HTLV-1-induced proliferative diseases induced by the transactivating Tax protein. □

Received 17 May; accepted 4 August 1989.


1. Poiesz, B. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77,** 7415–7419 (1980).
2. Robert-Guroff, M. *et al. Science* **215,** 975–982 (1982).
3. Yochida, M., Miyoshi, I. & Hinuma, Y. *Proc. natn. Acad. Sci. U.S.A.* **79,** 2031–2035 (1982).
4. Gessian, A. *et al. Lancet,* 407–410 (1985).
5. Rodgers-Johnson, P. *et al. Lancet ii,* 1247–1248 (1985).
6. Jacobson, S., Raine, C. S., Mingioli, E. S. & McFarlin, D. E. *Nature* **331,** 540–543 (1988).
7. Reddy, P. E. *et al. Science* **243,** 529–533 (1989).
8. Nerenberg, M., Hinrichs, S. H., Reynolds, R. K., Khoury, G. & Jay, G. *Science* **237,** 1324–1329 (1987).
9. Hinrichs, S. H., Nerenberg, M., Reynolds, R. K., Khoury, G. & Jay, G. *Science* **237,** 1340–1343 (1987).
10. Bloch, K. J., Buchanan, W. W., Wohl, M. J. & Bunim, J. J. *Medicine* **44**(3), 187–231 (1965).
11. Moutsopoulos, H. M., *et al. Ann. intern. Med.* **92**(1), 212–226 (1980).
12. Daniels, T. E. & Talal, N., in *Sjögren's Syndrome, Clinical and Immunologial Aspects* (eds Talal, N., Montsopoulos, H. M. & Kassan, S. S.) 193–199 (Springer, New York, 1987).
13. Anderson, L. G. & Talal, N. *Clin. expl. Immun.* **10,** 199–219 (1972).
14. Vernant, J. C. *et al. La Presse Medicale* **15**(9), 419–422 (1986).
15. Vernant, J.-C. *et al. Lancet,* 177 (1988).
16. Smith, F. B., Rajdeo, H., Panesar, N., Bhuta, K. & Stahl, R. *Arch. Pathol. lab. Med.* **112,** 742–745 (1988).
17. Robert-Guroff, M. *et al. J. Am. Med. Assoc.* **255,** 3133–3137 (1986).
18. Gordon, J. J., Golbus, J. & Kurtides, E. S. *New Engl. J. Med.* **311**(22), 1441–1442 (1984).
19. Couderc, L.-J. *et al. Arch. intern. Med.* **147,** 898–901 (1987).
20. Ulirsch, R. C. & Jaffe, E. S. *Hum. Path.* **18**(10), 1063–1068 (1987).
21. Batsakis, J. G. & Howard, D. R. *Clin. Lab. Ann.* **1,** 178–188 (1982).
22. Konttinen, Y. T., Segerberg-Konttinen, M., Bergroth, O. & Malmstrom, M. *Scand. J. Rheum.* **61,** 61–66 (1986).
23. Hsu, S. M., Raine, L. & Fanger, J. *J. Histochem. Cytochem.* **29,** 576–582 (1981).
24. Hinrichs, S. H. *et al., Cancer Genet. Cytogenet.* **8,** 19–23 (1985).
25. Towbin, H., Staehelin, J. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76,** 4350–4354 (1979).


ACKNOWLEDGEMENTS. We thank S. Oroszlan for providing antibody to the Tax protein and Ms Judie Ireland and Ms Maureen Preston for preparation of this manuscript. This study was supported in part by a grant from the N.I.H. (S.H.H.).

# Changing Fos oncoprotein to a Jun-independent DNA-binding protein with GCN4 dimerization specificity by swapping 'leucine zippers'

**Joan W. Sellers & Kevin Struhl**

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Massachusetts 02115, USA

**A STRUCTURAL motif for DNA-binding proteins, the 'leucine zipper', has been proposed for the *jun*, *fos* and *myc* gene products, the yeast transcriptional activator GCN4, and the C/EBP enhancer-binding protein[1]. These proteins all contain a region with four or five leucine residues spaced exactly seven amino acid residues apart whose sequence is consistent with the formation of an amphipathic α-helix. It has been proposed that the leucine zipper consists of two interdigitated α-helices, one from each monomer, that constitute the dimerization function necessary for high-affinity binding to DNA; an adjacent region of basic residues is thought to be responsible for specific protein–DNA contacts[1]. In support of this model, substitution of the leucine residues within the motif can abolish dimerization and DNA-binding[2–6], and a synthetic peptide corresponding to the GCN4 leucine zipper forms α-helical dimers[7]. Despite the conserved leucine residues, however, each protein has a distinct dimerization specificity. Specifically, GCN4 homodimer, Jun homodimer and Fos–Jun heterodimer proteins bind to the same DNA site, whereas Fos is unable to form homodimers, bind DNA, or interact with GCN4 (refs 8–14). Here, we alter the dimerization specificity of Fos by precisely replacing its leucine zipper with that from GCN4. This Fos-GCN4 chimaeric protein is able to bind to the target site in the absence of Jun, and can form DNA-binding heterodimers with GCN4 but not with Jun. These results indicate that the leucine zipper is sufficient to confer dimerization specificity and strongly suggest that Fos contacts DNA directly.**

The products of the nuclear proto-oncogenes *fos* and *jun* form a heterodimeric complex which binds to the transcriptional regulatory element known as the AP-1 site[2,4–6,12–14]. Yeast GCN4 protein binds as a homodimer to target sequences that are indistinguishable from AP-1 sites[8,9]. Extensive analysis indicates that the optimal GCN4 binding site is dyad-symmetric and that each GCN4 monomer directly contacts a half-site[15,16]. By analogy, each subunit of a Fos–Jun heterodimer would be expected to contact a half-site. This idea has been supported by the finding that Fos, Jun, and GCN4 contain a conserved stretch of basic residues equally spaced from their leucine zippers, and mutations in the Fos and Jun basic region prevent DNA-binding, but not dimerization[2,4–6]. Although the failure of Fos to bind DNA could reflect its inability to form homodimers, its contribution to specific DNA-binding could also be by an indirect effect mediated through Jun.

To determine whether the leucine zipper confers dimerization specificity and Fos has the potential to contact DNA, we created $Fos^G$, a chimaeric protein in which Fos residues 23–160 (including the basic region) are fused to the carboxy-terminal 33 amino acids of GCN4 (containing the leucine zipper, but unable to bind DNA; see ref. 17) (Fig. 1). The spacing of the leucine zipper and the basic domain in $Fos^G$ is identical to that in Fos and GCN4. The LexA domain at the amino-terminus of the protein lacks the dimerization function[18,19] and is irrelevant to these experiments (see below).

To test the DNA-binding properties of $Fos^G$, $^{35}S$-labelled $Fos^G$ was synthesized *in vitro*, incubated with the optimal GCN4/AP-1 site (ATGACTCAT), and electrophoresed in non-denaturing polyacrylamide gels (Fig. 2*a*). A band is seen which is dependent on the addition of both the protein and DNA, indicating that $Fos^G$ can bind to this DNA fragment. We tested the specificity of binding with a DNA containing a point mutation, AGGACTCAT; this mutation leads to a significant reduction in binding by GCN4 and by the Jun–Fos complex[9,10,15]. Complex formation by $Fos^G$ was also greatly reduced by this point mutation, indicating that the chimaeric protein has similar DNA sequence recognition properties to those of GCN4 and the Jun–Fos complex.

The fact that $Fos^G$ binds to the dyad-symmetric GCN4/AP-1 site suggests that it can form homodimers, presumably mediated by the GCN4 leucine zipper. If the leucine zipper confers dimerization specificity, then $Fos^G$ should form heterodimers with GCN4 but not with Jun. Indeed, when we synthesized GCN4 and $Fos^G$ together and incubated them with the target DNA, we observed a new protein–DNA complex with an intermediate mobility (Fig. 2*b*) indicative of heterodimer formation. The $Fos^G$–GCN4 heterodimer and the GCN4 homodimer bands were of equal intensity, but the presumed $Fos^G$ homodimer band was much fainter. This observation could reflect a decrease in either the dimerization ability or in the affinity for target DNA of the $Fos^G$ homodimer compared with the heterodimer; we favour the latter explanation. In any event, our results exclude the possibility that dimerization by the LexA domain influences binding by $Fos^G$, because although the heterodimer contains only one LexA moiety, it functions more effectively than the homodimer.

Additional observations indicate that $Fos^G$ is a derivative of Fos with an altered dimerization specificity (Fig. 2*c*). First, when $Fos^G$ and a Jun derivative were co-synthesized, only the $Fos^G$ complex was seen, indicating that $Fos^G$ and Jun cannot form heterodimers. The failure to observe new bands was not due to the co-migration of heterodimer and homodimer complexes, because the relative molecular masses of $Fos^G$ and the Jun derivative differ (under our conditions, the Jun homodimer complex cannot be seen[13]). Second, unlike $Fos^G$, a Fos derivative does not bind DNA or form a heterodimer complex after co-synthesis with GCN4, but it will form a DNA-binding complex when co-synthesized with Jun (refs 2 and 12–14; data not shown).

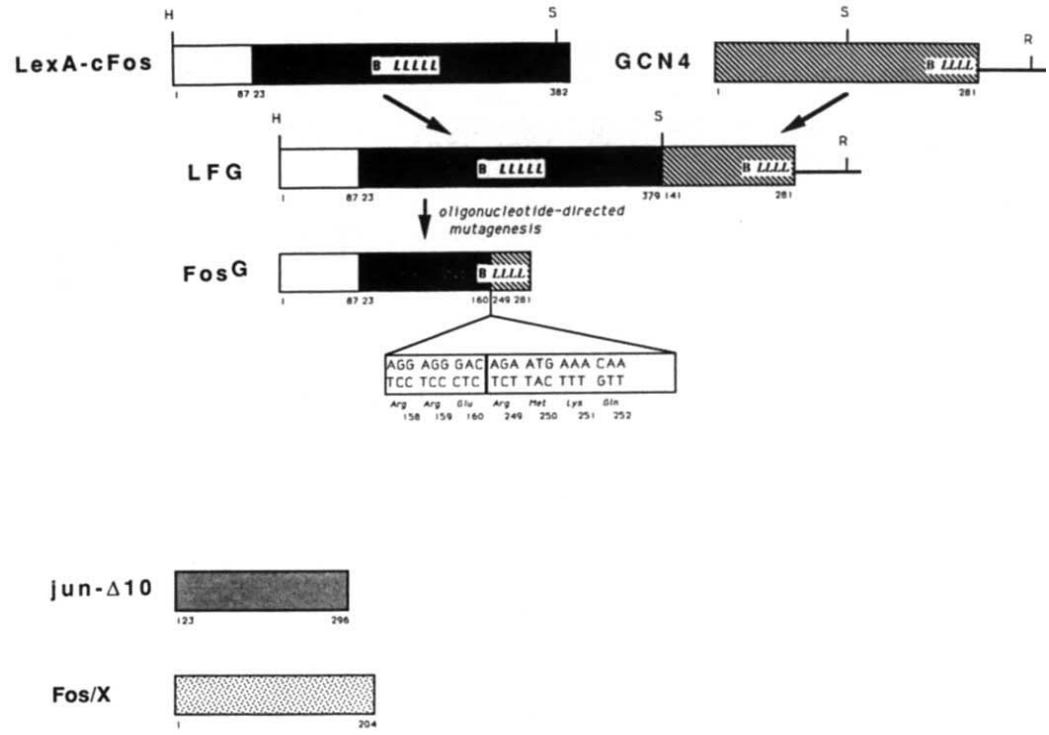


FIG. 1 Structure of proteins. Top line, structure of LexA–c-Fos, which is composed of the LexA DNA-binding domain (open box) fused to the Fos oncoprotein (black box), and the structure of GCN4 (hatched box). Leucine zippers, L; adjacent basic regions, B; H, R and S indicate restriction sites (H, *Hin*dIII site; S, *Sal*I site; R, *Eco*RI site) used in the cloning of LFG, which is shown in the second line from the top. (The wild-type *GCN4* gene does not contain a *Sal*I site; the site shown occurs in *LexA–gcn4–Δ29* (ref. 20), a deleted derivative used in the construction of LFG.) Shown below LFG is Fos$^G$, with the sequence of the junction between Fos$^G$ and GCN4 indicated. Jun–Δ10 is a deleted Jun derivative. The Fos derivative, Fos/X, was obtained from T. Halazonetis[13]. Numbers below protein structures indicate amino-acid residues. The thick line represents non-coding sequences.

METHODS. The DNA LFG was constructed by joining the *Hin*dIII–*Sal*I fragment of the gene encoding LexA–c-Fos (ref. 21) with the *Sal*I–*Eco*RI fragment of *GCN4* (the *Eco*RI site is downstream of the structural gene) and cloning the resulting fusion into *Hin*dIII/*Eco*RI-digested YCm90, a vector derived from YCp88 (ref. 17) that contain an SP6 promoter and an M13 origin of replication. The DNA molecule encoding the chimaeric Fos$^G$ protein was generated from LFG by oligonucleotide-directed mutagenesis using uracil-substituted single-stranded templates[22]; the junction between *fos* and *GCN4* was confirmed by sequencing. The *jun* deriative, *jun–Δ10* was generated by cloning the *Sal*I–*Eco*RI fragment of *LexA–jun–Δ5* (ref. 23) into pSP64, a vector containing the SP6 promoter.

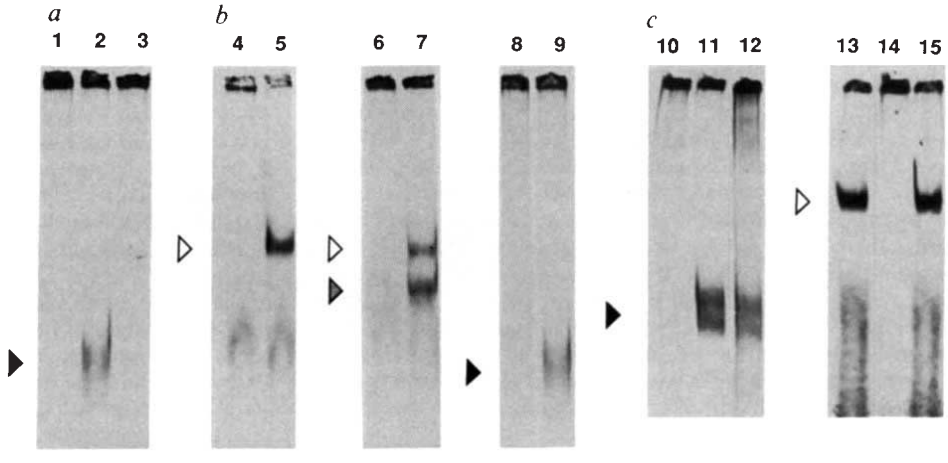


FIG. 2 Specificity of complex formation. *a*, $^{35}$S-labelled, *in vitro*-translated Fos$^G$ protein was tested for ability to bind to the GCN4/AP-1 site in the absence of DNA (lane 1), in the presence of a 631-base pair DNA fragment containing the optimal GCN4/AP-1 site ATGACTCAT (lane 2), or in the presence of a similar fragment containing the point mutation AGGACTCAT (lane 3). A DNA–protein complex is observed only in the reaction containing the optimal GCN4/AP-1 site (lane 2, black arrow). *b*, To test for heterodimer formation, GCN4 (lane 5), Fos$^G$ (lane 9) or a co-synthesis of both proteins (lane 7), were each incubated with the 631-base pair fragment containing the optimal GCN4/AP-1 binding site. When synthesized alone, GCN4 (lane 5, open arrow) and Fos$^G$ (lane 9, black arrow) each form complexes. When synthesized together, a new complex of intermediate mobility is seen (lane 7, stippled arrow; see text). In the absence of DNA, no complexes are evident (lanes 4, 6 and 8). The diffuse band at the bottom of lanes 4 and 5 represents a minor artefactual translation product made only in syntheses of full-length GCN4 protein; it is unrelated to any DNA-binding phenomenon because it is found in the absence of added DNA (lane 4; refs 8 and 17). Because this artefactual band is associated only with synthesis of full-length GCN4, it does not interfere with the results, except in lane 7 where it obscures the Fos$^G$ complex band due to almost similar electrophoretic gel mobilities; *c*, To test for specificity of dimerization, Jun (lane 10), Fos$^G$ (lane 11) or a co-synthesis of both Jun and Fos$^G$ proteins (lane 12) were incubated with the optimal binding-site-containing fragment. Only Fos$^G$ complexes are seen (lanes 11 and 12, black arrows). When GCN4 protein (lane 13) or a co-synthesis of GCN4 and c-Fos proteins (lane 15) are incubated with the target DNA, only GCN4 complexes are found. No complex was observed when c-Fos protein alone was incubated with the target DNA (lane 14). The Jun and c-Fos proteins synthesized in these experiments were functional; as expected[1,2–14], when both proteins are mixed and incubated with the DNA fragment containing the optimal GCN4/AP-1 site, they generate a protein–DNA complex (data not shown). The experiments represented in the separate panels involve different protein preparations, gels and autoradiograph exposure, which accounts for any small differences in the Fos$^G$ complex band intensity between panels.

METHODS. Labelling of protein and analysis of DNA–protein complexes have been described[8,17], except that the binding buffer contained poly(dI · dC) at a final concentration of 100 μg ml$^{-1}$ instead of salmon sperm DNA. Cellular-Fos protein was generated from pGEM*Fos X* (from T. Halazonetis). Yields and purity of protein preparations were routinely established by SDS-PAGE and autoradiography.

An important conclusion from our results is that the leucine zipper is sufficient to confer dimerization specificity. Specifically, $Fos^G$ contains only the 33 C-terminal residues of GCN4 (the leucine zipper region) but, unlike Fos, can form heterodimers with GCN4 but not with Jun. Moreover, these observations together with the fact that $Fos^G$ alone binds the GCN4/AP-1 site strongly suggest that the leucine zipper is sufficient for dimerization *per se*. (The possibility that other regions of Fos and GCN4 contribute a common, non-specific dimerization function cannot be excluded, but seems unlikely.) These results agree with an earlier observation that a synthetic peptide containing the 33 C-terminal residues of GCN4 forms stable dimers[7]: the conditions here were more physiological, however (concentrations $10^4$–$10^5$ times lower) and experiments were performed in the context of an intact DNA-binding domain.

The leucine-zipper motif was initially defined by the presence of four to five leucine residues spaced seven amino-acid residues apart in a region of the protein that could permit the formation of an $\alpha$-helix; other than this, there was little overall amino-acid sequence similarity[1]. The fact that swapping the leucine zipper of Fos for that of GCN4 yields a specific DNA-binding protein indicates that these regons are functionally homologous. Nevertheless, it is clear that although the conserved leucines are important for the dimerization of C/EBP (ref. 3), Fos[2,4–6], Jun[4,5], and GCN4 (J.W.S., W. J. van Heeckeren and K.S., unpublished results), other non-conserved residues in the various zipper regions must be involved in dimerization specifiicity. Thus, the ability of this class of proteins to form homodimers or heterodimers will depend on the association properties of individual leucine-zipper regions.

Another conclusion from our experiments is that Fos has an inherent specific DNA-binding activity. Because the GCN4 leucine-zipper region alone does not bind DNA[17], some region of Fos (probably the basic region) must contribute to the ability of $Fos^G$ to interact directly with specific DNA sequences. Furthermore, because the GCN4 leucine zipper can convert Fos into a DNA-binding protein that bypasses the requirement for Jun, the failure of native Fos to bind to DNA almost certainly reflects its inability to form homodimers. In terms of the Fos–Jun heterodimer, these considerations strongly argue against the model that Fos indirectly affects DNA-binding by Jun. Instead, they indicate that Fos and Jun monomers in the heteromeric complex interact with adjacent half-sites.

Finally, chimaeric proteins such as $Fos^G$ should be useful in elucidating the function *in vivo* of proteins involved in heteromeric complexes. For example, transcription involving the AP-1 site is regulated by phorbol esters[10,11], a process which might be mediated by Jun and/or Fos. As $Fos^G$ derivatives can bind to the AP-1 site without Jun, it should be possible to separate the effects of these two oncoproteins. Chimaeric proteins containing the GCN4 leucine zipper could be used to investigate how proteins with leucine zippers might bind DNA specifically. For example, specific DNA-binding by the Myc oncoprotein has yet to be demonstrated, but perhaps by replacing its leucine zipper with that from GCN4, Myc could be converted to a DNA-binding protein whose target sites might be identified by the random selection method[16]. □

Received 12 June; accepted 2 August 1989.

1. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **240,** 1759–1764 (1988).
2. Kouzarides, T. & Ziff, E. *Nature* **336,** 646–651 (1988).
3. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **243,** 1681–1688 (1989).
4. Turner, R. & Tjian, R. *Science* **243,** 1689–1964 (1989).
5. Gentz, R., Rauscher III, F. J., Abate, C. & Curran, T. *Science* **243,** 1695–1699 (1989).
6. Schuermann, M. *et al. Cell* **56,** 507–516 (1989).
7. O'Shea, E. K., Rutkowski, R. & Kim, P. *Science* **243,** 538–542 (1989).
8. Hope, I. A. & Struhl, K. *EMBO J.* **6,** 2781–2784 (1987).
9. Struhl, K. *Cell* **50,** 841–846 (1987).
10. Bohmann, D. *et al. Science* **238,** 1386–1392 (1987).
11. Angel, P. *et al. Nature* **332,** 166–171 (1988).
12. Nakabeppu, Y., Ryder, K. & Nathans, D. *Cell* **55,** 907–915 (1988).
13. Halazonetis, T. D., Georgopoulos, K., Greenberg, M. E. & Leder, P. *Cell* **55,** 917–924 (1988).
14. Sassone-Corsi, P., Ransone, L. J., Lamph, W. W. & Verma, I. M. *Nature* **336,** 692–695 (1988).
15. Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. *Science* **234,** 451–457 (1986).
16. Oliphant, A. R., Brandl, C. J. & Struhl, K. *Molec. Cell. Biol.* **9,** 2944–2949 (1989).
17. Hope, I. A. & Struhl, K. *Cell* **46,** 885–894 (1986).
18. Little, J. W. & Hill, S. A. *Proc. natn. Acad. Sci. U.S.A.* **82,** 2301–2305 (1985).
19. Hurstel, S., Granger-Schnarr, M. & Schnarr, M. *EMBO J.* **7,** 269–275 (1988).
20. Hope, I. A., Mahadevan, S. & Struhl, K. *Nature* **333,** 635–640 (1988).
21. Lech, K., Anderson, K. & Brent, R. *Cell* **52,** 179–184 (1988).
22. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. *Meth. Enzym.* **154,** 367–382 (1987).
23. Struhl, K. *Nature* **332,** 649–650 (1988).

# Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific splice acceptor

**Ronald B. Emeson*, Farah Hedjran*, Joanne M. Yeakley*†, Jeffrey W. Guise* & Michael G. Rosenfeld*§**

§ Howard Hughes Medical Institute, * Eukaryotic Regulatory Biology Program, † Department of Biology, School of Medicine, University of California, San Diego, La Jolla, California 92093, USA

**ALTERNATIVE splicing of eukaryotic messenger RNA precursors represents a common mechanism for generating multiple transcripts from a single gene[1,2]. Although there has been increasing information concerning the sequence requirements and the biochemical mechanisms involved in the constitutive splicing of primary RNA transcripts, very little is known about the sequences or mechanisms which determine alternative RNA-processing events in complex transcription units[3–6]. The calcitonin/calcitonin gene-related peptide (CGRP) primary RNA transcript undergoes tissue-specific alternative processing, resulting in the differential production of calcitonin mRNA in thyroid C cells and CGRP mRNA in neurons of the central and peripheral nervous systems[7,8]. To elucidate the molecular mechanisms underlying these alternative RNA processing events, we have examined the nucleotide sequences involved in the production of calcitonin and CGRP mRNAs. Analyses of HeLa and F9 cell lines transfected with a variety of mutant calcitonin/CGRP transcription units have demonstrated that alternative splice-site selection is primarily regulated by *cis*-active element(s) near the calcitonin-specific 3′-splice junction. We suggest that the tissue-specific pattern of alternative RNA processing is conferred by sequence information at the calcitonin-specific acceptor which serves to inhibit the production of calcitonin transcripts in CGRP-producing cells.**

The rat calcitonin/CGRP gene is comprised of six exons; calcitonin mRNA is produced by splicing of the first three exons to the fourth exon, accompanied by cleavage and polyadenylation at the 3′ end of the fourth exon. CGRP mRNA production results from the splicing of the first three exons to the fifth and sixth exons and use of a distal poly(A) site at the 3′ end of the sixth exon (Fig. 1*a*)[7,9]. Previous studies on the expression of a metallothionein-calcitonin/CGRP fusion gene in transgenic mice demonstrated that almost all tissues produced predominantly calcitonin mRNA[10]. In most neurons however, CGRP transcripts represented the main RNA species. The simplest interpretation of these results is that the ability to splice the calcitonin/CGRP primary transcript to produce mature CGRP mRNA requires a neuron-specific splicing machinery, thereby suggesting that calcitonin mRNA is likely to represent the unregulated splicing choice. Alternatively, production of calcitonin-specific transcripts could require a tissue-specific factor which is widely distributed, but absent from neuronal tissues.

Two potential mechanisms have been proposed for the regulation of alternative splice-site use in the calcitonin/CGRP gene. One mechanism involves sequence or site-specific poly(A) site