STAR-studded circuitry

Samuel E Butcher & Marvin Wickens

A recent study reports the RNA sequences that bind to the translational repressor protein GLD-1. The data suggest that a network of developmental genes may be regulated by GLD-1 or related STAR proteins through silencing or alternative splicing.

Many genes are regulated post-transcriptionally, through effects on RNA processing, stability and translation. The specific manner in which regulatory proteins recognize their target mRNAs is an important and rapidly developing area of investigation. In this issue of *Nature Structural & Molecular Biology*, the Williamson and Goodwin groups report the RNA-binding specificity of the translational repressor GLD-1, a STAR domain protein[1]. Their findings suggest that a complex circuitry of regulatory proteins may be post-transcriptionally controlled by GLD-1 or similar STAR proteins that specifically recognize a hexanucleotide RNA sequence within untranslated regions of mRNA. Additionally, the binding specificity of GLD-1 indicates that a subset of intronic branch site signals may also be targeted by GLD-1 with high affinity[1]. These findings suggest how regulatory proteins can target a network of mRNAs for translational repression and alternative splicing.

Cellular differentiation requires a delicate balance of gene expression, activation and repression. GLD-1 is a key developmental protein that controls germline development in *Caenorhabditis elegans*, in part by repressing the translation of *tra-2* mRNA[2]. For proper male development, the translation of *tra-2* mRNA must be specifically repressed by GLD-1 while other genes are translated. Although GLD-1 binding had been isolated to a 3′ untranslated region (UTR) of *tra-2* mRNA[2], it was unclear how this certain 3′ UTR was specifically recognized by GLD-1. Ryder *et al.*[1] have focused on the sequence specificity of GLD-1 as a representative of a much larger group of proteins, the STAR family. The hallmark features of STAR proteins include a common KH domain RNA-binding motif flanked by two additionally conserved domains known as Qua1 and Qua2. Qua1 domain mediates dimerization[1], whereas Qua2 probably participates with the KH

The authors are in the Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706, USA.
e-mail: butcher@nmrfam.wisc.edu or wickens@biochem.wisc.edu

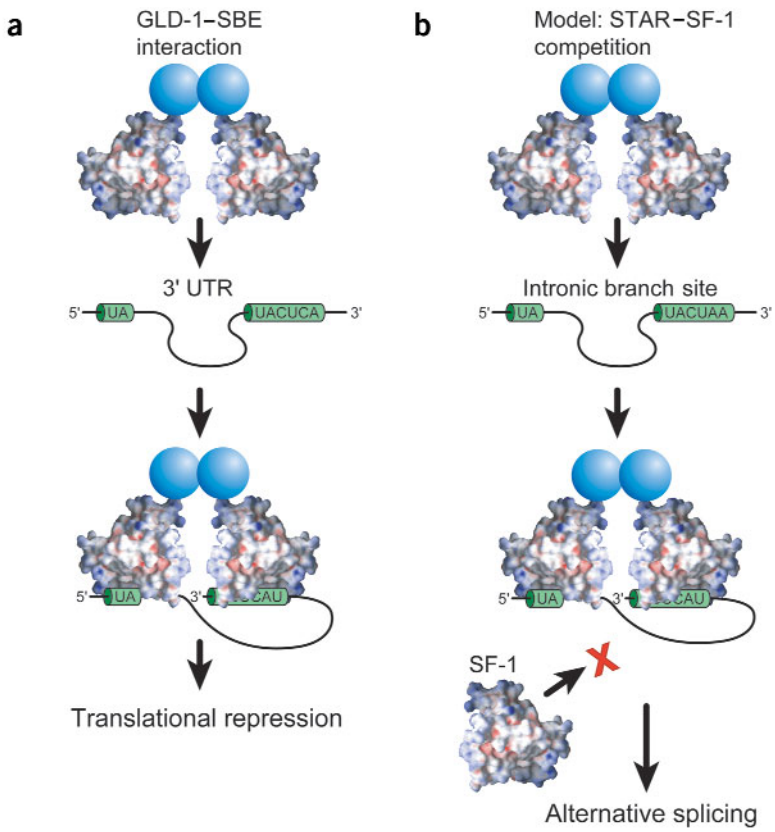

Figure 1 STAR domain proteins can post-transcriptionally regulate gene expression through translational repression or alternative splicing. The Qua1 dimerization domain of STAR proteins is a blue sphere. The Qua2-KH RNA-binding domain, represented as an electrostatic surface (blue, positive; white, neutral; red, negative) is modeled from the SF-1 NMR structure[3]. (**a**) The GLD-1 protein mediates translational repression by binding the STAR-binding element within the 3′ UTR of mRNA. (**b**) Proposed model for competition between STAR proteins and SF-1. STAR proteins are predicted to mediate alternative splicing by masking the intronic branch site and blocking access by the splicing factor SF-1 (ref. 1).

domain in forming an extended RNA-binding surface[3].

GLD-1 has substantial homology to the Quaking translational repressor[4], the How and Sam68 regulatory proteins involved in alternative pre-mRNA splicing[5,6] and the SF-1 mammalian intronic branch site RNA-binding protein[7]. Like GLD-1, the Quaking translational repressor functions as a homodimer[8], whereas SF-1 lacks the Qua1 dimerization domain and is therefore a monomer[3]. Ryder *et al.*[1] report that each subunit in the GLD-1 homodimer can bind to RNA with high affinity via the KH domain[1] (**Fig. 1a**). The high-affinity binding site, or STAR-binding element, was identified as the hexanucleotide consensus sequence UACU (C/A)A. Two copies of the hexanucleotide STAR-binding element exist within the 3′ UTR of *tra-2* mRNA.

Reasoning that GLD-1 might be responsible for the translational repression of other mRNAs, Ryder *et al.* inspected a set of 3′ UTRs from *C. elegans* for the hexanucleotide consensus sequence, and identified more than a dozen mRNAs carrying GLD-1-binding sites. Several of these targets were found within 3′ UTRs that were previously shown to interact with GLD-1 *in vivo*[9,10]. Ryder *et al.* also verified GLD-1 interactions *in vivo* with previously known STAR-binding elements[1].

The hexanucleotide STAR-binding element consensus sequence UACU(C/A)A corresponds to a subset of branch site signals within pre-mRNA introns that share the consensus YNCURAY (where Y indicates a pyrimidine, N any nucleotide and R a purine). This is notable because at least two STAR proteins are known to have a role in alternative splicing[5,6]. Furthermore, a required early step in pre-mRNA splicing is branch site recognition by the splicing factor SF-1 (ref. 11), a protein closely related to STAR domain proteins. The authors speculate that STAR proteins may therefore mask the branch site signal within an intron such that SF-1 cannot bind and target the intron for splicing by the spliceosome (**Fig. 1b**). The resulting competition between SF-1 and STAR proteins provides a model for the regulation of alternative splicing[1]. The dual roles of STAR proteins in translation and splicing may be akin to those of sex-lethal (Sxl) protein. This single protein binds to splice sites and inhibits splicing while also binding to UTRs to repress translation[12–14].

How does GLD-1 repress translation? The answer is unknown, but some data suggest that GLD-1 recruits other factors that shorten the length of the *tra-2* mRNA poly(A) tail[2]. The biochemical groundwork for STAR proteins has been laid, suggesting additional questions for structural and molecular biologists, such as: what are the atomic-level details of the interaction between STAR proteins and RNA? The conserved identity of amino acids involved in RNA recognition among STAR proteins suggests that they bind RNA in similar fashion[1]. The closely related SF-1 structure is known[3], but it is not yet clear why SF-1 is a monomer whereas GLD-1 and other STAR proteins presumably exist as homodimers. Perhaps the additional RNA-binding capacity of the GLD-1 homodimer allows it to more effectively sequester intronic branch sites away from SF-1 (**Fig. 1b**).

The identification of the hexanucleotide STAR-binding element has predictive power that can be used to elucidate complex modes of gene expression and regulation, as Ryder *et al.*[1] demonstrate in several instances. The mRNAs found to have STAR-binding element consensus sequences include an intriguing collection of regulatory proteins, phosphatases and RNA-binding proteins, as well as genes involved in sex determination, early embryogenesis and X-chromosome silencing[1]. As the authors conducted only a preliminary search of a few genes from *C. elegans*, it is likely that many more genes will be identified as STAR protein targets. Already, the identified STAR-binding elements suggest that STAR proteins may regulate an intricate network of genes at the post-transcriptional level, either by translational repression or by alternative splicing.

STAR proteins are not alone in this regard. In fact, the principle of coordinated control of functionally related mRNAs is an emerging hallmark of post-transcriptional regulators. Iron regulatory protein (IRP) regulates several mRNAs involved in iron metabolism, including those encoding ferritin and the transferrin receptor[15], and CPEB regulates multiple mRNAs crucial for control of the embryonic cell cycle[16]. The emerging post-transcriptional web of interconnected factors and co-regulation circuitry is akin to the sophisticated combinatorial regulation of transcription.

1. Ryder, S.P., Frater, L., Abramovitz, D.L., Goodwin, E.L. & Williamson, J.R. *Nat. Struct. Mol. Biol.* **11**, 20–28 (2004).
2. Jan, E., Motzny, C.K., Graves, L.E. & Goodwin, E.B. *EMBO J.* **18**, 258–269 (1999).
3. Liu, Z. *et al. Science* **294**, 1098–1102 (2001).
4. Saccomanno, L. *et al. Proc. Natl. Acad. Sci. USA* **96**, 12605–12610 (1999).
5. Nabel-Rosen, H., Dorevitch, N., Reuveny, A. & Volk, T. *Mol. Cell* **4**, 573–584 (1999).
6. Matter, N., Herrlich, P. & Konig, H. *Nature* **420**, 691–695 (2002).
7. Berglund, J. A., Chua, K., Abovich, N., Reed, R. & Rosbash, M. *Cell* **89**, 781–787 (1997).
8. Zorn, A.M. & Krieg, P.A. *Genes Dev.* **11**, 2176–2190 (1997).
9. Lee, M.H. & Schedl, T. *Genes Dev.* **15**, 2408–2420 (2001).
10. Marin, V.A. & Evans, T.C. *Development* **130**, 2623–2632 (2003).
11. Peled-Zehavi, H., Berglund, J.A., Rosbash, M. & Frankel, A.D. *Mol. Cell. Biol.* **21**, 5232–5241 (2001).
12. Bashaw, G.J. & Baker, B.S. *Cell* **89**, 789–798 (1997).
13. Grskovic, M., Hentze, M.W. & Gebauer, F. *EMBO J.* **22**, 5571–5581 (2003).
14. Penalva, L.O. & Sanchez, L. *Microbiol. Mol. Biol. Rev.* **67**, 343–359 (2003).
15. Eisenstein, R.S. *Annu. Rev. Nutr.* **20**, 627–662 (2000).
16. Mendez, R. & Richter, J.D. *Nat. Rev. Mol. Cell Biol.* **2**, 521–529 (2001).