

REVIEW

RNA-binding proteins as developmental regulators

Raymond J. Bandziulis, Maurice S. Swanson, and Gideon Dreyfuss

Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, Illinois 60208 USA.

Many RNA-binding proteins of the nucleus and cytoplasm, including pre-mRNA-, mRNA-, snRNA-, and pre-rRNA-binding proteins, contain a putative RNA-binding domain of approximately 90 amino acids whose amino acid sequence is conserved from yeast to man. The most highly conserved motif within this RNA-binding domain is an octapeptide, termed the ribonucleoprotein consensus sequence (RNP-CS), which is an identifying feature of this group of proteins. Frequently, these proteins contain several similar, but nonidentical, RNP-CS-type RNA-binding domains. All of these proteins also contain at least one auxiliary domain that is unique to each type of protein and most likely functions in protein-protein interactions. Many, if not all, of the RNP-CS-type proteins display binding preferences for specific RNA sequences, and several have been shown to interact with pre-mRNA sequences important for pre-mRNA processing. Recent work has shown that the proteins encoded by several developmental loci in *Drosophila* and maize contain RNP-CS and, therefore, are most likely RNA-binding proteins. Here we provide an overview of the structural characteristics of these proteins and speculate on how the modular structure of RNP-CS-type RNA-binding proteins may facilitate their participation in pathways that regulate development at the post-transcriptional level.

Identification of the RNP consensus sequence

The RNP-CS (Adam et al. 1986; Swanson et al. 1987) was identified initially in the yeast mRNA poly(A)-binding protein (PABP) (Adam et al. 1986; Sachs et al. 1986) and in the mammalian heterogeneous nuclear ribonucleoprotein particle (hnRNP) protein A1 (Cobianchi et al. 1986; Riva et al. 1986), and subsequently has been found in many other proteins (Lahiri and Thomas 1986; Theissen et al. 1986; Grange et al. 1987; Haynes et al. 1987; Jong et al. 1987; Lapeyre et al. 1987; Sillikens et al. 1987; Amrein et al. 1988; Bell et al. 1988; Chambers et al. 1988; Deutscher et al. 1988; Gomez et al. 1988; Preugschat and Wold 1988; Robinow et al. 1988; Gorski et al. 1989). This octapeptide consensus sequence, Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe-Val-X-Phe/Tyr, is the most highly conserved segment in a generally conserved region consisting of about 90 amino acids present at

least once in all of these proteins. This region, because of its general conservation as a unit and because of experimental findings discussed below, is considered to be an RNA-binding domain. In this review, we shall refer to the 90-amino-acid region as an RNP-CS-type RNA-binding domain and to the more highly conserved octapeptide sequence as RNP-CS.

Figure 1 lists the proteins in which the RNP-CS sequence has been found thus far. All of the RNP-CS-type proteins are known RNA-binding proteins, except for the developmental control genes discussed below. The RNP-CS motif defines only one class of RNA-binding protein, because many well-characterized RNA-binding proteins exist (e.g., ribosomal proteins, viral nucleocapsid proteins) that do not contain this motif. In addition to the RNP-CS (RNP 1 in Fig. 1), alignment of the relevant RNP-CS-containing segments of each of the proteins reveals that several other amino acids at specific positions in the RNP-CS-type RNA-binding domain also are conserved, for example, a 6-amino-acid segment, RNP 2 (Dreyfuss et al. 1988), located about 30 amino acids amino-terminal to the RNP-CS. RNP 2 is less well conserved than the RNP-CS, yet the general character of its amino acids is preserved. Sequence similarities also have been reported between the eukaryotic RNP-CS and several prokaryotic ssDNA-binding proteins (including T4 gp32, M13 gp5, and *E. coli* ssb) and bacterial σ transcription factors (Helmann and Chamberlin 1988).

What is the evidence that the RNP-CS-type RNA-binding domain is responsible for the RNA-binding properties displayed by each of these proteins? Characterization of UP1, the mammalian single-stranded DNA-binding protein (Herrick and Alberts 1976), has revealed that this 24-kD polypeptide is an amino-terminal fragment of the hnRNP protein A1 that results from proteolysis in vitro (Kumar et al. 1986; Riva et al. 1986). UP1, which is composed of two similar domains of approximately 90 amino acids each, retains the ability to bind single-stranded DNA or RNA (Kumar et al. 1986; Riva et al. 1986), and recent photo-cross-linking experiments have demonstrated that residues of the RNP-CS and RNP 2 within each of these two domains is directly in contact with RNA (Merrill et al. 1988). Furthermore, experiments using fragments of the human hnRNP C1 protein that contain only the first 150 amino-terminal

Protein	Domain	RNA Binding Domain				
hnRNP						
		RNP 2	RNP 1			
A1 rat	1	SPKEPEQLRK LFVGG .SFETIDESLSRSHFEQ.WGTLTDCVVMRDPNPKRS.	RGFGVTF ATVEEVDAAAMNARPHK.VDGRVVEPKRAVS			
A1 <i>drosophila</i>	1	SITEPEHMRK LFVGG .DYRTIDENLKAHFEK.WGNIVDVVMKDPRTKRS.	RGFGVTF SHSSMIDEAQKSRPHK.IDGRVVEPKRAVP			
A1 rat	2	RPGAHLTVKK IFVGGI .KEDTEHHLRDVFQK.YGKIEVIEIMTDRGSGKK.	RGFAFVF DDHSDVDKIVIQKYHT.VNGHNCVEVRKALS			
A1 <i>drosophila</i>	2	SPNAGATVKK IFVGGAL .KDDHDEQSIIRDYFQH.FGNIVDINIVIDKETGKK.	RGFAFVF DDYDPPVDKVVLQKQHQ.LNGKMDVVKKALP			
C1 human	-	KTDP RSMNSR VFVGNL NTLVVKKSDVEAIFSK.YGKIV..... GCSVH.	RGFAFVF VNERNARA VAGEDGRMIAGQVLDINLAAE			
C1 <i>xenopus</i>	-	KTDP RSMNSR VFVGNL NTLVVKKSDVEAIFSK.YGKIV..... GCSVH.	RGFAFVF SNERTARTAVAGEDGRMIAGQVLDINLAAE			
E (UP2) human	-	AMKTKEPVKK IFVGG SPDTPEEK.IREYFGG.FGEVESIE LPMDNKTKNR.	RGFCFITF NQEEP			
snRNP						
U1A human	-	AVPETRNPH IYINNL NEKIKKDE.LKKSLYAIFSFQGGIILDIVSRSLKM.	RGQAFVIF KEVSSATNALRSMQGFPPFYDKPMRIQYAKT			
U1B human	-	MDI..RPNHT IYINNM NDKIKKKE.LKRSLYALFSQGHVVDIVALTKMK.	RGQAFVIF KELGSSTNALRQLQGFPFYDKPMRIQYAKT			
U1 70K human	-	PNAQGDAFKT LFVARV N.YDTTESKLRREFEV.YGPIKRIHMVYSKRSKGP.	RGYAFIEY EHERDMHSAYKHADGKKIDGRRLVLDVVERG			
pre-rRNP						
		RNP 2	RNP 1			
Nucleolin hamster	1	EGSESTTPFN LFVGNL N.PNKSVAE LKVAISEPFAKNDLA.VVDVRT.GTN.	RKFGYVDF ESAEDLEKALELTGLK.VFGNEIKLEKPKG			
	2	DSKKVRAART LLAKNL .SFNITDELKKEVFEDAL.EIRLVSDQ....GKS.	KGIAYIEF KSEADAENKLEEKQGAIEDGRSVSLYTYGTE			
	3	NSTWSGESKT LVLNL .SYSATEETLQEVFEKATF.IKVPQNOQ....GKS.	KGYAFIEF ASFEDAKEALNSCNKMEIEGRTIRLELQGP			
	4	PNARSQPSKT LFVGNL .SEDTEETLKESEFEG...SVRARIVTDRET.GSS.	KGFGVDF NSEDAKAAKEAMEDEIDGNKVITLDWAKP			
SSB1 yeast	-	HKTVIDPEDT IFVGNV .AHECTEDDLKQLFVEEFGDEVSEIPIKE//SAN.	RGMAFVTF SGENVDIEAKAEFKGKVGFDRELTVDVAV			
mRNP						
PABP yeast	1	SQSVENSSAS LYVGD .EPSVSEAHLYDIFSPI.GSVSSIRVCRDAITKTS.	LGAYVVF NDHEAGRKAIEQLNYTPIKGRLCRIMWSQR			
PABP human	1	PSAPSYPMAS LYVGD .HPDVTKAMLYEKFSFA.GPILSIRVCRDMITRRS.	LGAYVVF QQPADAERALDTMNFVDVYKPKVPRIMWSQR			
PABP yeast	2	PSLRKKGSGN IFIKNL .HPDIDNKALYDTFSV.FGDIILSKKATADEN.GKS.	KGFGVVF EEEGAKEAIDALNGMLLNQGEIYVAPHLS			
PABP human	2	PSLRKSGVGN IFIKNL .DKSIDNKALYDTFSA.FGNILSKVVCDEEN.G.S.	KGYGFVF ETQGAERKAEIKMNGMLNDRKVFVGRFNS			
PABP yeast	3	LEETKAHYTN LYVKN NS.ETTDEQFQELFAK.FGPIVSALEKAD.D.GKL.	KGFGVVF EKHEDAVKAVALNDSELNGEKLYVGRAQK			
PABP human	3	EAELGARAKE FTNVYI .KNFGEDMDDERLKD.L.FGPAISVKVMTDES.GKS.	KGFGVVF ERHEDAQKAVDEMNGKELNGKQIYVGRAQK			
PABP yeast	4	EKMAYQGVN LFVKNL .DDSVDEKLEEEFAP.YGTTISAKVMRTEN.GKS.	KGFGVVF STPEATKAITEKNQVI VAGKPLVVAIAQR			
PABP human	4	DRITRYQGVN LYVKNL .DDGIDDERLKEKFSF.FGTITS AKVM.MEG.GRS.	KGFGVVF SSPEATKAVTEMNGRIVATKPLVYVLAQR			
other						
ABAI-maize	-	MAAADVEYR CFVGG .AWATSNESENENAFAS.YGEILDSKVIITDRETGRS.	RGFGVTF SSENSMLDAIENMNGKELDGRNITVNAQOS			
La human	-	EYKNDVKNRS VYIKGF PTDATLDD.IKE.WLEDKQVNLNIQMRRLH...KAF	KGIIFVVF DSIESAKKFVETPGQKYLETDLLILFKDDY			
Ro human	-	EGRTTKQEPM LFALAI .CSQCSDISTKQA...AFKAVSEVCRPT.....	HLVFIQF KKDLSKMSKCGMW.GRALRKAIAADWYNEKG			
<i>elav</i>	1	QTMTEDEIRS LFSSVG ..EIESVKLIRDKSVQ.VYDPLNPQAP...SKGQS.	LGYGFVVF ..VRPQDAEQAV.N...VLNGLRLQNKTIKV			
<i>drosophila</i>	2	KGAN LYVSG .PKTMTQOELEAIFAP.FGAIITSRI LQNGADNTQT	KGYGFVVF DKREEATRAI...ALNGTTPSSCTDPI			
	3	LASGPGGAYP IFIYNL .APETEAAALWQLFQP.FGAVQSVKIVKDPPTN.QC	KGYGFVVF TNYDEAAMAIR...ALNGYTMGNRVLVQ			
<i>sex-lethal</i>	1	MNDRPRASNTN LIVNYL .PQDMTDRRELYALFRA.IGPINTCRIM....RDY	KTYGSFGY AFVDFTESEMSQRAIKVNLGITVNRKRLKV			
<i>drosophila</i>	2	PGGESIKDTN LYVNL .PRTITDDQLDTIFGK.YGSIVQKNILRDKLTGRP.	RGVAFVVF NKREEAQEAISALNVIPEGGSQPSLVRLA			
<i>tra-2 dros.</i>	-	SREHPQASRC IGVFGL N.TNTSQHKVRELFNK.YGPIERIQMVIDAQTGRS.	RGFCFITF EKLSDARA AKDCSGSIEVDGRRIRVDFSFIT			
RNP CONSENSUS		LFVGNL	E L F F G I K	RGFGVVF	A	L G
		IYIKG	D Y V R	R Y A Y		I

Figure 1. The 90-amino-acid RNA binding domains of RNP-CS-type proteins. The most highly conserved segments of these domains are denoted RNP 1 (RNP-CS) and RNP 2 and are in boldface type, as are other conserved amino acid residues located throughout the domain. Conserved positions listed under RNP consensus, which were detected by pair-wise comparisons using the UWGCG GAP program and further refined by visual inspection, were selected by having no more than two different amino acids in each position in at least 20 out of the 32 domains; only one amino acid appears under RNP consensus when it alone fulfills this criterion. Protein sequences were obtained from the following references: hnRNP A1 from rat (Cobianchi et al. 1986) and *Drosophila* (Haynes et al. 1987); human (Swanson et al. 1987) and *Xenopus* (Preugschat and Wold 1988) hnRNP C; human hnRNP E (UP2) (Lahiri and Thomas 1986); human snRNP A and B' (Sillikens et al. 1987); human snRNP U1 70K (Theissen et al. 1986); hamster pre-rRNP nucleolin (Lapeyre et al. 1987); yeast SSB1 (Jong et al. 1987); yeast (Adam et al. 1986; Sachs et al. 1986) and human (Grange et al. 1987) polyadenylate binding protein (PABP); maize abscisic acid-induced protein (ABAI-maize) (Gomez et al. 1988); human La (Chambers et al. 1988); human Ro (Deutscher et al. 1988); and *Drosophila elav* (Robinow et al. 1988), *Sxl* (Bell et al. 1988), and *tra-2* (Amrein et al. 1988).

amino acids (including the RNA-binding domain), and studies using RNP-CS-containing peptide fragments of nucleolin, indicate that this domain is sufficient to retain the specific RNA-binding properties of the native protein (C. Burd, E. Mortenson, and G. Dreyfuss, unpubl.; Bugler et al. 1987, respectively). Taken together, these data indicate that the RNP-CS and RNP 2 segments are integral components of an RNA-binding domain. The contribution of particular residues in the RNA-binding domain to RNA binding is not yet known. It is even possible that in some cases only part of the RNA-binding domain is required for binding (Sachs et al. 1987). Nevertheless, the RNP-CS is a strong indicator of RNA-binding function.

Modular structure of RNP-CS-type proteins

Several RNP-CS-type RNA-binding domains often occur in an individual protein (Fig. 1) and usually are located in the amino-terminal portion of the protein. Examination of these repeats in corresponding proteins from divergent organisms, especially with regard to evolutionary relationships, has implications for the possible activities of these proteins. Comparison of the RNA-binding domains within a single protein containing multiple RNA-binding domains typically shows only 30% amino acid identity. However, comparison of the corresponding domain in the same protein from divergent species (e.g., alignment of domain 1 of the PABPs from yeast and

man) demonstrates strong amino acid similarities, typically greater than 60% identity (Dreyfuss et al. 1988). Even more striking, the corresponding domains of the yeast and human PABPs contain segments of 8–10 amino acids that are completely identical (see pairwise comparison, Fig. 4 in Dreyfuss et al. 1988). These segments are characteristic of the individual RNA-binding domain alone, and they are most likely indicative of functional specialization. In the A1 protein from mammals (rat) and *Drosophila*, domain 1 has an amino acid segment downstream of the RNP-CS in which 15 out of 16 amino acids are identical (Fig. 1, underlined region); overall these corresponding domains share a 65% identity (Dreyfuss et al. 1988). A schematic model depicting the modular assembly of RNP-CS-type proteins is presented in Figure 2. The patterns of evolutionary conservation among these domains suggest that they evolved by duplication and diversification of a common, ancestral RNA-binding protein gene bearing the prototypical RNP-CS. Subsequently, each RNA-

binding domain, for example in the PABP or in the hnRNP protein A1, evolved independently. The diversification of the RNA-binding domains is particularly evident in isoforms of the A1 protein, where point mutations cause conservative amino acid substitutions in the first RNA-binding domain of this protein (Buvoli et al. 1988).

The remainder of a RNP-CS-type protein is characterized typically by a distinctive auxiliary domain(s). For example, the auxiliary domain is glycine-rich in the hnRNP protein A1, proline-rich in the polyadenylate binding protein (PABP), and glutamic and aspartic acid-rich in the hnRNP C proteins. The abundant nucleolar pre-rRNA-binding protein, nucleolin, has a glycine-rich domain similar to that of the hnRNP proteins A1 and A2 (Kumar et al. 1986) and an acidic domain similar to that of the hnRNP C proteins (Lapeyre et al. 1987). These auxiliary domains appear likely to function primarily in protein–protein interactions, since all of these proteins are components of multiprotein ribonucleoprotein com-

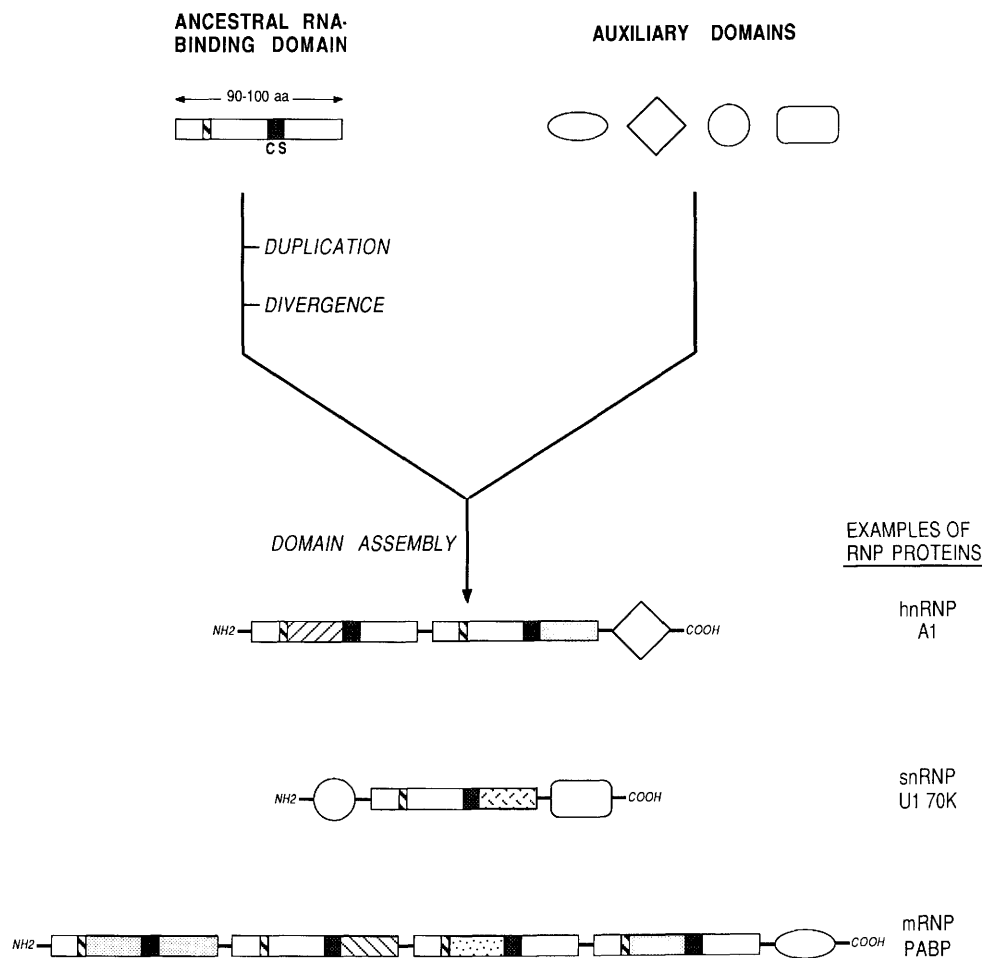


Figure 2. Schematic model of the evolution and modular assembly of RNP-CS-type RNA-binding proteins. It is postulated that a prototype RNP-CS protein gene, encoding the ancestral RNA-binding domain protein, was subject to duplication and divergence, thereby expanding the range of RNA-binding specificities of RNP-CS-type proteins. The individual character of the diverged RNA-binding domains is represented by slanted lines or by stippled areas, whereas the conserved RNP-CS (RNP 1) is shown by the black box and RNP 2 by a smaller slashed box. In addition, auxiliary domains, composed of a distinctive set of amino acids, contribute to the domain character of RNP-CS-type proteins.

plexes. However, the auxiliary domains may also influence the RNA-binding activity of the protein. A comparison of the binding of A1 to single-stranded (ss) DNA and RNA with that of UP1, the truncated and auxiliary domain-less derivative of A1, suggests that the auxiliary domains may also modify the polynucleotide binding properties of RNP-CS-type RNA-binding domains (Kumar et al. 1986). This combination of different RNA-binding domains with auxiliary domains imparts a modular structure to RNP-CS-type proteins.

RNA-binding specificity

Several of the proteins listed in Figure 1 have binding specificity for certain RNA sequences. For example, the PABP binds preferentially to the poly(A) segments of mRNA. Recently, several of the hnRNP proteins, A1, C, and D have been shown to have RNA-binding specificity. The amino acid sequence of the hnRNP D proteins has not been determined yet, but the hnRNP A1 and C proteins, as well as the mRNP PABP, contain RNP-CS-type RNA-binding domains. Experimental efforts to purify individual hnRNP proteins led to the observation that several of these proteins display very high binding affinities for specific ribonucleotide homopolymers (Swanson and Dreyfuss 1988a). For example, the hnRNP C proteins have a striking affinity for poly(U), since they remain bound to this homopolymer at high salt concentrations (2 M NaCl). As reference, the yeast cytoplasmic PABP remains selectively bound to poly(A) in vitro only up to 1 M NaCl.

The finding that hnRNP C proteins are poly(U)-binding proteins led to the prediction that they would compete effectively for binding to the uridine-rich sequences often found in the polypyrimidine stretches near the 3' end of introns and downstream of the cleavage site for polyadenylation in pre-mRNAs. Moreover, evidence that suggested a role for hnRNP proteins in pre-mRNA splicing has been described (Choi et al. 1986; Sierakowska et al. 1986). Experiments using an in vitro RNA-binding assay demonstrated that three hnRNP proteins, A1, C, and D (Piñol Roma et al. 1988), show intrinsic binding specificities on authentic pre-mRNAs (Swanson and Dreyfuss 1988b). These proteins bind specifically to the 3' end of introns within a region containing the conserved polypyrimidine stretch between the branch site and the 3' splice site. The hnRNP C proteins also were found to cross-link by UV light preferentially to uridine-rich sequences downstream of the polyadenylation cleavage site on two different viral pre-mRNAs (Wilusz et al. 1988; Moore et al. 1988); this binding was sensitive to mutations that interfered with aspects of the in vitro cleavage-polyadenylation reaction (Wilusz et al. 1988).

In sum, contrary to an earlier view that hnRNP proteins coat pre-mRNA in a non-sequence-specific manner (Pullman and Martin 1983; Thomas et al. 1983; Wilk et al. 1983; Conway et al. 1988), these experiments demonstrate that several RNP-CS-containing hnRNP proteins do have RNA-binding specificities and can bind to sites

on pre-mRNAs that are critical for the formation of mature and functional mRNA. Since the A1, C, and D proteins are confined to the nucleus, and appear to have the highest affinity for pre-mRNA sequences that are removed during the formation of mRNA, it is possible that these pre-mRNA segments constitute nuclear retention sequences (Swanson and Dreyfuss 1988a,b; Wilusz et al. 1988).

Developmental loci in *Drosophila* encode RNP-CS-type proteins

Several genes that appear to have critical roles in establishing certain developmental pathways in *Drosophila* and maize recently have been cloned molecularly and their sequences have predicted protein products that contain RNP-CS-type RNA-binding domains (Amrein et al. 1988; Bell et al. 1988; Gómez et al. 1988; Robinow et al. 1988; Goralski et al. 1989). In maize, the plant hormone abscisic acid induces in the embryo an RNP-CS containing protein whose function is unknown (Gómez et al. 1988; Mortenson and Dreyfuss 1989). The *Drosophila* genes, *embryonic lethal abnormal visual system (elav)*, *sex-lethal (Sxl)*, and *transformer-2 (tra-2)*, are critical in the early determination, as well as in the maintenance, of two terminally differentiated states, the neuron and the somatic sexual state. Although it has not been shown yet that these gene products are RNA-binding proteins in vivo, considerable genetic and molecular data suggest that these loci exert their regulatory effects at the post-transcriptional level. Indeed, it appears that the entire sex-determination pathway in *Drosophila* is regulated at the level of sex-specific splicing (Boggs et al. 1987; Baker and Wolfner 1988; Nagoshi et al. 1988). *Sxl* positively regulates the activity of the feminizing gene, *transformer (tra)*; the protein product of the *Sxl* locus is necessary to produce a correctly spliced, functional *tra* mRNA in females (Cline 1979; McKeown et al. 1988; Nagoshi et al. 1988). Although *Drosophila* males and females express *Sxl* transcripts, only females encode a mRNA capable of producing a full-length, functional polypeptide; male transcripts, on the other hand, harbor an exon with an in-frame translation termination codon. Bell et al. (1988) propose that *Sxl* transcripts are most likely subject to an autoregulatory, alternative splicing mechanism that maintains the appropriate sexually differentiated state, producing functional *Sxl* protein and correctly spliced *tra* mRNA in females only.

Role in alternative splicing

Considering that RNP-CS-containing proteins such as *Sxl*, *tra-2*, *elav*, and the abscisic acid-induced maize protein are most probably RNA-binding proteins, they could have any one of many possible functions in RNA metabolism. However, as the foregoing discussion suggests, particularly for *Sxl* and *tra-2*, a role in alternative splicing is likely and interesting models on how this might be accomplished by such proteins can be envi-

sioned. The occurrence of multiple, nonidentical RNA-binding domains in a single protein suggests that, like any macromolecule that has multiple ligand binding sites, RNP-CS-type RNA-binding proteins may be able to bind to multiple segments of RNA, with each domain possibly binding independently of the others. Together with the finding that RNP-CS-containing proteins can have RNA binding specificity, it appears that such proteins have the potential to bring into proximity segments of RNA that are otherwise far apart on the same RNA molecule or even on different RNA molecules. In this context, the binding preference of some hnRNP proteins for sites on the pre-mRNA that are important for pre-mRNA processing—the 3' splice site and sequences downstream from the polyadenylation signal—is particularly tantalizing. An extension of these findings is the possibility that such proteins may play a role in pre-mRNA splicing, and if their modular structure contains the appropriate combination of binding specificities, also in alternative splicing. RNP-CS-type proteins potentially may affect alternative splicing by negative regulation as well. In this scheme, a protein with even a single RNA-binding domain could compete with other nuclear factors for a specific set of 3' splice sites, and it could contain an auxiliary domain(s) that prevents, rather than promotes, splicing complex formation at that site.

Post-transcriptional control mechanisms, including a myriad alternative splicing events, appear to operate in a growing number of cellular processes and pathways (for review, see Kozak 1988). As the above discussion suggests, RNP-CS-type proteins are excellent candidates for a novel class of regulatory proteins affecting these mechanisms. To date, the best characterized examples of alternative splicing pathways that have major developmental consequences are in *Drosophila*, among them *Sxl* and *tra-2*. It is tempting to speculate that the protein product of *Sxl*, an RNP-CS-type protein with multiple RNA-binding domains, effects sex-specific alternative splicing by binding to different sets of sequences on an individual pre-mRNA, consequently promoting or repressing the selection of splice sites. By being expressed appropriately and by using the potential for sequence preference within its multiple, independent RNA-binding domains, a regulator protein such as the *Sxl* gene product could influence the production of novel mRNAs significantly.

Similarity to transcription factors

Finally, the emerging picture of the structure and function of RNP-CS-type RNA-binding proteins bears an intriguing similarity to some characteristics of transcription factors. The binding of RNP-CS-type proteins to RNA may not require recognition of a strict consensus binding site because they bind to the 3' end of introns and to sequences downstream of the polyadenylation signal that vary in nucleotide sequence. The binding of OBP100, the enhancer-octamer DNA-binding protein, to degenerate consensus binding sites is reminiscent of this

situation (Baumruker et al. 1988). In addition, one consequence of the hypothetical role of RNP-CS-type proteins in alternative splicing is that differential sequence binding by such proteins on a RNA molecule could lead to a looping out of RNA between these binding sites. This would be similar to the ability of a well-characterized specific DNA-binding protein, phage λ repressor, to bind cooperatively at distal, nonadjacent sites upstream of operator DNA in the λ genome, with the result that DNA in between is looped out. The inhibition of transcriptional activity that follows recently was shown to be a direct consequence of distal repressor-repressor interaction in vitro and in vivo (Hochschild and Ptashne 1988). Finally, the auxiliary domains of several of the RNP-CS-type RNA-binding proteins are similar in amino acid character to some activation domains of eukaryotic transcription factors. For example, there are similarities in amino acid character between the highly acidic domains of the hnRNP C proteins and the transcription factor GCN 4 (Hope and Struhl 1986), the glutamine-rich amino-terminal domain of *elav* and SP1 (Courey and Tjian 1988), and the proline-rich domains of the PABP and CTF/NF1 (Bohmann et al. 1987).

If the proteins of the *Drosophila elav* and *Sxl* loci are shown, as expected, to be bona fide RNA-binding proteins, a new and experimentally tractable function of RNP-CS-type proteins will have been uncovered. It remains an exciting possibility that the repertoire of developmental regulators now has to be expanded to include RNA-binding proteins.

Acknowledgments

We thank Drs. Robert Lamb, Daniel Linzer, and Richard Morimoto, and members of our laboratory for comments on the manuscript. This work was supported by grants (GM-31888 and GM-32175) from the National Institutes of Health. R.J.B. is postdoctoral fellow of the Anna Fuller Fund. G.D. is an Established Investigator of The American Heart Association.

References

- Adam, S.A., T.Y. Nakagawa, M.S. Swanson, T. Woodruff, and G. Dreyfuss. 1986. mRNA polyadenylate-binding protein: Gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol. Cell Biol.* 6: 2932–2943.
- Amrein, H., M. Gorman, and R. Nöthinger. 1988. The sex-determining gene *tra-2* of *Drosophila* encodes a putative RNA-binding protein. *Cell* 55: 1025–1035.
- Baker, B.S. and M.F. Wolfner. 1988. A molecular analysis of *doublesex*, a bifunctional gene that controls both male and female sexual differentiation in *Drosophila melanogaster*. *Genes Dev.* 2: 477–489.
- Baumruker, T., R. Sturm, and W. Herr. 1988. OBP 100 binds remarkably degenerate octamer motifs through specific interactions with flanking sequences. *Genes Dev.* 2: 1400–1413.
- Bell, L.R., E.M. Maine, P. Schedl, and T.W. Cline. 1988. *Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* 55: 1037–1046.

- Boggs, R.T., P. Gregor, S. Idriss, J.M. Belote, and M. McKeown. 1987. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the *transformer* gene. *Cell* **50**: 739–747.
- Bohmann, D., T.J. Bos, A. Admon, T. Nishimura, P.K. Vogt, and R. Tjian. 1987. Human proto-oncogene *c-jun* encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. *Science* **238**: 1386–1392.
- Bugler, B., H. Bourbon, B. Lapeyre, M.O. Wallace, J.-H. Chang, F. Amalric, and M.O.J. Olson. 1987. RNA binding fragments from nucleolin contain the ribonucleoprotein consensus sequence. *J. Biol. Chem.* **262**: 10922–10925.
- Buvoli, M., G. Biamanti, P. Tsoulfas, M.T. Bassi, A. Ghetti, S. Riva, and C. Morandi. 1988. cDNA cloning of human hnRNP protein A1 reveals the existence of multiple mRNA isoforms. *Nucleic Acids Res.* **16**: 3751–3770.
- Chambers, J.C., D. Kenan, B.J. Martin, and J.D. Keene. 1988. Genomic structure and amino acid sequence domains of the human La autoantigen. *J. Biol. Chem.* **263**: 18043–18051.
- Choi, Y.D., P.J. Grabowski, P.A. Sharp, and G. Dreyfuss. 1986. Heterogeneous nuclear ribonucleoproteins: Role in RNA splicing. *Science* **231**: 1534–1539.
- Cline, T.W. 1979. A male-specific lethal mutation in *Drosophila melanogaster* that transforms sex. *Dev. Biol.* **72**: 266–275.
- Cobianchi, R., D.N. SenGupta, B.Z. Zmudzka, and S.H. Wilson. 1986. Structure of rodent helix-destabilizing protein revealed by cDNA cloning. *J. Biol. Chem.* **261**: 3536–3543.
- Conway, G., J. Wooley, T. Bibring, and W.M. LeSturgeon. 1988. Ribonucleoproteins package 700 nucleotides of pre-mRNA into a repeating array of regular particles. *Mol. Cell. Biol.* **8**: 2884–2895.
- Courey, A.J. and R. Tjian. 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* **55**: 887–898.
- Deutscher, S.L., J.B. Harley, and J.D. Keene. 1988. Molecular analysis of the 60-kDa human Ro ribonucleoprotein. *Proc. Natl. Acad. Sci.* **85**: 9479–9483.
- Dreyfuss, G. 1986. Structure and function of nuclear and cytoplasmic ribonucleoprotein particles. *Annu. Rev. Cell Biol.* **3**: 459–498.
- Dreyfuss, G., M.S. Swanson, and S. Piñol-Roma. 1988. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Sci.* **13**: 86–91.
- Gómez, J., D. Sánchez-Martínez, V. Stiefel, J. Rigau, P. Puigdomènech, and M. Pagès. 1988. A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. *Nature* **334**: 262–264.
- Goralski, T.I., J.-E. Edström, and B.S. Baker. 1989. The sex determination locus *transformer-2* of *Drosophila* encodes a polypeptide with similarity to RNA binding proteins. *Cell* (in press).
- Grange, T., C. Martins de Sa, J. Oddos, and R. Pictet. 1987. Human mRNA polyadenylate-binding protein: evolutionary conservation of a nucleic acid binding motif. *Nucleic Acids Res.* **15**: 4771–4787.
- Haynes, S.R., M.L. Rebbert, B.A. Mozer, F. Forquignon, and I.G. Dawid. 1987. *pen* repeat sequences are GGN clusters and encode a glycine-rich domain in *Drosophila* homologous to the rat helix destabilizing protein. *Proc. Natl. Acad. Sci.* **84**: 1819–1823.
- Helmann, J.D. and M.J. Chamberlin. 1988. Structure and function of bacterial σ factors. *Annu. Rev. Biochem.* **57**: 839–872.
- Herrick, G. and B. Alberts. 1976. Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus. *J. Biol. Chem.* **251**: 2124–2132.
- Hochschild, A. and M. Ptashne. 1988. Interaction at a distance between repressors disrupts gene activation. *Nature* **336**: 353–357.
- Hope, I.A. and K. Struhl. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**: 885–894.
- Jong, A.Y.-S., M.W. Clark, M. Gilbert, A. Oehm, and J.L. Campbell. 1987. *Saccharomyces cerevisiae* SSB1 protein and its relationship to nucleolar RNA-binding proteins. *Mol. Cell. Biol.* **7**: 2947–2955.
- Kozak, M. 1988. A profusion of controls. *J. Cell Biol.* **107**: 1–7.
- Kumar, A., K.R. Williams, and W. Szer. 1986. Purification and domain structure of core hnRNP proteins A1 and A2 and their relationship to single-stranded DNA-binding proteins. *J. Biol. Chem.* **261**: 11266–11273.
- Lahiri, D.K. and J.O. Thomas. 1986. A cDNA clone of the hnRNP C proteins and its homology with the single-stranded DNA binding protein UP2. *Nucleic Acids Res.* **14**: 4077–4094.
- Lapeyre, B., H. Bourbon, and F. Amalric. 1987. Nucleolin, the major nucleolar protein of growing eukaryotic cells: An unusual protein structure revealed by the nucleotide sequence. *Proc. Natl. Acad. Sci.* **84**: 1472–1476.
- McKeown, M., J.M. Belote, and R.T. Boggs. 1988. Ectopic expression of the female *transformer* gene product leads to female differentiation of chromosomally male *Drosophila*. *Cell* **53**: 887–895.
- Merrill, B.M., K.L. Stone, F. Cobianchi, S.H. Wilson, and K.R. Williams. 1988. Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 hnRNP protein. *J. Biol. Chem.* **263**: 3307–3313.
- Moore, C.L., J. Chen, and J. Whoriskey. 1988. Two proteins crosslinked to RNA containing the adenovirus L3 poly(A) site require the AAUAAA sequence for binding. *EMBO J.* **7**: 3159–3169.
- Mortenson, E. and G. Dreyfuss. 1989. RNP in maize protein. *Nature* **337**: 312.
- Nagoshi, R.N., M. McKeown, K.C. Burtis, J.M. Belote, and B.S. Baker. 1988. The control of alternative splicing at genes regulating sexual differentiation of *D. melanogaster*. *Cell* **53**: 229–236.
- Piñol-Roma, S., Y.D. Choi, M.J. Matunis, and G. Dreyfuss. 1988. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev.* **2**: 215–227.
- Pullman, J.M. and T.E. Martin. 1983. Reconstitution of nucleoprotein complexes with mammalian heterogeneous nuclear ribonucleoprotein (hnRNP) core proteins. *J. Cell Biol.* **97**: 99–111.
- Preugschat, F. and B. Wold. 1988. Isolation and characterization of a *Xenopus laevis* C protein cDNA: Structure and expression of a heterogeneous nuclear ribonucleoprotein core protein. *Proc. Natl. Acad. Sci.* **85**: 9669–9673.
- Riva, S., C. Morandi, P. Tsoulfas, M. Pandolfo, G. Biamonti, B. Merrill, K.R. Williams, G. Multhaup, K. Beyreuther, H. Werr, B. Henrich, and K.P. Schäfer. 1986. Mammalian single-stranded DNA binding protein UP1 is derived from the hnRNP core protein A1. *EMBO J.* **5**: 2267–2273.
- Robinow, S., A.R. Campos, K.-M. Yao, and K. White. 1988. The *elav* gene product of *Drosophila*, required in neurons, has three RNP consensus motifs. *Science* **242**: 1570–1572.
- Sachs, A.B., W.M. Bond, and R.D. Kornberg. 1986. A single gene from yeast for both nuclear and cytoplasmic polyadenylate-

- binding protein: Domain structure and expression. *Cell* **45**: 827–835.
- Sachs, A.B., R.W. Davis, and R. Kornberg. 1987. A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. *Mol. Cell. Biol.* **7**: 3268–3276.
- Sierakowska, H., W. Szer, P.J. Furdon, and R. Kole. 1986. Antibodies to hnRNP core proteins inhibit in vitro splicing of human β globin pre-mRNA. *Nucleic Acids Res.* **14**: 5241–5254.
- Sillekens, P.T.G., W.J. Habets, R.P. Beijer, and W.J. van Venrooij. 1987. cDNA cloning of the human U1 snRNA-associated A protein: Extensive homology between U1 and U2 snRNP-specific proteins. *EMBO J.* **6**: 3841–3848.
- Swanson, M.S. and G. Dreyfuss. 1988a. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol. Cell. Biol.* **8**: 2237–2241.
- . 1988b. RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO J.* **11**: 3519–3529.
- Swanson, M.S., T.Y. Nakagawa, K. LeVan, and G. Dreyfuss. 1987. Primary structure of human nuclear ribonucleoprotein particle C proteins: Conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA and pre-rRNA-binding proteins. *Mol. Cell. Biol.* **7**: 1731–1739.
- Theissen, H., M. Etzerodt, R. Reuter, C. Schneider, F. Lottspeich, P. Argos, R. Lührmann, and L. Philipson. 1986. Cloning of the human cDNA for the U1 RNA-associated 70K protein. *EMBO J.* **5**: 3209–3217.
- Thomas, J.O., S.K. Glowacka, and W. Szer. 1983. Structure of complexes between a major protein of heterogeneous nuclear ribonucleoprotein particles and polyribonucleotides. *J. Mol. Biol.* **171**: 439–455.
- Wilk, H.-E., G. Angeli, and K.P. Schäfer. 1983. In vitro reconstitution of 35S ribonucleoprotein complexes. *Biochemistry* **22**: 4592–4600.
- Wilusz, J., D.I. Feig, and T. Shenk. 1988. The C proteins of heterogeneous nuclear ribonucleoprotein complexes interact with RNA sequences downstream of polyadenylation cleavage sites. *Mol. Cell. Biol.* **8**: 4477–4483.



RNA-binding proteins as developmental regulators.

R J Bandziulis, M S Swanson and G Dreyfuss

Genes Dev. 1989, **3**:

Access the most recent version at doi:[10.1101/gad.3.4.431](https://doi.org/10.1101/gad.3.4.431)

References

This article cites 51 articles, 25 of which can be accessed free at:
<http://genesdev.cshlp.org/content/3/4/431.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Dharmacon[™] Reagents
Custom synthesis, RNAi, and CRISPR solutions

Infinite
Reliability

More

horizon
a PerkinElmer company