments. We also thank the *C. elegans* Genetic Center for worm strains. S.P.R. was supported by a Damon Runyon Fellowship Award. This work was funded by The Skaggs Institute for Chemical Biology and by US National Institutes of Health grants to J.R.W. and E.B.G.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 11 September; accepted 22 October 2003
Published online at http://www.nature.com/natstructmolbiol/

1. Curtis, D., Lehmann, R. & Zamore, P.D. Translational regulation in development. *Cell* **81**, 171–178 (1995).
2. Johnstone, O. & Lasko, P. Translational regulation and RNA localization in *Drosophila* oocytes and embryos. *Annu. Rev. Genet.* **35**, 365–406 (2001).
3. Wickens, M., Goodwin, E.B., Kimble, J., Strickland, S. & Hentze, M. Translational control in developmental decisions. in *Translational Control of Gene Expression* (eds. Sonenberg, N., Hershey, J.W.B. & Matthews, M.B.) 295–370 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 2000).
4. Jan, E., Motzny, C.K., Graves, L.E. & Goodwin, E.B. The STAR protein, GLD-1, is a translational regulator of sexual identity in *Caenorhabditis elegans*. *EMBO J.* **18**, 258–269 (1999).
5. Saccomanno, L. *et al.* The STAR protein QKI-6 is a translational repressor. *Proc. Natl. Acad. Sci. USA* **96**, 12605–12610 (1999).
6. Nabel-Rosen, H., Dorevitch, N., Reuveny, A. & Volk, T. The balance between two isoforms of the *Drosophila* RNA-binding protein How controls tendon cell differentiation. *Mol. Cell* **4**, 573–584 (1999).
7. Matter, N., Herrlich, P. & Konig, H. Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature* **420**, 691–695 (2002).
8. Di Fruscio, M. *et al.* Kep1 interacts genetically with dredd/caspase-8, and kep1 mutants alter the balance of dredd isoforms. *Proc. Natl. Acad. Sci. USA* **100**, 1814–1819 (2003).
9. Berglund, J.A., Chua, K., Abovich, N., Reed, R. & Rosbash, M. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**, 781–787 (1997).
10. Vernet, C. & Artzt, K. STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet.* **13**, 479–484 (1997).
11. Chen, T., Damaj, B.B., Herrera, C., Lasko, P. & Richard, S. 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 (1997).
12. Chen, T. & Richard, S. Structure-function analysis of Qk1: a lethal point mutation in mouse quaking prevents homodimerization. *Mol. Cell Biol.* **18**, 4863–4871 (1998).
13. Berglund, J.A., Fleming, M.L. & Rosbash, M. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA* **4**, 998–1006 (1998).
14. Liu, Z. *et al.* Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098–1102 (2001).
15. Jones, A.R., Francis, R. & Schedl, T. 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 (1996).
16. Francis, R., Maine, E. & Schedl, T. Analysis of the multiple roles of gld-1 in germline development: interactions with the sex determination cascade and the glp-1 signaling pathway. *Genetics* **139**, 607–630 (1995).
17. Francis, R., Barton, M.K., Kimble, J. & Schedl, T. *gld-1*, a tumor suppressor gene required for oocyte development in *Caenorhabditis elegans*. *Genetics* **139**, 579–606 (1995).
18. Kuwabara, P.E. & Perry, M.D. It ain't over till it's ova: germline sex determination in *C. elegans*. *Bioessays* **23**, 596–604 (2001).
19. Clifford, R. *et al.* FOG-2, a novel F-box containing protein, associates with the GLD-1 RNA-binding protein and directs male sex determination in the *C. elegans* hermaphrodite germline. *Development* **127**, 5265–5276 (2000).
20. Crittenden, S.L. *et al.* A conserved RNA-binding protein controls germline stem cells in *Caenorhabditis elegans*. *Nature* **417**, 660–663 (2002).
21. Zhang, B. *et al.* A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* **390**, 477–484 (1997).
22. Jan, E., Yoon, J.W., Walterhouse, D., Iannaccone, P. & Goodwin, E.B. Conservation of the *C. elegans* tra-2 3′UTR translational control. *EMBO J.* **16**, 6301–6313 (1997).
23. Ramos, A. *et al.* Role of dimerization in KH/RNA complexes: the example of Nova KH3. *Biochemistry* **41**, 4193–4201 (2002).
24. Berglund, J.A., Abovich, N. & Rosbash, M. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* **12**, 858–867 (1998).
25. Peled-Zehavi, H., Berglund, J.A., Rosbash, M. & Frankel, A.D. Recognition of RNA branch point sequences by the KH domain of splicing factor 1 (mammalian branch point binding protein) in a splicing factor complex. *Mol. Cell Biol.* **21**, 5232–5241 (2001).
26. Recht, M.I. & Williamson, J.R. Central domain assembly: thermodynamics and kinetics of S6 and S18 binding to an S15-RNA complex. *J. Mol. Biol.* **313**, 35–48 (2001).
27. Xu, L., Paulsen, J., Yoo, Y., Goodwin, E.B. & Strome, S. *Caenorhabditis elegans* MES-3 is a target of GLD-1 and functions epigenetically in germline development. *Genetics* **159**, 1007–1017 (2001).
28. Fong, Y., Bender, L., Wang, W. & Strome, S. Regulation of the different chromatin states of autosomes and X chromosomes in the germline of *C. elegans*. *Science* **296**, 2235–2238 (2002).
29. Lee, M.H. & Schedl, T. Identification of *in vivo* mRNA targets of GLD-1, a maxi-KH motif containing protein required for *C. elegans* germ cell development. *Genes Dev.* **15**, 2408–2420 (2001).
30. Grant, B. & Hirsh, D. Receptor-mediated endocytosis in the *Caenorhabditis elegans* oocyte. *Mol. Biol. Cell* **10**, 4311–26 (1999).
31. Thatcher, J.D., Fernandez, A.P., Beaster-Jones, L., Haun, C. & Okkema, P.G. The *Caenorhabditis elegans* peb-1 gene encodes a novel DNA-binding protein involved in morphogenesis of the pharynx, vulva, and hindgut. *Dev. Biol.* **229**, 480–493 (2001).
32. Ashcroft, N. & Golden, A. CDC-25.1 regulates germline proliferation in *Caenorhabditis elegans*. *Genesis* **33**, 1–7 (2002).
33. Capowski, E.E., Martin, P., Garvin, C. & Strome, S. Identification of grandchildless loci whose products are required for normal germline development in the nematode *Caenorhabditis elegans*. *Genetics* **129**, 1061–1072 (1991).
34. Mello, C.C. *et al.* The PIE-1 protein and germline specification in *C. elegans* embryos. *Nature* **382**, 710–712 (1996).
35. Seydoux, G. *et al.* Repression of gene expression in the embryonic germ lineage of *C. elegans*. *Nature* **382**, 713–716 (1996).
36. Reese, K.J., Dunn, M.A., Waddle, J.A. & Seydoux, G. Asymmetric segregation of PIE-1 in *C. elegans* is mediated by two complementary mechanisms that act through separate PIE-1 protein domains. *Mol. Cell* **6**, 445–55 (2000).
37. Marin, V.A. & Evans, T.C. Translational repression of a *C. elegans* Notch mRNA by the STAR/KH domain protein GLD-1. *Development* **130**, 2623–2632 (2003).
38. Li, Z., Zhang, Y., Li, D. & Feng, Y. Destabilization and mislocalization of myelin basic protein mRNAs in quaking dysmyelination lacking the QKI RNA-binding proteins. *J. Neurosci.* **20**, 4944–4953 (2000).
39. Ainger, K. *et al.* Transport and localization elements in myelin basic protein mRNA. *J. Cell Biol.* **138**, 1077–87 (1997).
40. Wu, J.I., Reed, R.B., Grabowski, P.J. & Artzt, K. Function of quaking in myelination: regulation of alternative splicing. *Proc. Natl. Acad. Sci. USA* **99**, 4233–4238 (2002).
41. Nabel-Rosen, H., Volohonsky, G., Reuveny, A., Zaidel-Bar, R. & Volk, T. Two isoforms of the *Drosophila* RNA-binding protein, How, act in opposing directions to regulate tendon cell differentiation. *Dev. Cell* **2**, 183–193 (2002).
42. Zaffran, S., Astier, M., Gratecos, D. & Semeriva, M. 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 (1997).
43. Reddy, T.R. *et al.* Inhibition of HIV replication by dominant negative mutants of Sam68, a functional homolog of HIV-1 Rev. *Nat. Med.* **5**, 635–642 (1999).
44. Lin, Q., Taylor, S.J. & Shalloway, D. Specificity and determinants of Sam68 RNA binding. Implications for the biological function of K homology domains. *J. Biol. Chem.* **272**, 27274–27280 (1997).
45. Batey, R.T. & Williamson, J.R. Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with 16 S rRNA: I. Defining the minimal RNA site. *J. Mol. Biol.* **261**, 536–549 (1996).