Molecular Cell, Vol. 14, 841–847, June 18, 2004, Copyright ©2004 by Cell Press

The *C. elegans Polycomb* Gene *sop-2* Encodes an RNA Binding Protein

Short Article

Hong Zhang,[1] Andrea Christoforou,[1] L. Aravind,[2] Scott W. Emmons,[3] Sander van den Heuvel,[1] and Daniel A. Haber[1,*]

[1]Massachusetts General Hospital Cancer Center and Harvard Medical School
Charlestown, Massachusetts 02129
[2]Computational Biology Branch
National Center for Biotechnology Information
Bethesda, Massachusetts 20894
[3]Department of Molecular Genetics
Albert Einstein College of Medicine
Bronx, New York 10461

Summary

Epigenetic silencing of Hox cluster genes by Polycomb group (PcG) proteins is thought to involve the formation of a stably inherited repressive chromatin structure. Here we show that the *C. elegans*-specific PcG protein SOP-2 directly binds to RNA through three nonoverlapping regions, each of which is essential for its localization to characteristic nuclear bodies and for its in vivo function in the repression of Hox genes. Functional studies indicate that the RNA involved in SOP-2 binding is distinct from either siRNA or microRNA. Remarkably, the vertebrate PcG protein Rae28, which is functionally and structurally related to SOP-2, also binds to RNA through an FCS finger domain. Substitution of the Rae28 FCS finger for the essential RNA binding region of SOP-2 partially restores localization to nuclear bodies. These observations suggest that direct binding to RNA is an evolutionarily conserved and potentially important property of PcG proteins.

Introduction

Distinct RNA components have recently been recognized as essential for epigenetic silencing of genes packaged within heterochromatin (Grewal and Moazed, 2003). The Xist structural RNA is required for the initiation of X chromosome inactivation in mammals, and small interference RNAs (siRNA) are involved in silencing genes within heterochromatic domains in *S. pombe* and *Drosophila* (Dernburg and Karpen, 2002; Volpe et al., 2002; Pal-Bhadra et al., 2004). siRNA appears to be involved in the sequence-specific targeting of histone deacetylase and H3 lysine 9 methyltransferase, and in the subsequent recruitment of the heterochromatin binding protein (HP1) in *Drosophila* and *S. pombe* (Hall et al., 2002; Pal-Bhadra et al., 2004). Loss of function of components of the RNAi pathway results in reduction of H3 lysine 9 methylation and delocalization of *Drosophila* HP1 and its yeast homolog Swi6. In *S. pombe*, the RITS complex required for heterochromatin assembly has been shown to contain siRNA derived from centromeric repeats, the yeast Argonaute homolog, implicated in siRNA processing, and the Chp1 protein, which binds to centromeric repeats and is required for H3 lysine 9 methylation and for recruitment of Swi6 (Verdel et al., 2004). Heterochromatin assembly therefore involves protein-RNA complexes, whose components presumably mediate sequence recognition, initiation, and maintenance of inactive chromatin structure.

Epigenetic transcriptional repression of the Hox cluster genes mediated by the Polycomb group (PcG) proteins is also thought to involve the formation of localized heterochromatic-like structures, including deacetylation and methylation of histone tails (Francis and Kingston, 2001). Of the two distinct complexes formed by the PcG proteins, the ESC-E(Z) complex associates with histone deacetylases and the SET domain of E(Z) possesses H3 specific methyltransferase activity, while the PRC1 complex prevents access of the SWI/SNF nucleosome-remodeling factors (Otte and Kwaks, 2003). Components of the PRC1 complex are also remarkable for their localization to distinct nuclear speckles, called PcG bodies, whose role in transcriptional repression is not well understood (Francis and Kingston, 2001). Recently a link between the PcG complex and some RNA-dependent silencing processes has emerged in contexts other than Hox gene regulation. For example, the mammalian ESC-E(Z) complex, EED/ENX1, is recruited to the inactivated X chromosome in an Xist RNA-dependent manner, and it is required for X chromosome inactivation (Plath et al., 2003; Silva et al., 2003). Silencing of tandem transgenic arrays in *C. elegans* germ cells and cosuppression in *Drosophila* require both the PcG complex and components of the RNAi pathway (Kelly and Fire, 1998; Tabara et al., 1999; Pal-Bhadra et al., 2002). However, it is unclear whether these effects result from direct RNA-PcG protein interactions and whether they contribute to Hox gene repression, a primary property of PcG proteins that is conserved across the phylogenetic spectrum.

We have recently isolated the *C. elegans*-specific *PcG* gene, *sop-2*, which is involved in maintaining the repressed state of Hox genes (Zhang et al., 2003). *sop-2* encodes a protein-protein interaction domain, the SAM domain, which is also present in Polyhomeotic (PH) and Sex comb on midleg (SCM), both of which are components of the PRC1 PcG complex. Here we demonstrate that SOP-2 is an RNA binding protein and that its three RNA binding domains are required for its function in Hox gene repression. RNA binding by PcG proteins is conserved across the phylogenetic spectrum, pointing to its physiological importance in Hox gene repression.

Results and Discussion

RNA Binding by SOP-2

The postulated role of RNA intermediates in PcG-mediated transcriptional silencing led us to test whether SOP-2 directly binds to RNA. Full-length GST-SOP-2 was purified from *E. coli* and incubated with a radiolabeled single-stranded RNA (ssRNA) probe derived from

*Correspondence: haber@helix.mgh.harvard.edu

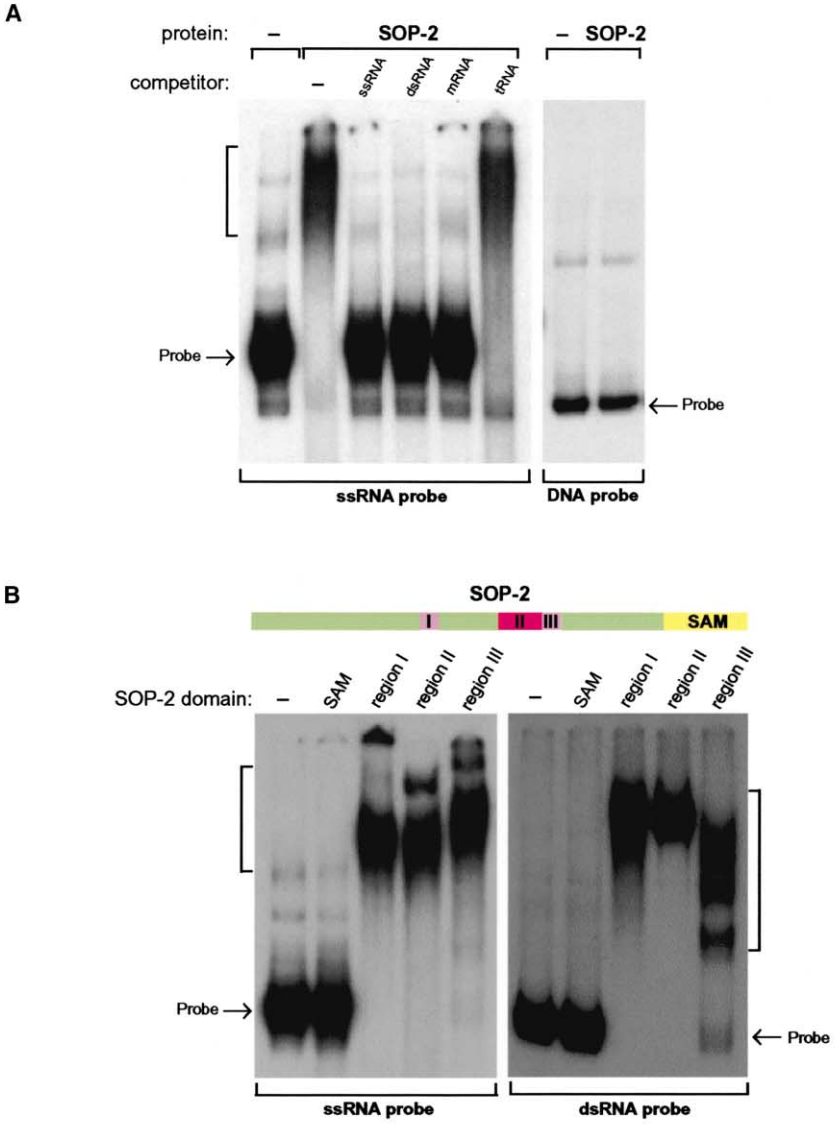


Figure 1. RNA Binding by SOP-2

(A) Direct RNA binding by SOP-2. The GST-SOP-2 was incubated with radiolabeled RNA (5′ UTR of *egl-5*) in EMSA. Binding of SOP-2 to RNA is competed by the cold unrelated single-stranded RNA (ssRNA) (10× excess) (lane 3), double-stranded RNA (dsRNA) (lane 4), mRNA (lane 5), but not by yeast tRNA (lane 6). SOP-2 does not bind to the corresponding DNA probe. Protein-RNA complexes are marked by bracket.

(B) The RNA binding regions of SOP-2 identified by EMSA (I, II, and III) are shown schematically, N-terminal to the protein interaction SAM domain. Each region binds strongly to both ssRNA and dsRNA. No RNA binding is observed with the SAM domain.

(C) Amino acid motifs within RNA binding regions of SOP-2. Region I and III contain RGG repeats similar to those found in other RNA binding proteins. Both of these RNA binding regions are associated with a downstream proline-rich region (PRR). RNA binding region II is rich in arginine and lysine residues, which is also characteristic of other RNA binding proteins (see text).

the 5′ UTR of the Hox gene *egl-5* transcript, followed by electrophoretic mobility shift analysis (EMSA). Indeed, purified SOP-2 binds efficiently to the RNA probe (Figure 1A). Binding of SOP-2 to RNA templates is effectively competed by cold unrelated ssRNA or by double-stranded RNA (dsRNA), but not by tRNA, suggesting that this interaction is saturable, and affected by the tertiary structure of the RNA template (Figure 1A). RNA

binding by SOP-2 is not sequence specific, since several unrelated RNA templates bind with comparable efficiency. SOP-2 also binds to dsRNA, but not to a corresponding DNA probe (Figure 1A). As for other RNA binding proteins, the apparent lack of sequence specificity in vitro may not reflect differential affinity for more physiologically relevant RNA targets in vivo (Akhtar et al., 2000).

The specific RNA binding regions of SOP-2 were mapped by EMSA, using a series of 16 overlapping SOP-2 fragments. The RNA binding domain of SOP-2 does not map to its C-terminal SAM domain, implicated in protein-protein interaction. Instead, RNA binding by SOP-2 is mediated by three nonoverlapping regions, which we have designated RNA binding regions I (amino acids 224–254), II (amino acids 324–414), and III (amino acids 416–455) (Figure 1B). All three regions bind strongly to both ssRNA and dsRNA (Figure 1B). Computational analysis of these regions indicates that they are low entropy sequences that are likely to adopt a nonglobular structure. Regions I and III are enriched in charged amino acids corresponding to repeats of the RGG motif which is known to bind RNA (Figure 1C) (Krecic and Swanson, 1999). Interestingly, both regions I and III are followed by a proline-rich region (PRR), which is potentially involved in protein-protein interactions (Kay et al., 2000), and they demonstrate significant sequence similarity to each other over a stretch of 89 amino acid (32% identity and 40% similarity) (Figure 1C). In contrast, SOP-2 RNA binding region II is rich in arginine and lysine residues, and is compositionally similar to the low entropy RNA binding regions of certain ribosomal proteins, the Tat protein of HIV, the Epstein-Barr virus (EBV) mRNA export factor EB2, and the RNA binding region of HP1 (Hiriart et al., 2003; Muchardt et al., 2002).

Contribution of SOP-2 RNA Binding Domains to Hox Gene Repression

To explore the physiological consequences of RNA binding by SOP-2, we made chimeric GFP-tagged constructs with an in-frame deletion of each minimal RNA binding region, and examined their cellular localization and functional properties in vivo. Like the vertebrate PRC1 PcG complex, SOP-2 is localized to characteristic nuclear speckles, called SOP-2 bodies, postulated to be enriched in components of the transcriptional repression machinery (Zhang et al., 2003). SOP-2(Δregion I)::GFP does form nuclear bodies, although they are slightly reduced in number, compared to the nuclear bodies formed by wild-type SOP-2::GFP (Figures 2A, 2B, and 2E). In some cells, these nuclear bodies appear irregular and fused together (Figure 2B). SOP-2(Δregion II)::GFP shows a more significant decrease in the number of nuclear bodies, with highly variable shape and size (Figures 2C and 2E). SOP-2(Δregion III)::GFP is localized within one or two large nuclear bodies, with some homogeneous nuclear staining (Figures 2D and 2E). While extending the deletion of region I to include the adjacent PRR region has no further effect on localization, deleting both region III and its adjacent PRR region results in the appearance of cytoplasmic bodies (data not shown). Taken together, these results indicate that each of the RNA binding segments contributes to the appropriate nuclear localization of SOP-2, with region III being essential for the formation of these nuclear bodies.

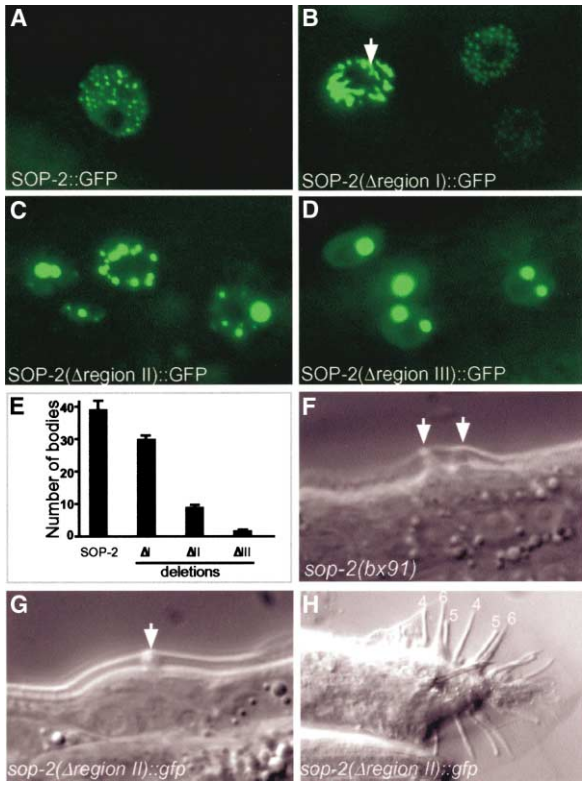


Figure 2. Requirement of SOP-2 RNA Binding Regions for Nuclear Localization and Hox Gene Repression

Amino acids 224–254, 324–414, or 416–455 were deleted to make *sop-2(Δregion I)::gfp, sop-2(Δregion II)::gfp*, and *sop-2(Δregion III)::gfp* reporter, respectively, within the previously described functional *sop-2::gfp* reporter (Zhang et al., 2003).

(A) Formation of nuclear bodies by wild-type SOP-2::GFP in hypodermal cells of a transgenic animal.

(B) Reduction in number of nuclear bodies by SOP-2(Δregion I)::GFP in hypodermal cells. In some cells, the bodies are fused together (arrow).

(C) Expression of SOP-2(Δregion II)::GFP showing defects in formation of SOP-2 bodies. The number of bodies is significantly decreased, and their shape and size are highly variable.

(D) Expression of SOP-2(Δregion III)::GFP in hypodermal cells. GFP is present in one or few bodies within the nucleus, with some diffuse staining.

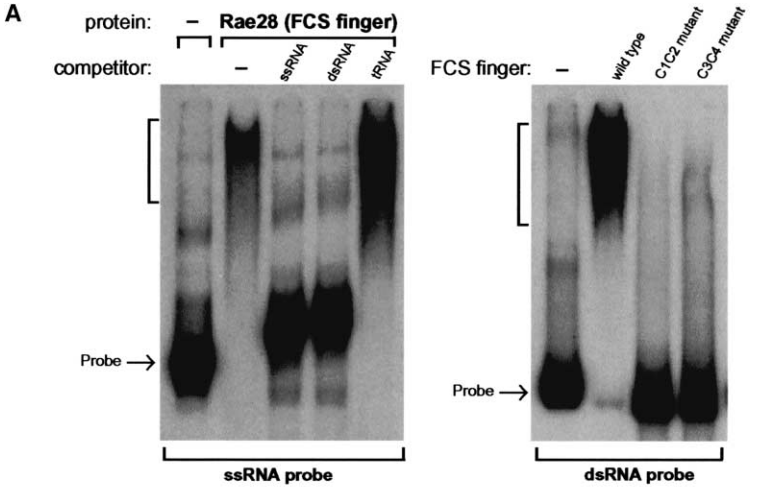
(E) The average number of bodies/hypodermal cells is shown with standard error, following analysis of multiple cells expressing SOP-2::GFP (36.7 ± 1.2, number of cells counted, n = 32), SOP-2(Δregion I)::GFP (29.9 ± 1.2, n = 22), SOP-2(Δregion II)::GFP (8.2 ± 0.5, n = 78), SOP-2(Δregion III)::GFP (2.6 ± 0.2, n = 61).

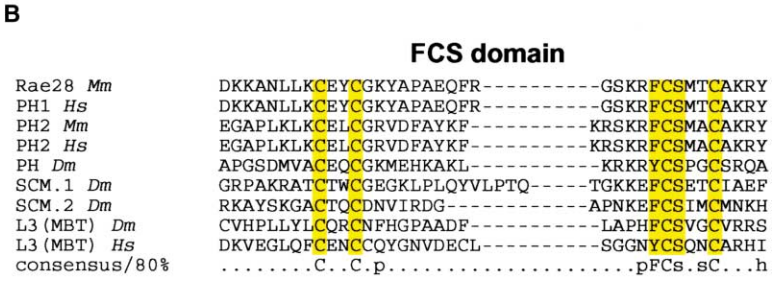
(F) Ectopic rays, indicating homeotic transformation, are generated from the anterior seam cells in *sop-2(bx91)* mutants (arrow).

(G) Expression of *sop-2(Δregion II)::gfp* in wild-type animals phenocopies the *sop-2* mutant phenotypes. An ectopic ray is generated from anterior seam cells in the animal shown (arrows). The frequency of generation of ectopic rays is correlated with the expression level of the transgene.

(H) Expression of *sop-2(Δregion II)::gfp* in wild-type animals leads to duplication of V6 rays. One side of the male tail is shown with extra rays 4, 5, and 6, duplicated at the expense of rays 2 and 3.

The functional consequence of deleting individual SOP-2 RNA binding regions was examined with respect





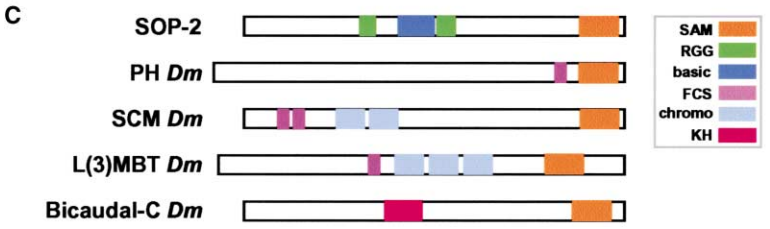


Figure 3. RNA Binding by Mouse PH Homolog Rae28

(A) Binding of RNA by the mouse PH homolog Rae28. GST fusion fragment of Rae28 (amino acids 775 to 860) binds RNA by EMSA. Binding is competed by unlabeled ssRNA (10×) and dsRNA, but not tRNA. The RNA binding activity of Rae28 FCS domain is abrogated in mutants with cysteines to serines substitutions at either the first two conserved residues (C1C2) or last two residues (C3C4).

(B) Presence of conserved FCS domains in PcG and related proteins. Conserved cysteines and the FCS signature residues are highlighted in yellow. *Dm, Drosophila melanogaster; Mm*, *Mus musculus; Hs, Homo sapiens*.

(C) Shared architectural structure of PcG and related proteins, containing a protein interaction SAM domain at their C terminus and distinct RNA binding modules at their N terminus. The FCS and KH RNA binding motifs are readily recognized from their consensus residues. The RNA binding version of chromodomains is evident by computational analysis in SCM (amino acids 175–273 and 281–382) and in L(3)MBT (amino acids 819–918, 926–1028, and 1035–1130).

to in vivo repression of Hox gene expression. Hox genes in *C. elegans* are well characterized for their role in patterning the anterior-posterior identity adopted by a row of epidermal seam cells V1 to V6 during male development (Emmons, 1999). In *sop-2(bx91)* mutants, ectopic expression of the Hox genes *mab-5* or *egl-5* in anterior seam cells V1 to V4 leads to the generation of male sensory organs called rays (Figure 2F), rather than the normal cuticle stripes called alae (ie. homeotic transformation) (Zhang et al., 2003). Transgenic expression of wild-type SOP-2::GFP effectively rescues the homeotic transformations seen in *sop-2* mutants. In contrast, SOP-2 proteins lacking either RNA binding region I, II, or III fail to rescue these homeotic transformations (data not shown). Moreover, ectopic expression of *sop-2(Δregion II)::gfp* or *sop-2(Δregion III)::gfp* in wild-type animals phenocopies *sop-2(bx91)* mutants, including generation of ectopic rays from anterior seam cells and Male abnormal (Mab) phenotype (Figures 2G and 2H). This apparent dominant-negative effect may be associated with the mislocalization of SOP-2 within abnormal nuclear complexes. Taken together, these observations indicate that deletion of any of the minimal RNA binding domains within SOP-2 leads to abnormal protein function.

Independence of SOP-2-Mediated Hox Gene Repression from siRNA and miRNA Pathways

While these observations point to a critical role for RNA binding in SOP-2-mediated repression of Hox genes, the specific type of RNA involved is unknown. Small interfering RNAs (RNAi) appear to be required for triggering epigenetic gene silencing in plants and *S. pombe* (Mette et al., 2000; Volpe et al., 2002). Furthermore, PcG proteins have been shown to be involved in several RNAi-dependent processes. The ESC-E(Z) complex and components of the RNAi pathway are required for silencing of tandem transgenic arrays in *C. elegans* germ cells (Kelly and Fire, 1998; Tabara et al., 1999). *PcG* and *piwi*, implicated in the RNAi pathway, are involved in cosup-

pression in *Drosophila* (Pal-Bhadra et al., 2002). However, we did not observe a role of the RNAi pathway in Hox gene repression in *C. elegans*. None of a series of mutants with defects in components of the RNAi pathway (including *dcr-1, rde-1, rde-2, rde-4*, *ego-1*, *drh-1*, and *mut-7*) showed ectopic expression of Hox genes, as demonstrated by the *egl-5* reporter, or homeotic transformations, scored using the ray-specific marker *pkd-2::gfp* (Zhang et al., 2003). *dcr-1* is also essential for the generation of microRNAs (miRNAs), a second form of small RNAs that have been linked to temporal cell fate changes (Grishok et al., 2001). Targeting other components of the miRNA pathway, *alg-1/alg-2* using RNAi, also does not affect the expression of Hox genes, indicating that miRNAs are unlikely to be involved in SOP-2 function. No such definitive tests are available for structural RNAs, which represent a third possible RNA target for SOP-2 binding. In fact, a role for structural RNAs would be consistent with observations in vertebrate systems, including the recruitment of the ESC-E(Z) PcG complex to the inactive X chromosome by Xist RNA (Plath et al., 2003; Silva et al., 2003). Hence, binding of SOP-2 to structural RNAs that mark loci targeted for silencing provides a viable hypothesis.

Evolutionary Conservation of RNA Binding by Vertebrate PcG Proteins

To determine whether RNA binding is an evolutionarily conserved property of PcG proteins, we extended our studies to the more complex and functionally redundant PcG proteins of arthropods and vertebrates. An obvious ortholog of SOP-2 is not observed among these, but both PH and SCM have similar domain architecture and functional properties, including a C-terminal SAM domain and localization to nuclear bodies. We therefore tested the mouse PH homolog Rae28 for RNA binding activity in EMSA experiments, using 11 overlapping fragments produced as GST-fusion proteins. Strong RNA binding activity was mapped to amino acids 775 to 860 of Rae28 (Figure 3A). As for SOP-2, RNA binding by this fragment of Rae28 is effectively competed by single- or double-stranded RNA, but not by tRNA (Figure 3A). Computational analysis of this minimal RNA binding region revealed a previously unappreciated Zn-chelating domain with four conserved cysteines, called the FCS finger (after the signature FCS residues associated with the third cysteine), which is thought to assume a Zn-ribbon like fold. Consistent with its role in RNA binding, mutating the first two or last two conserved cysteines to serines within the Rae28 FCS finger abolishes RNA binding activity (Figure 3A). Interestingly, the Rae28 FCS finger also exhibits non-sequence-specific binding to DNA, an effect that is primarily dependent on the first two conserved cysteine residues (data not shown). While the nucleic acid binding properties of FCS fingers have not been studied, the Cys_2His_2-type zinc fingers of transcription factor TFIIIA are known to bind both RNA and DNA (Lu et al., 2003). Thus, vertebrate PH proteins demonstrate strong RNA binding activity through a functional domain that is distinct from that of the *C. elegans* protein SOP-2.

The demonstration of RNA binding activity by FCS fingers is particularly significant since this domain is

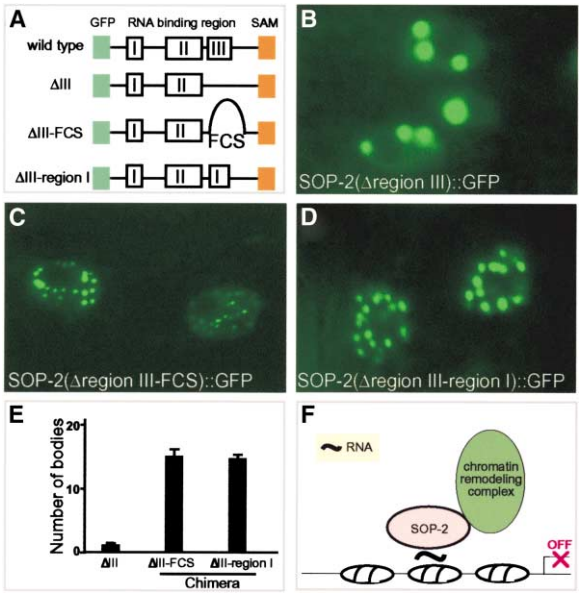


Figure 4. Restoration of the Nuclear Localization of SOP-2(Δregion III)::GFP by the FCS Domain of Rae28 and the RNA Binding Region I of SOP-2

(A) Schematic representation of chimeric GFP-tagged SOP-2 constructs in which RNA binding region III is replaced with either the mouse Rae28-derived FCS domain or by RNA binding region I of SOP-2 itself.

(B) Abrogation of nuclear body formation as previously shown in hypodermal cells expressing SOP-2(Δregion III)::GFP.

(C) Partial restoration of nuclear bodies formed by SOP-2(Δregion III-FCS)::GFP chimeric reporter.

(D) Partial restoration of nuclear bodies formed by SOP-2(Δregion III-region I)::GFP chimeric reporter.

(E) The average number of bodies/hypodermal cells is shown with standard error, following analysis of multiple cells expressing SOP-2(Δregion III)::GFP (2.6 $\pm$ 0.2, n = 61), SOP-2(Δregion III-FCS)::GFP (15.7 $\pm$ 1.8, n = 23), SOP-2(Δregion III-region I)::GFP (15.2 $\pm$ 0.9, n = 40).

(F) Schematic representation of SOP-2-mediated transcriptional repression. According to this model, SOP-2 may be recruited to the appropriate chromatic locus by its interaction with a structural RNA. SOP-2 in turn recruits the chromatin-remodeling complex which confers a repressive chromatin structure, such as the NURD histone deacetylation complex.

widely conserved, being present in the *Drosophila* proteins PH, SCM, and Lethal 3 (MBT) (L3(MBT)), their vertebrate homologs, the vertebrate FIM protein, and the A1L-like transcription factors of poxviruses and other large DNA viruses (Figure 3B and data not shown). Moreover, SCM and L3(MBT) also contain repeats of RNA binding version of chromodomains, such as those in MSL-3 and MOF (Akhtar et al., 2000), suggesting that they may in fact contain multiple RNA binding motifs. It is noteworthy that SOP-2, PH, SCM, and L3(MBT) all have a C-terminal protein interaction SAM domain linked with an N-terminal RNA binding module; this combination is also present in more distant protein families, such as the Bicaudal-C protein and its orthologs, which have a KH RNA binding motif (Figure 3C) (Mahone et al., 1995). The combination of a C-terminal SAM domain involved in protein-protein interactions with different N-terminal RNA binding domains may therefore define a common architectural theme across the phylogenic spectrum.

To determine whether the RNA binding motifs of Rae28 and SOP-2 share some functional similarity, we substituted RNA binding region III of SOP-2, which is essential for the formation of nuclear bodies, with the FCS domain of mouse Rae28 (Figure 4A). Although the chimeric construct encoding the mouse-derived FCS RNA binding domain is not functional in rescuing the homeotic defects seen in *sop-2(bx91)* mutants, it partially restores the nuclear localization of the SOP-2(Δregion III)::GFP protein. While SOP-2(Δregion III)::GFP forms 2.6 ± 0.2 nuclear bodies (n = 61), the number of bodies in the chimeric SOP-2(Δregion III-FCS)::GFP is dramatically increased to 15.7 ± 1.8 (n = 23) (Figures 4B, 4C, and 4E). A similar partial rescue of nuclear body formation is evident following substitution of RNA binding region III with RNA binding region I of SOP-2, which forms 15.2 ± 0.9 bodies (n = 40) (Figures 4D and 4E). Taken together, these studies further support the importance of RNA binding for the formation of SOP-2 nuclear bodies. The ability of the divergent mouse and *C. elegans* RNA binding domains to confer nuclear localization to SOP-2 suggests that formation of nuclear bodies may not involve sequence-specific RNA binding.

Concluding Remarks

An RNA-dependent step in Hox gene repression by PcG proteins has been postulated (Grewal and Moazed, 2003; Richards and Elgin, 2002), but direct RNA binding by any of these proteins has not been demonstrated. RNA binding by PcG proteins may contribute to transcriptional silencing in a number of ways. By analogy to its roles in initiating epigenetic silencing in *S. pombe* and X chromosome inactivation in mammalian cells (Volpe et al., 2002; Plath et al., 2003; Silva et al., 2003), RNA may direct PcG proteins to the appropriate target locus, where they in turn recruit chromatin-remodeling complexes through their protein interaction domains, such as the SAM domain (Figure 4F). Moreover, as demonstrated by the deletion of RNA binding domains and by chimeric constructs derived from heterologous RNA binding motifs, RNA binding also contributes to the nucleation of SOP-2 bodies. A similar nucleation function for RNA has been proposed for the silencing protein HP1, which is localized to pericentromeric heterochromatin domains (Maison et al., 2002; Muchardt et al., 2002).

PcG-mediated Hox gene repression is highly conserved in *Drosophila* and vertebrates. However, *C. elegans* lacks a defined homolog of the components of the PRC1 complex, and our studies indicate that SOP-2 may have assumed this role. In spite of limited sequence similarity between SOP-2 and the PRC1 components PH and SCM, they share a C-terminal SAM domain, and the SAM domain of SOP-2 can be functionally replaced by the SAM domains derived from *Drosophila* or mammalian PcG proteins, but not by those of non-PcG proteins (including TEL, L(3)MBT) (Zhang et al., 2004). The presence of RNA binding motifs in SOP-2 and PH proteins further supports their functional similarity. Taken together, the combination of SAM domain and RNA binding motifs define a class of PcG proteins, characterized by their nuclear localization as well as their functional properties. *C. elegans* SOP-2 and vertebrate PcG proteins may thus have evolved separately toward shared functional properties.

Experimental Procedures

Strains

The following strains were used in this work: LGI, *rde-2(ne221)*; LGII, *sop-2(bx91)*; LGIII, *mut-7(pk204), pha-1(e2123ts)*; LGV, *him-5(e1490), bxIs14(pkd-2::gfp, pha-1(+))*; LGX, *bxIs13(egl-5::gfp, lin-15(+)), let-7(n2853)*.

EMSA and RNA Binding Assays

RNA binding reactions contained 20 mM HEPES (pH 7.6), 100 mM KCl, 2 mM EDTA, 0.01% NP40, 1 mM DTT, labeled RNA fragment (20,000 c.p.m.). Reactions were incubated on ice for 20 min and eletrophoresed on 4% native TBE PAGE gel.

RNAi Experiments

The following DNA templates were used for RNA synthesis: *ego-1* (F26A3.3, nt 17721–18460); *rde-1* (K08H10.7, nt 3261–4236); *drh-1* (F15B10.2, nt 10239–11119); *rde-4* (T20G5.11, nt 45635–46467); *alg-1* (F48F7.1, nt 661–1522); *dcr-1* (K12H4.8, nt 37410–38147); *alg-2* (T07D3.7, nt 2298–3207).

Sequence Analysis

The nonredundant (NR) database of protein sequences (National Center for Biotechnology Information, NIH, Bethesda) was searched using the BLASTP program. Profile searches were conducted using the PSI-BLAST program with either a single sequence or an alignment used as the query, with a default profile inclusion expectation (E) value threshold of 0.01 (unless specified otherwise), and was iterated until convergence. For all searches of compositionally biased proteins, we used a statistical correction for this bias to reduce false positives in these searches. Multiple alignments were constructed using the T_Coffee or PCMA programs, followed by manual correction based on the PSI-BLAST results. The FCS domain was detected by comparing the region determined to bind RNA with a library of profiles, which were constructed for use with the RPS-BLAST program based on alignments of conserved domains. A significant hit, e value = 10^{-8} was obtained for the profile of the FCS domain, while all other hits were not statistically significant. Protein secondary structure was predicted using a multiple alignment as the input for the PHD and PROF programs. Similarity based clustering of proteins was carried out using the BLASTCLUST program.

Acknowledgments

This work was supported by the NIH (2T32CA09216, H.Z.; GM63081, S.v.d.H.; CA58596, D.A.H.).

Received: March 1, 2004
Revised: April 23, 2004
Accepted: April 27, 2004
Published: June 17, 2004

References

Akhtar, A., Zink, D., and Becker, P.B. (2000). Chromodomains are protein-RNA interaction modules. Nature *407*, 405–409.

Dernburg, A.F., and Karpen, G.H. (2002). A chromosome RNAissance. Cell *111*, 159–162.

Emmons, S.W. (1999). In Cell Lineage and Fate Determination, S.A. Moody, ed. (San Diego: Academic Press), pp. 139–155.

Francis, N.J., and Kingston, R.E. (2001). Mechanisms of transcriptional memory. Nat. Rev. Mol. Cell Biol. *2*, 409–421.

Grewal, S.I., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. Science *301*, 798–802.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell *106*, 23–34.

Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232–2237.

Hiriart, E., Bardouillet, L., Manet, E., Gruffat, H., Penin, F., Montserret, R., Farjot, G., and Sergeant, A. (2003). A region of the Epstein-Barr virus (EBV) mRNA export factor EB2 containing an arginine-rich motif mediates direct binding to RNA. J. Biol. Chem. *278*, 37790–37798.

Kay, B.K., Williamson, M.P., and Sudol, M. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J. *14*, 231–241.

Kelly, W.G., and Fire, A. (1998). Chromatin silencing and the maintenance of a functional germline in *Caenorhabditis elegans*. Development *125*, 2451–2456.

Krecic, A.M., and Swanson, M.S. (1999). hnRNP complexes: composition, structure, and function. Curr. Opin. Cell Biol. *11*, 363–371.

Lu, D., Searles, M.A., and Klug, A. (2003). Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. Nature *426*, 96–100.

Mahone, M., Saffman, E.E., and Lasko, P.F. (1995). Localized Bicaudal-C RNA encodes a protein containing a KH domain, the RNA binding motif of FMR1. EMBO J. *14*, 2043–2055.

Maison, C., Bailly, D., Peters, A.H., Quivy, J.P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., and Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat. Genet. *30*, 329–334.

Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. *19*, 5194–5201.

Muchardt, C., Guilleme, M., Seeler, J.S., Trouche, D., Dejean, A., and Yaniv, M. (2002). Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. EMBO Rep. *3*, 975–981.

Otte, A.P., and Kwaks, T.H. (2003). Gene repression by Polycomb group protein complexes: a distinct complex for every occasion? Curr. Opin. Genet. Dev. *13*, 448–454.

Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. (2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. Mol. Cell *9*, 315–327.

Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C. (2004). Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. Science *303*, 669–672.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135.

Richards, E.J., and Elgin, S.C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell *108*, 489–500.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev. Cell *4*, 481–495.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. Cell *99*, 123–132.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672–676.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833–1837.

Zhang, H., Azevedo, R.B., Lints, R., Doyle, C., Teng, Y., Haber, D., and Emmons, S.W. (2003). Global regulation of Hox gene expression in *C. elegans* by a SAM domain protein. Dev. Cell *4*, 903–915.

Zhang, H., Smolen, G.A., Palmer, R., Christoforou, A., van den Heuvel, S., and Haber, D.A. (2004). SUMO modification is required for *in vivo* Hox gene regulation by the *C. elegans* Polycomb group protein SOP-2. Nat. Genet. *36*, 507–511.