Mini-Review

An Analysis of Vertebrate mRNA Sequences: Intimations of Translational Control

Marilyn Kozak
Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Piscataway, NJ 08854

Abstract. Five structural features in mRNAs have been found to contribute to the fidelity and efficiency of initiation by eukaryotic ribosomes. Scrutiny of vertebrate cDNA sequences in light of these criteria reveals a set of transcripts—encoding oncoproteins, growth factors, transcription factors, and other regulatory proteins—that seem designed to be translated poorly. Thus, throttling at the level of translation may be a critical component of gene regulation in vertebrates. An alternative interpretation is that some (perhaps many) cDNAs with encumbered 5' noncoding sequences represent mRNA precursors, which would imply extensive regulation at a posttranscriptional step that precedes translation.

Initiation of translation in multicellular eukaryotes is influenced by five aspects of mRNA structure: (a) the m7G cap (355); (b) the primary sequence or context surrounding the AUG codon (187, 190, 194); (c) the position of the AUG codon, i.e., whether it is the first AUG in the message (186); (d) leader length (198, 199); and (e) secondary structure both upstream (188, 195) and downstream (196) from the AUG codon. Elsewhere (200) I have reviewed the evidence for these five features and explained how they work together to determine the fidelity and efficiency of initiation. A scanning mechanism for initiation can explain many of the effects of cap, context, position, etc. The scanning model (193) in its simplest form postulates that a 40S ribosomal subunit, carrying Met-tRNA**, and an imperfectly defined set of initiation factors (302), enters at the 5' end of the mRNA and migrates linearly until it reaches the first AUG codon, whereupon a 60S subunit joins and the first peptide bond is formed. Evidence in support of the model has been adduced previously (62, 193, 197). More recent evidence for scanning includes the apparent queuing of 40S ribosomal subunits on long leader sequences (199) and the stalling of 40S subunits on the 5' side of a stable hairpin structure introduced between the cap and the AUG codon (195). The possibility of initiation by a mechanism other than scanning has been proposed (158) and is evaluated elsewhere (197).

The trick to identifying elements in 5' noncoding sequences that can modulate translation was to isolate each feature (200), an approach made possible by the techniques of genetic engineering. For example, by devising a transcript in which the first AUG codon was in an unfavorable context and hence "leaky," we were able to show that downstream secondary structure enhances recognition of the preceding AUG codon, apparently by preventing the 40S ribosomal subunit from scanning too fast or too far (196). The contribution of downstream secondary structure would have been missed had the primary sequence around the first AUG codon been more favorable, and vice versa. Having used the reductionist approach to identify several features that can modulate initiation, I attempt herein to put the story back together by examining the extent to which natural mRNAs conform to the experimentally determined requirements for initiation. A surprising realization is that, although most vertebrate mRNAs have features that ensure the fidelity of initiation (i.e., selection of the correct AUG codon), many do not appear to be designed for efficient translation. This would seem to have important implications for gene regulation.

Leader Sequences on Vertebrate mRNAs: An Overview

In considering the extent to which natural mRNAs conform to the five recognized requirements for initiation, I will focus on mRNAs from vertebrate cells where the rapidly expanding catalogue of sequences provides grist for analysis. mRNAs from animal and plant viruses and yeasts are mentioned only incidentally when they uniquely illustrate a point.

Every cellular mRNA that has been examined is capped (355). Not every mRNA has been examined, of course, but it seems unlikely that uncapped cellular mRNAs will be found inasmuch as the cap is crucial not only for translation (355) but also for mRNA stability and transport (126). Although the uncapped mRNAs from picornaviruses are a vogue topic for discussion (158), it should be remembered that picornaviruses are exceptions. All other animal viruses produce capped mRNAs, even when doing so requires the virus to encode its own capping and modifying enzymes (189).

The requirement for a favorable context around the AUG initiator codon is also met by nearly all mRNAs from higher eukaryotes. The consensus sequence for initiation derived from a compilation of 699 vertebrate mRNAs (191) is GCCG<sup>5</sup>CCCAUGG, the same as the experimentally derived optimal sequence (187, 190, 194). While the full consensus sequence is found in only a small number of vertebrate mRNAs, the two positions most critical for function (187) are highly conserved: 97% of vertebrate mRNAs have a purine (usually A) in position -3 and 46% have G in position -4.
Bifunctional Genes and Bifunctional mRNAs

The importance of position in determining the functional initiator codon is illustrated by a family of genes that are required to produce two versions of the encoded protein (Fig. 1 and Table I). The general idea is that ribosomes need to initiate translation from the first and second AUG codons in each of these genes. Although the longer mRNA from the model gene in Fig. 1A contains both AUG codons, the presence of a good context around the first AUG codon precludes access to the second. The solution is that the gene produces a second form of mRNA, the 5' end of which maps between the two AUG codons. In each mRNA ribosomes initiate at the first and only the first AUG codon.

There is a way for two proteins encoded in overlapping open reading frames (ORFs) to be translated from a single mRNA; namely, by introducing a poor context around the first AUG codon. In the bifunctional mRNAs listed in Table I, the first AUG is in a strong context, as is usually the case, one gene can produce two proteins only by producing two mRNAs, i.e., by initiating one transcript (P1) upstream from the first AUG codon and initiating a second transcript (P2) downstream from that AUG. Often the NH2-terminal amino acid extension targets the long form of the protein to a special intracellular compartment. Examples are given in Table I. (B) Leaky scanning permits synthesis of two proteins from one mRNA when the context around the first AUG codon is unfavorable; i.e., when a pyrimidine occurs in position -3, or when there is a G in position -3 and something other than G in position +4. Examples of genes that use leaky scanning are given in Table II.

A. PROMOTER SWITCHING

B. LEAKY SCANNING

Figure 1. Two mechanisms that enable one gene to produce two versions of the encoded protein. (A) When the first AUG codon is in a strong context, as is usually the case, one gene can produce two proteins only by producing two mRNAs, i.e., by initiating one transcript (P1) upstream from the first AUG codon and initiating a second transcript (P2) downstream from that AUG. Often the NH2-terminal amino acid extension targets the long form of the protein to a special intracellular compartment. Examples are given in Table I. (B) Leaky scanning permits synthesis of two proteins from one mRNA when the context around the first AUG codon is unfavorable; i.e., when a pyrimidine occurs in position -3, or when there is a G in position -3 and something other than G in position +4. Examples of genes that use leaky scanning are given in Table II.
by initiating transcription from two promoters, as illustrated in Fig. 1 A

Table I. Genes that Produce Two Overlapping Proteins by Initiating Transcription from Two Promoters, as Illustrated in Fig. 1 A

| Gene | Description |
|------|-------------|
| Val-tRNA synthetase (VAS1, yeast) | (49) |
| His-tRNA synthetase (HTS1, yeast) | (282) |
| α-Isopropylmalate synthase (LEU4, yeast) | (18) |
| tRNA dimethyltransferase (TRM1, yeast) | (90) |
| Serine:pyruvate aminotransferase (SPT, rat) | (291) |
| Cyclophilin (N. crassa) | (398, 399) |
| tRNA dimethyltransferase (TRM1, yeast) | (90) |
| a-Isopropylmalate synthase (LEU4, yeast) | (18) |
| His-tRNA synthetase (HTS1, yeast) | (282) |
| Val-tRNA synthetase (VAS1, yeast) | (49) |
| Gelsolin (human) | (207) |
| Surface antigen, Hepatitis B virus | (307) |
| E2 protein, bovine papillomavirus | (211) |
| Invertase (SUC2, yeast) | (43) |

Superscripts refer to ways in which the function of the long isoform differs from that of the shorter protein: (a) import into mitochondria, (b) secretion, (c) control of transcription. The expression of porphobilinogen deaminase, erythroid membrane protein 4.1, and Gsα protein require alternative splicing as well as promoter switching; they are included because the net effect is activation of an internal AUG initiator codon by making it the first AUG in the mRNA. Not listed are some interesting genes that produce small amounts of 5′ truncated transcripts in extraneous tissues (63, 160, 405, 410). Even when such mRNAs can be shown to direct synthesis of a polypeptide fragment (63), which is almost inevitable if the transcript enters the cytoplasm, the phenomenon might reflect inadvertent expression; what needs to be established is that the NH2 terminally truncated polypeptide serves a unique function in the ectopic tissue.

II, the first AUG codon deviates from the consensus sequence in either or both of positions −3 and +4. (Three exceptions are discussed in the Table II legend.) The result is “leaky scanning” in which some 40S ribosomes bypass the first AUG codon; initiation occurs from the first and second AUG codons in these mRNAs (Fig. 1 B). Curiously, most of these bifunctional mRNAs are of viral origin. Only two cellular mRNAs are listed in Table II, and in neither of those cases has the short protein been shown to mediate a function distinct from the long isoform. Thus, as a practical device for producing two proteins from one gene, cells rely mostly on a transcriptional device (Fig. 1 A) while viruses use a translational ploy (Fig. 1 B). Leaky scanning does not require virus-induced modifications of the translational machinery, however, inasmuch as the isolated revcorev S1 gene, when expressed in uninfected COS cells, produces the expected two proteins (95). Leaky scanning may also result when the first AUG codon resides close to (within 12 or so nucleotides of) the cap (113, 198), although leakiness due to an unfavorable context is the more common mechanism.

Not included in Table II are a few bifunctional mRNAs (344, 347, 397) in which two proteins are produced from nonoverlapping ORFs. In such cases, leaky scanning may also result when the first AUG codon resides close to (within 12 or so nucleotides of) the cap (113, 198), although leakiness due to an unfavorable context is the more common mechanism.

Leaky scanning may also result when the first AUG codon resides close to (within 12 or so nucleotides of) the cap (113, 198), although leakiness due to an unfavorable context is the more common mechanism.

Table II. Genes that Produce Two Proteins from One mRNA by Leaky Scanning, as Illustrated in Fig. 1 B

| Gene | Description |
|------|-------------|
| Sendai (paramyxovirus) | (P, C) |
| Reovirus S1 mRNA | (92, 95) |
| BUNyavirus s-RNA | (349) |
| Adenovirus, region E1B | (21K, 55K) |
| Human T-Cell Leukemia Virus (HTLV-I) | (p27, p40) |
| Potato leafroll virus | (17K) |
| Satellite tobacco mosaic virus | (6.8K) |
| Rotavirus S1 mRNA | (349) |
| Potyvirus, 200K RP | (170) |
| Rubella virus | (243) |
| Erythroid membrane protein 4.1 (human) | (68) |

In several cases the connection between leaky scanning and a suboptimal context around the first AUG codon has been confirmed by mutation analysis (78, 95, 346, 370). Only three cases have been described in which ribosomes initiate at the first and second AUG codons despite a favorable context around the first AUG. The most important of these exceptions is influenza virus B, where the proximity of the second AUG codon to the first AUG seems to allow leaky scanning (414). The other exceptions are barley stripe mosaic virus (308) and cowpea mosaic virus RNA-M (144), where the absence of secondary structure downstream from the first “strong” AUG codon might account for the leakiness. Leaky scanning in these two plant viruses might be inadvertent, inasmuch as the second protein isoform does not contribute to viral infectivity. In contrast, for most of the other viral entries, both proteins produced from the bifunctional mRNA are required for infectivity. The fact that the overlapping arrangement of ORFs is conserved among different members of the paramyoxovirus, reovirus, bunyavirus, adenovirus, rotavirus, and potyvirus families constitutes additional evidence that the synthesis of two proteins by leaky scanning is not accidental in those cases. In the case of HIV-1, wuv functions more efficiently in promoting the processing of env when the two proteins are translated from the same mRNA than when they are expressed experimentally from separate transcripts (M. Martin, K. Strebel and R. Willey, personal communication). An asterisk, preceding some entries in the table, means that only one of the two proteins predicted by the mRNA sequence has been detected so far.

Kozak An Analysis of Vertebrate mRNA Sequences 889
of the LP and EBNA-2 proteins of Epstein-Barr virus (5).
The functionality of an interesting bicistronic transcript for
growth/differentiation factor 1 has not yet been established.
(Although the only detected transcript for GDF-1 in 14.5-d
mouse embryos is a 3-kb bicistronic transcript in which
GDF-1 is the downstream cistron, the GDF-1 protein de-
tected by immunohistochemical analysis in 14.5-d embryos
might actually have been synthesized a few days earlier,
when a 1.4-kb transcript was the predominant form (218).)
The ability to reinitiate to some extent after translating a
small 5' ORF, and the tendency of 40S ribosomes to scan past
an AUG codon in a weak context, explain how ribosomes can
initiate from an AUG codon that is not first. Nevertheless,
the occurrence of upstream AUG codons nearly always
reduces the efficiency of initiating from downstream. Thus,
mRNA (or, more correctly, cDNA) sequences that are pep-
pered with small upstream ORFs pose a problem.

5' Noncoding Exons, Introns, and Upstream AUG Codons

The simple question of whether the mRNA from a particular
gene has upstream AUG codons cannot always be answered
simply. Some of the complexities are due to 5' noncoding
exons and associated phenomena such as alternative splic-
ing, inefficient removal of a 5' intron, and the presence of al-
ternative promoters. I will first address those complications
and then try to assess the frequency and significance of up-
stream AUG codons in vertebrate mRNAs.

Nearly one-fourth of the entries in a recent survey of 699
vertebrate mRNA sequences (191) have turned out to have
an intron between the promoter and the start of the major
ORF. The high incidence of 5' introns has theoretical as well
as practical consequences. The first intron in a gene some-
times contains sequences that facilitate transcription (26, 42,
60, 73, 150, 174, 176, 185, 269, 280, 287, 303, 368), an effect
that sometimes requires the intron to be near the 5' end (42,
287, 303). Some genes that have retained a 5' intron thereby
have the ability to switch promoters, in response to hormonal
or tissue-specific inducers, for example, and thus to ex-
change an inefficiently translated 5' noncoding exon for one
that appears more favorable (9, 54, 77, 106, 289). (The pre-
dicted improvement in translation has not yet been verified
for all of those genes.) Another kind of regulation takes the form of allowing a gene to be transcribed in an ec-
topic tissue but preventing its translation by not removing the
5' intron. The expression of gonadotropin-releasing hor-
monene mRNA in extra-hypothalamic tissues is a striking ex-
ample (135). Another might be the expression of the tyrosine
kinase fer gene in testis versus other tissues (101).

From a practical perspective, the frequent presence of 5'
introns (sometimes directly abutting or even interrupting the
AUG codon) necessitates caution in picking a probable start
site for translation, and caution in scoring upstream AUG
triplets. Many claims of mRNAs with AUG-burdened 5'
non-coding sequences (46, 114, 125, 213, 236, 238, 292, 348,
387) have been resolved by finding that the 5' portion of the
cDNA corresponds not to the mature mRNA but to an intron-containing form (36, 135, 225, 235, 284, 316, 388,
407, 408). (see reference 193 for additional examples.)
The growing evidence of incomplete (20, 23, 28, 74, 76, 80, 138,
151, 168, 222, 259, 343, 354, 411) or regulated (417) RNA
processing in mammalian cells underscores the point that
cDNA sequences cannot invariably be equated with func-
tional mRNAs. Some intron-containing transcripts are abun-
dant (70), some enter the cytoplasm (328), and some are
even found on polysomes (411). These problems complicate
attempts to deduce the real structures of vertebrate mRNAs.
While a cDNA sequence that retains an unspliced intron
within the coding domain is easily recognizable as a process-
ing intermediate, the presence of unspliced intron(s) in the
5' noncoding domain is much harder to recognize.

Some genes are transcribed in ectopic tissues from an ill-
placed promoter that burdens the 5' noncoding sequence
with AUG codons, thereby impairing translation in that par-
ticular tissue. In the tissue that constitutes the major site
of expression, however, a different promoter produces a 5' non-
coding sequence that is not so encumbered. Examples in-
clude murine complement-B mRNA in hepatic versus extra-
hepatic tissues (286), rat preproenkephalin mRNA in
tests versus brain (172, 175), rat α-crystallin mRNA in
brain versus other tissues (157), and rat farnesyl pyrophos-
phate synthetase mRNA in liver versus tests (390). In the
last three cases, the predicted difference in translational
efficiency has been verified experimentally. Another ploy in-
volves switching to a shorter, more efficiently translated
leader sequence in response to some developmental (61, 425)
or environmental cue, such as stimulation with serum
(409) or retinoic acid (61) or endotoxin (286), or during T
cell maturation (324). Because of promoter switching and/or
alternative splicing, many other vertebrate genes produce
multiple transcripts that differ near the 5' end, and failure to
detect all pertinent forms has sometimes led to false conclu-
sions. The suggestion of "internal initiation" in the chicken
progesterone receptor mRNA (71) is one example of a wrong
conclusion that was righted upon discovering other forms of
mRNA (166, 167). Detecting alternative transcripts is not al-
ways easy! Competition for the primer may cause a minor
transcript to be missed (175). Even the major transcript has
been missed when the primer was positioned inappropriately
(72) or when hybridization conditions were too stringent
(338).

Because of the difficulties described above and various
other complications in cloning or interpretation (6, 32, 37,
68, 69, 146, 226, 430), the frequency of spurious upstream
AUG codons in vertebrate mRNAs is difficult to estimate;
but clearly it is not as high as superficial reading of the lit-
erature might suggest. When upstream AUG codons do occur,
the AUG-burdened leader sequence impairs translation (9a,
12, 105, 157, 247, 272, 277, 281, 326, 390, 409, 412, 424),
as expected if initiation occurs by the conventional scanning
process.

A partial listing of cDNAs with AUG-burdened leader se-
quences is given in Table III. It includes many proto-
 Oncogenes as well as genes for transcription factors, a variety
of receptor proteins, signal transduction components, and
many proteins involved in the immune response. One con-
clusion might be that mRNAs that encode critical regulatory
proteins are intended to be translated poorly. I suspect that
conclusion is true for some entries in the table, but some
(perhaps many) entries might reflect a different type of regu-
lation. For example, the repeated finding of incompletely
spliced transcripts in lymphocytes (20, 107, 389, 411) and re-
cent evidence that undefined posttranscriptional processes
improve upon mitogen activation of lymphocytes (67 en-
| Tumor associated (proto-oncogenes, etc.) | Transcription factors and DNA-binding proteins | Receptors for the following ligands |
|-----------------------------------------|-----------------------------------------------|-----------------------------------|
| * abl, human (21)                      | NFI-B (TGGCA), chicken (330)                  | * Acetylcholine, rat (224)        |
| * bcl-2, human (401)                   | NFI-X, hamster (112)                          | * Angiotensin II, rat (276)       |
| ear-7, human (266)                     | cAMP response (CRE-BP1) (278)                | * Atrial natriuretic peptide (110) |
| * erb-A, human (413)                   | IFN response (IRE-BF1) (421)                  | D1 dopamine, mouse (268)         |
| erg, human (320)                       | PRDII-BF1 (96)                                | Estrogen, chicken (203)          |
| Evi-1, human (271)                     | † DBP, rat liver (272)                        | GABA $\alpha_1$, mouse (171)     |
| Evi-2, mouse (38)                      | HOX 2G, human (2)                             | GABA A $\gamma_2$, mouse (115)   |
| * fgr, human (122)                     | † HOX 5.1, human (61)                         | Gastrin-releasing peptide (13)    |
| * fos-B, mouse (427)                   | Hox 2.9, mouse (274)                          | Glycine, human brain (119)        |
| HCK, human (318)                       | Hox 3, mouse (31)                             | Heparin-binding GF (K-sam) (133)  |
| * int-2, human (34)                    | Hox 3.1, mouse (10)                           | Interferon $\alpha$, human (402)  |
| * lck, mouse (324)                     | Hox 3.2, mouse (93)                           | Progesterone, rabbit (262)        |
| * mos, mouse (314)                     | † BTFT3 (general) (428)                       | Prolactin, rabbit (89)            |
| ROS-1, human (22)                      | OTF-2, human (339)                            | Retinoic acid (hRAR-$\alpha$) (30) |
| * sis, (PDGF-2) (98, 319)              | TFE, canine (161)                             | Retinoic acid (mRAR-$\beta$) (426) |
| * sno, human (285)                     | KUP, human (48)                               | Retinoic acid (hRAR-$\gamma$) (204) |
| syn (silk), human (352)                | pol II factor UBF, rat (296)                  | Serotonin ic, rat (164)           |
| * T-cell 11p15 (25)                    | HNF-1$\beta$, mouse (256)                     | Serotonin SHT-2, rat & CHO (45, 313) |

**Table III. Vertebrate cDNA Sequences that Have Three or More AUG Codons Upstream from the Major Open Reading Frame**

**Immune/inflammation mediators**

| Immune/inflammation mediators | Signal transduction | Growth factors |
|-------------------------------|---------------------|----------------|
| † Interleukin-7, mouse (239, 281) | chick embryo tyr kinase (304) | Keratinocyte GF (99) |
| IL-1 receptor, mouse (358)    | mouse liver tyr kinase (244) | * Insulin-like GF-1 (17, 105) |
| IL-2 receptor-$\beta$, human (132) | FER tyr kinase (130) | Platelet-derived GF-A (328) |
| IL-3 receptor, mouse (116)    | tyk2 tyr kinase (100) | Epidermal GF (16) |
| IL-5 receptor, mouse (382)    | elk tyr kinase (223) | † Transforming GF-$\beta$ (9a, 209) |
| IL-6 receptor, human (420)    | ERK3 ser/thr kinase (29) | See also: proto-oncogenes |
| IL-7 receptor, mouse (309)    | p58 protein kinase (173) | |
| G-CSF receptor, mouse (108)   | rp-S6 kinase, chick (4) | |
| C3b receptor (Mac-1a) (315)   | protein tyr phosphatase (PTPase) LRP (293) | |
| CD28, human T cells (217)     | PTPase, megakaryocyte (121a) | |
| CD75, human B cells (365)     | Phospholipase C-1, rat (375) | |
| * Ly-5 (CD45, CALLA) (334)    | cAMP phosphodiesterase (230) | |
| Tgr-1, T cells (252)          | insulin receptor substrate-1 (IRS-1, pp185) (376a) | |
| Surface antigen 114/A10 (87)  | Tyrosine kinase, leucocyte (19) | |
| Tyrosine kinase, leucocyte (19) | bFGF receptor II, mouse (143) | |
| * IgE receptor (high affinity) | (229) | |

Some, perhaps many, of these cDNA sequences are likely to represent mRNA precursors rather than functional mRNAs. The literature contains scattered reports of AUG-burdened cDNA sequences in addition to those listed here. * The gene produces multiple transcripts with alternative 5' noncoding sequences; †, translation is more efficient with transcripts (natural or derived) that lack the encumbered leader sequence.

Some, perhaps many, of these cDNA sequences from immune cells might correspond to precursors rather than to functional mRNAs. The same may be true of AUG-burdened transcripts from transcription factor genes, since some of those transcripts are restricted to the nucleus (61, 70, 275); in other cases, a transcript is detectable but the corresponding protein is not (256). The first report of mitogen-regulated splicing of 5' introns in vertebrate genes has just been published (417), giving substance to the hypothesis that nontranslatable transcripts may be synthesized and stored for later processing.

Proto-oncogenes, on the other hand, might be genuine candidates for translational modulation via an encumbered leader sequence. Interpretation is complicated by the fact that many proto-oncogenes produce transcripts with alternative 5' sequences (these are marked by asterisks in Table III), but several observations support the idea that proto-oncogene mRNAs are meant to be translated inefficiently: c-mos transcripts are found on very small polysomes (314); some activated oncogenes produce transcripts with simpler 5' noncoding sequences than the corresponding proto-oncogenes (321, 337); and the experimental expression of many proto-oncogenes improves dramatically upon deleting portions of the leader sequence (12, 50, 247, 319). To propose that proto-oncogene mRNAs might be translated by a mechanism other than scanning (253), inasmuch as their
AUG-burdened leader sequences seem incompatible with efficient scanning, is to miss the point that these potent proteins probably have to be translated inefficiently.

**Occurrence and Consequences of Secondary Structure**

The catalogue of vertebrate mRNAs with GC-rich (hence highly structured) leader sequences again includes many mRNAs for oncogenes, growth factors, transcription factors, signal transduction components, and a wide variety of receptor proteins (Table IV). Again, the presence of an encumbered leader sequence suggests that production of these critical regulatory proteins is throttled at the level of translation. The GC-cohort also includes many housekeeping genes, which are generally recognized to be expressed at low levels. While it is easy to show that many of these leader sequences support translation poorly (see below), delineating the cause is not simple. The extraordinarily high GC content (70 to 90%) predicts many alternative base pairings, making it impossible to pinpoint a target for mutagenesis. Consequently, our understanding of how particular base-paired structures affect translation relies heavily on experiments carried out with synthetic transcripts (188, 195, 196) in which discrete stem-and-loop structures have been introduced and their existence documented by genetic techniques.

The best evidence that mRNAs in Table IV are translationally impaired is the dramatic improvement in expression when the GC-rich leader sequences (some of which also contain upstream AUG codons) are truncated experimentally (53, 82, 182, 221, 246, 273, 319, 395). (Discrepancies between mRNA levels and protein accumulation in some stages or conditions of cell growth may be another indication of translational control of transcription factor and other such genes (66, 81, 83, 205, 340, 357, 373); but in most of those cases alternative explanations, such as compartmentalization of the mRNA or accelerated degradation of the protein, have not been ruled out.) Some genes in Table IV actually produce two versions of mRNA, on one of which the leader is shorter and less encumbered than on the other (178, 283, 317, 377). In the few cases where long- and short-leader mRNAs from the same gene have been put to the test, the short-leader transcript nearly always supports translation more efficiently (153, 283, 337, 409). Indeed, the discrepancy in translatability dependent on leader sequences can be so profound that a minor transcript from certain genes appears to be the major functional mRNA (149, 264, 283). Other genes in the GC-rich cohort produce transcripts with so many different leader sequences (75, 85a, 165, 220, 245, 426) that it is impossible to guess, and no small task to test, their functionality.

Notwithstanding those caveats, the extraordinary number of mRNAs with GC-burdened leader sequences forces the idea that synthesis of critical cellular proteins is probably throttled at the level of translation. Under constitutive conditions, the synthesis of a single molecule of such a protein could conceivably take hours as a 40S subunit slowly maneuvers its way to the downstream AUG codon. If slow initiation of translation is a key to limiting the production of proteins that would be lethal if overproduced, one should not be surprised that such mRNAs are virtually untranslatable in standard in vitro assays in which mRNAs are expected to produce a product in minutes! A compelling rationale for the cumbersome 5' noncoding sequences on so many regulatory genes is that those transcripts should respond as a cohort to shifts in the cell's translational capacity. As for how hypothetical shifts in translational capacity might be accomplished, changes in the extent of phosphorylation of initiation factors and ribosomal proteins have often been remarked (137, 393). With the notable exception of eIF-2 (65), however, hard evidence for the functional consequences of phosphorylation remains elusive.

A structured leader sequence may have qualitative as well as quantitative effects on translation. In a small number of vertebrate mRNAs, ribosomes initiate at a non-AUG codon, such as ACG, CUG, or GUG (3, 20, 86, 103, 127, 219, 232a, 336). The list is slightly longer if one counts viral mRNAs (15, 78, 312, 367). It is not valid, however, to count mRNAs in which the use of alternative initiator codons has been documented only in vitro, where inappropriate reaction conditions can activate cryptic sites that would not be used in vivo (194). Initiation at non-AUG codons is usually inefficient and usually occurs in addition to using the first AUG codon. The result is synthesis of an "extra" NH2-terminally extended version of the protein. (There are only two instances in which a protein derives uniquely from initiation at an upstream non-AUG codon and not, at least in part, from the first in-frame AUG codon. One occurs in cells transfected with ltk tyrosine kinase cDNA, in which five out-of-frame AUG triplets occur between the putative CUG initiation site and the first in-frame AUG (20); initiation at the far-upstream CUG codon thus circumvents the problem of getting past out-of-frame AUG codons. As yet, however, initiation at the upstream CUG codon has not been demonstrated with the endogenous ltk gene in untransfected cells. The other very intriguing example is the apparently unique use of an AUU codon to initiate translation of the human enhancer factor TEF-1 (416).) All of the vertebrate mRNAs that use a nonstandard initiator codon have GC-rich leader sequences, prompting the speculation that the slow transit of scanning 40S ribosomes across a highly structured 5' noncoding sequence might be responsible for activating cryptic upstream sites (196). Indeed, initiation at upstream non-AUG codons in synthetic transcripts was considerably enhanced upon introducing secondary structure in an appropriate position 3' of the cognate initiator codon (196). While the NH2-terminally extended polypeptides initiated from non-AUG sites in viral and cellular mRNAs occasionally have distinct functions (15, 312, 367) or distributions (3, 40, 232a), it would be simplistic to assume that every instance of initiation from a cryptic upstream site is functionally important. Given the GC-richness of leader sequences on mammalian mRNAs, spurious upstream initiation events may be unavoidable byproducts of the way eukaryotic ribosomes arrive at the AUG codon. In avian cells, the efficiency of initiating at the upstream CUG codon in c-myc mRNA is regulatable by culture conditions (Stephen Hann, personal communication). This suggests interesting modulation of the translational machinery, but it does not aid the so-far unsuccessful effort to ascribe functional significance to the NH2-terminally extended form of c-myc.

Many vertebrate mRNAs that have highly structured leader sequences also have upstream AUG codons (see the entries marked t in Table IV). This coincidence might be viewed in either of two ways. One rationalization invokes the
The leader sequences on cDNAs from these genes have a G+C content of 70 to 90%, which would seem to imply extensive secondary structure. Most but not all genes for oncogenes, receptors, transcription factors and signal transduction components belong to this GC-cohort. Overall, only 19% of the vertebrate mRNAs compiled in reference 191 have GC-rich leader sequences. Of these, ~40% fall into the categories listed in Table IV, although oncogenes, receptors, transcription factors, signal transduction, and growth factor genes constitute only 13% of the total sequences in the compilation. Another ~30% of the GC-burdened leader sequences derive from mRNAs for cytoskeletal and housekeeping proteins. Thus, GC-rich 5'-noncoding sequences are not randomly distributed among vertebrate genes. The frequency of GC-rich mRNAs does appear to be increasing, however, now that technical improvements enable the routine cloning of cDNAs derived from scarce transcripts. *, The gene produces multiple transcripts with alternative 5' noncoding sequences; †, the GC-rich leader sequence also contains upstream AUG codon(s); ‡, translation improves upon deleting portions of the leader sequence.

Table IV. Some Vertebrate Genes Predicted to Have Highly Structured 5' Noncoding Sequences

| Tumor associated (proto-oncogenes, etc.) | Receptors for the following ligands |
|----------------------------------------|-----------------------------------|
| † bcl-3, human (294)                   | † n-acetylcholine, a5 (28)        |
| †† BCR, human (353)                   | † n-acetylcholine, β2 (7)         |
| bmk, mouse (145)                      | † α2B-adrenergic, rat (102)       |
| DBL, human (94)                       | † β-adrenergic, rat (356)         |
| †† erbA-1, human THRA (215)           | † β2-adrenergic, human (91)       |
| erbB (HER1) (124)                     | † atrial natriuretic peptide (53) |
| eph, human (139)                      | † dopamine D2, rat (267)          |
| ets-2, human (250)                    | †* estrogem, human (118)          |
| for, human (372)                      | † HER3 (THR family) (265)         |
| † hck, mouse (232)                    | † insulin, human (351)            |
| † HER2 (new) human (384)              | IGF-II, human (270)               |
| † kat, human (423)                    | † integrin, chicken (385)         |
| † int-1, human (406)                  | † interferon-γ (117)              |
| †† int-2, human (34, 120)             | † interleukin-1 (59)              |
| jun, human (8)                        | mannose-6-phosphate (231)         |
| † KS (Kaposi) (79)                    | N10 (TH receptor family) (333)    |
| lyl-1, human (255)                    | nerve growth factor (350)         |
| lyn, human (419)                      | poliovirus, human (184)           |
| myb, mouse (362)                      | † progestrone, chicken (162)      |
| † L-myc, human (169)                  | †* retinoic acid (hRARα) (204)    |
| † N-MYC, human (366)                  | ryanodine, rabbit (383)           |
| pim-1, human (323)                    | syndecan, human (242)             |
| H-ras, human (148)                    | thrombin, human (404)             |
| Rel, human (118)                      | †* transferrin, human (341)       |
| rer (381)                             | tumor necrosis factor (183)       |
| sis (PDGF-2) human (98, 319)          | (See also: proto-oncogenes)       |
| skl, human (285)                      |                                  |
| ret, human (120)                      |                                  |
| TFII (elongation) (422)               |                                  |
| †* TCF-1, murine T-cells (298)        |                                  |
| TEF-1, human EBP (416)                |                                  |
| SCIP rat nerve (204a)                 |                                  |

The leader sequences on cDNAs from these genes have a G+C content of 70 to 90%, which would seem to imply extensive secondary structure. Most but not all genes for oncogenes, receptors, transcription factors and signal transduction components belong to this GC-cohort. Overall, only 19% of the vertebrate mRNAs compiled in reference 191 have GC-rich leader sequences. Of these, ~40% fall into the categories listed in Table IV, although oncogenes, receptors, transcription factors, signal transduction, and growth factor genes constitute only 13% of the total sequences in the compilation. Another ~30% of the GC-burdened leader sequences in reference 191 derive from mRNAs for cytoskeletal and housekeeping proteins. Thus, GC-rich 5'-noncoding sequences are not randomly distributed among vertebrate genes. The frequency of GC-rich mRNAs does appear to be increasing, however, now that technical improvements enable the routine cloning of cDNAs derived from scarce transcripts. *, The gene produces multiple transcripts with alternative 5' noncoding sequences; †, the GC-rich leader sequence also contains upstream AUG codon(s); ‡, translation improves upon deleting portions of the leader sequence.
adage that nothing bad can happen to a rotten eggplant: a
highly structured 5′ sequence is so inhibitory to translation
that the further slight diminution attributable to one or two
small upstream ORFs should hardly matter. A more inter-
esting view is that upstream ORFs (initiating at AUG or AUG-
cognate codons) might actually be necessary to mitigate the
inhibitory effects of a GC-rich leader sequence. The argu-
ment here is that 80S ribosomes engaged in translating the
upstream ORFs might be able to penetrate duplex structures
that are too stable to be penetrated by scanning 40S ribosomal subunits—an idea which is supported by some evi-
dence from experimental constructs (179, 195). If a smatter-
ing of upstream initiator codons indeed facilitates the trans-
lation of mRNAs with highly structured leader sequences,
they probably provide only a small measure of relief. The
major experimental finding is that mRNAs with long, GC-
rich leader sequences are translated inefficiently. And a sur-
prising number of vertebrate mRNAs fit that bill.

Coda
The usually favorable context around the AUG start site in
vertebrate mRNAs ensures the fidelity of initiation. Because
recognition of the AUG codon is a late event in the initiation
process, however, a good context should not, and does not
(194), affect the ability of one mRNA to outcompete another.
Translational efficiency, defined as competitive ability, is
probably determined instead by accessibility of the capped
5′ end of the mRNA, since the 5′ end constitutes the apparent
entry site for the 40S ribosome/40S ribosome complex. Effective
competition for the 40S ribosome/40S ribosome complex is not
sufficient, however; translational efficiency (defined now as
actual production of the intended protein) can still be re-
duced by upstream AUG codons or by base-paired structures
that constitute barriers to the scanning 40S ribosome. The
particulars, such as how much secondary structure is re-
quired to inhibit scanning, and the available evidence are
summarized elsewhere (200). If one accepts the general no-
tion that base-paired structures and upstream AUG codons
can block ribosome entry and/or scanning, then the encum-
ered leader sequences described herein pose problems.

Time will tell which of the AUG-burdened cDNA se-
tquences described above represent functional mRNAs and
which represent mRNA precursors. The likelihood of the lat-
er explanation increases as the number of upstream AUG
codons increases: cDNAs with a dozen or so upstream AUG
triplets (45, 85, 140, 151a, 161, 256, 309, 315, 421) almost
certainly do not represent translatable transcripts! The im-
portance of using appropriate primers to search for alterna-
tive 5′ noncoding sequences cannot be overemphasized.
Positioning a primer near the 5′ end of the longest cDNA (85,
260) will nicely pinpoint the start site of the longest tran-
script, but alternative mRNAs with shorter leader sequences
inevitably will be missed unless the primer is positioned
close to the AUG initiator codon. The fact that so many
AUG-burdened 5′-noncoding sequences have already been
traced to retained introns or to other irregularities (documented above and in reference 193) encourages the
view that many of the cDNAs in Table III may correspond
to nonfunctional transcripts rather than to functional
mRNAs. On the other hand, when the correct form of the
mRNA is eventually deduced, genes thereby eliminated
from Table III often move into Table IV! One way or another,
the mRNAs for oncogenes, transcription factors, growth
factors, etc., seem destined to be translated poorly.

For GC-burdened cDNAs, the solution of switching
promoters to produce an alternative, less encumbered 5′ non-
coding sequence has been documented in only a few cases
(178, 283, 317, 377). Because alternative leader sequences
can easily be missed, as mentioned above, their frequency
might be higher than presently appears. Nevertheless, be-
cause of the consistency with which GC-rich leader se-
quences occur, it seems farfetched to argue that most of those
cDNAs (Table IV) derive from nonfunctional transcripts
rather than from functional mRNAs. Unlike the tabulation
of AUG-burdened leader sequences, which in time tends to
be whittled down by corrections, the tabulation of GC-
burdened leader sequences keeps growing. It includes
mRNAs for many cytoskeletal and housekeeping proteins in
addition to the regulatory proteins mentioned in Table IV.
Thus, it seems likely that some (probably many) vertebrate
mRNAs have enough secondary structure at the 5′ end to
throttle translation. The biggest uncertainty may be whether
these mRNAs invariably are translated poorly or whether
their translation is "derepressed" in response to mitogens, for
example, by modifications of the translational machinery or
induction of helicases. (Although numerous modifications of
the translational machinery correlate with a serum-induced
increase in translation, no causal connection has yet been
established.) The widespread occurrence of 5′ noncoding
sequences that appear unfavorable for translation might be
rationalized by the ability of GC motifs to promote transcrip-
tion. In some vertebrate genes, sequence elements located
downstream from the cap site indeed augment the yield of
mRNA (51, 97, 131, 131a, 163, 210, 327, 392).

Whatever the explanation, the encumbered leader se-
tquences described herein represent a minority of the ver-
tebrate cDNA sequences that have been analyzed, and they
are a distinctly nonrandom set. AUG-burdened/GC-rich
leader sequences virtually never occur on mRNAs that en-
code globins, albumins, caseins, immunoglobulins, histo-
tones, or other highly expressed proteins. The fact that genes
for growth factors, cytokine receptors, proto-oncogenes,
etc., often produce transcripts with encumbered 5′ non-
coding sequences suggests extensive regulation of the regula-
tors, at the level of translation and/or RNA processing.

Research in the author's laboratory is supported by the National Institutes
of Health (grant GM-33915).

This paper is dedicated to Judith Shakespeare.

Received for publication 8 May 1991 and in revised form 12 July 1991.

References
1. Abraham, J. A., J. L. Whang, A. Tumolo, A. Mergia, J. Friedman, D.
Gospodarowicz, and J. C. Fiddes. 1986. Human basic fibroblast growth
factor: nucleotide sequence and genomic organization. EMBO (Eur.
Mol. Biol. Organ.) J. 5:2523–2528.
2. Acampora, D., M. D'Esposito, A. Faiella, M. Pannese, E. Migliaccio,
F. Morelli, A. Stornaiuolo, V. Nigro, A. Simeone, and E. Bonniei.
1989. The human HOX gene family. Nucleic Acids Res. 17:10385–
10402.
3. Acland, P., M. Dixon, G. Peters, and C. Dickson. 1990. Subcellular fate
of the int-2 oncogene is determined by choice of initiation codon. Na-
ture (Lond.). 343:662–665.
4. Alcorta, D. A., C. Crews, L. Sweet, L. Baeckston, S. Jones, and R. L.
Erikson. 1989. Sequence and expression of chicken and mouse rsk:

The Journal of Cell Biology, Volume 115, 1991
Kozak An Analysis of Vertebrate mRNA Sequences

28. Boulter, J., A.O’Shea-Greenfield, R. Duvoisin, J. Connolly, E. Wada, A. Jensen, P. Gardner, M. Ballivet, E. S. Deneris, D. McKinnon, S. Heinemann, and J. Patrick. 1990. Three members of the rat neuronal nicotinic acetylcholine receptor gene family form a gene cluster. J. Biol. Chem. 265:4472–4482.

29. Boulton, T. G., S. H. Nye, D. Robbins, N. Y. Ip, E. Radziejewska, S. Morgenbesser, R. DePinho, N. Panayotatos, M. Cobb, and G. D. Yancopolous. 1991. ERKs: a family of protein-tyrosine/threonine kinases that are activated and tyrosine phosphorylated in response to insulin and NGF. Cell 65:663–675.

30. Brand, N. J., M. Petkovich, and P. Chambon. 1990. Characterization of a functional promoter for the human retinoic acid receptor-alpha (hRAR- 

31. Breier, G. R. Dressler, and P. Gruss. 1988. Primary structure and development expression pattern of the hox 3.1, a member of the murine Hox 3 homeobox gene cluster. EMBO (Eur. Mol. Biol. Organ.) J. 7:1329–1335.

32. Brenner, M., K. Lampe, Y. Nakatani, J. Mill, C. Banner, K. Mearrow, M. Dohadwala, R. Lipsky, and E. Freese. 1990. Characterization of human cDNA and genomic clones for glial fibrillary acidic protein. Mol. Brain Res. 7:277–286.

33. Brinkman, A., C. Groffen, D. J. Kortlieve, A. G. van Kessel, and S. L. S. Drop. 1988. Isolation and characterization of a cDNA encoding the low molecular weight insulin-like growth factor binding protein (IBP-1). EMBO (Eur. Mol. Biol. Organ.) J. 7:2417–2423.

34. Brookes, S., R. Smith, G. Casey, C. Dickson, and G. Peters. 1989. Sequence organization of the human int-2 gene and its expression in teratocarcinoma cells. 1989. Oncogene. 4:429–436.

35. Brown, A. L., and M. M. Rechler. 1990. Cloning of the rat insulin-like growth factor-binding protein-2 gene and identification of a functional promoter lacking a TATA box. Mol. Endocrinol. 4:2039–2051.

36. Brown, J. W., S. G. and F. Feix. 1990. A functional splice site in the 5′ untranslated region of a novel gene. Nucleic Acids Res. 18:2121–2125.

37. Brumm, M., F. Faroni, E. Saul, and L. E. Rosenberg. 1991. Sequence analysis, biosynthesis, and mitochondrial import of the α-subunit of rat liver proopionyl-CoA carboxylase. (Correction) J. Biol. Chem. 266:14660–14666.

38. Buchberg, A. M., H. G. Bedigan, N. A. Jenkins, and N. G. Copeland. 1990. Evi-2, a common integration site in murine myeloid leukemogenesis. Mol. Cell. Biol. 10:4658–4666.

39. Bukler, A. J., J. Pelletier, D. Habb, T. Ginger, and D. E. Housman. 1991. Isolation, characterization, and expression of the murine Wilms' tumor gene (WTI) during kidney development. Mol. Cell Biol. 11:1707–1712.

40. Bugler, B., F. Amalic, and H. Prats. 1991. Alternative initiation of translation determines cytoplasmic or nuclear localization of basic fibroblast growth factor. Mol. Cell. Biol. 11:573–577.

41. Burgess, W. H., C. A. Diene, J. Kaplow, R. Mudel, R. Friesele, A. Zilberstein, J. Schlessinger, and M. Jaye. 1990. Characterization and cDNA cloning of phospholipase C-γ, a major substrate for heparin-binding growth factor 1 (acidic FGF)-activated tyrosine kinase. Mol. Cell. Biol. 10:4770–4777.

42. Callis, J., M. Fromm, and V. Wallot. 1987. Intromuse increase gene expression in cultured maize cells. Genes and Dev. 1:1183–1200.

43. Carlson, M., R. Tausig, S. Kustu, and D. Botstein. 1983. The secreted prosaposin gene family forms a gene cluster. Cell 33:672–676.

44. Carlson, M., R. Tausig, S. Kustu, and D. Botstein. 1983. The secreted prosaposin gene family forms a gene cluster. Cell 33:672–676.

45. Carlson, M., R. Tausig, S. Kustu, and D. Botstein. 1983. The secreted prosaposin gene family forms a gene cluster. Cell 33:672–676.

46. Carlson, M., R. Tausig, S. Kustu, and D. Botstein. 1983. The secreted prosaposin gene family forms a gene cluster. Cell 33:672–676.

47. Carlson, M., R. Tausig, S. Kustu, and D. Botstein. 1983. The secreted prosaposin gene family forms a gene cluster. Cell 33:672–676.

48. Carlson, M., R. Tausig, S. Kustu, and D. Botstein. 1983. The secreted prosaposin gene family forms a gene cluster. Cell 33:672–676.

49. Carlson, M., R. Tausig, S. Kustu, and D. Botstein. 1983. The secreted prosaposin gene family forms a gene cluster. Cell 33:672–676.
69. Conboy, J., Y. W. Kan, S. B. Shohet, and N. Mohandas. 1986. Molecular cloning of a protein with zinc finger sequences. Proc. Natl. Acad. Sci. USA. 83:7857-7861.

68. Chiu, A. O., and U. Gubler. 1989. Sequence of the cDNA for the human homolog of the mouse c-fms receptor. Biochem. Biophys. Res. Commun. 167:594-598.

67. Chua, A. A., and S. C. Tan. 1989. Alternative splicing of the human porphobilinogen deaminase genes results in the generation of two forms of mRNA coding for human heparin-binding protein. Proc. Natl. Acad. Sci. USA. 86:209-213.

66. Chin, I-M., W-P. Wang, and K. Lehtoma. 1990. Alternative splicing of the human TIMP gene in transfected cells: regulation at the level of elongation. Proc. Natl. Acad. Sci. USA. 84:6995-6999.

65. Christy, R. J., K. H. Kaestner, D. E. Geiman, and M. D. Lane. 1991. Tissue- and development-specific alternative RNA splicing regulates expression of mouse ribosomal protein gene rpL32. 1989. Mol. Cell. Biol. 9:2075-2082.

64. Clarke, B. E., D. V. Sangar, S. E. Newton, A. R. Carter, and K. Lehtoma. 1990. Alternative splicing of the human homolog of the mouse c-fms receptor. Biochem. Biophys. Res. Commun. 167:594-598.

63. Chung, S., and R. P. Perry. 1989. Intron- and intronless forms of introns for expression of mouse ribosomal protein gene rpL32. 1989. Mol. Cell. Biol. 9:2075-2082.

62. Chua, A. O., and U. Gubler. 1989. Sequence of the cDNA for the human homolog of the mouse c-fms receptor. Biochem. Biophys. Res. Commun. 167:594-598.

61. Christy, R. J., K. H. Kaestner, D. E. Geiman, and M. D. Lane. 1991. Tissue- and development-specific alternative RNA splicing regulates expression of mouse ribosomal protein gene rpL32. 1989. Mol. Cell. Biol. 9:2075-2082.

60. Chung, S., and R. P. Perry. 1989. Intron- and intronless forms of introns for expression of mouse ribosomal protein gene rpL32. 1989. Mol. Cell. Biol. 9:2075-2082.

59. Chua, A. O., and U. Gubler. 1989. Sequence of the cDNA for the human homolog of the mouse c-fms receptor. Biochem. Biophys. Res. Commun. 167:594-598.

58. Chrivia, J. C., M. D. Uhler, and G. S. McKnight. 1988. Characterization of cloning cDNA for the Cox and Csb subunits of mouse CAM-dependent protein kinase. J. Biol. Chem. 263:5739-5744.

57. Christy, R. J., K. H. Kaestner, D. E. Geiman, and M. D. Lane. 1991. Tissue- and development-specific alternative RNA splicing regulates expression of mouse ribosomal protein gene rpL32. 1989. Mol. Cell. Biol. 9:2075-2082.

56. Chretien, S., A. Dubart, D. Beaupain, N. Raich, B. Grandchamp, J. N. Burroughs, and S. E. Newton. 1990. Identification of the serum-responsive transcription factor 2. Nature (Land.). 296:93-95.

55. Chrestien, S., A. Dubart, D. Beaupain, N. Raich, B. Grandchamp, J. N. Burroughs, and S. E. Newton. 1990. Identification of the serum-responsive transcription factor 2. Nature (Land.). 296:93-95.

54. Chin, I-M., W-P. Wang, and K. Lehtoma. 1990. Alternative splicing of the human TIMP gene in transfected cells: regulation at the level of elongation. Proc. Natl. Acad. Sci. USA. 84:6995-6999.

53. Chua, A. O., and U. Gubler. 1989. Sequence of the cDNA for the human homolog of the mouse c-fms receptor. Biochem. Biophys. Res. Commun. 167:594-598.

52. Christy, R. J., K. H. Kaestner, D. E. Geiman, and M. D. Lane. 1991. Tissue- and development-specific alternative RNA splicing regulates expression of mouse ribosomal protein gene rpL32. 1989. Mol. Cell. Biol. 9:2075-2082.

51. Christy, R. J., K. H. Kaestner, D. E. Geiman, and M. D. Lane. 1991. Tissue- and development-specific alternative RNA splicing regulates expression of mouse ribosomal protein gene rpL32. 1989. Mol. Cell. Biol. 9:2075-2082.

50. Christy, R. J., K. H. Kaestner, D. E. Geiman, and M. D. Lane. 1991. Tissue- and development-specific alternative RNA splicing regulates expression of mouse ribosomal protein gene rpL32. 1989. Mol. Cell. Biol. 9:2075-2082.

49. Christy, R. J., K. H. Kaestner, D. E. Geiman, and M. D. Lane. 1991. Tissue- and development-specific alternative RNA splicing regulates expression of mouse ribosomal protein gene rpL32. 1989. Mol. Cell. Biol. 9:2075-2082.

48. Christy, R. J., K. H. Kaestner, D. E. Geiman, and M. D. Lane. 1991. Tissue- and development-specific alternative RNA splicing regulates expression of mouse ribosomal protein gene rpL32. 1989. Mol. Cell. Biol. 9:2075-2082.
103. Florkiewicz, R. Z., and A. Sommer. 1989. Human basic fibroblast growth factor gene encodes four polypeptides: three initiate translation from non-AUG codons. Proc. Natl. Acad. Sci. USA. 86:3978-3981.

104. Fong, H. K. W., T. T. Amatruda, B. W. Birren, and M. I. Simon. 1987. Distinct forms of the β subunit of c-fos, a cAMP-binding regulatory protein, are identified by molecular cloning. Proc. Natl. Acad. Sci. USA. 84:3792-3796.

105. Hoyt, H. L., D. LeRoith, and C. T. Roberts, Jr. 1991. Differential association ofinsulin-like growth factor I mRNA variants with polyosomes in vivo. J. Biol. Chem. 266:7300-7305.

106. Fraser, P., P. Cummings, and P. Curtis. 1989. The mouse carbonic anhydrase I gene contains two tissue-specific promoters. Mol. Cell. Biol. 9:3386-3393.

107. Froussard, P., P. Chatagner, G. Soumne, A. Abadie, W. Greene, J. Theze, and S. Longacre. 1988. p55 IL-2 receptor mRNA precursors in murine T lymphocyte nuclei. J. Immunol. 141:1358-1364.

108. Fukunaga, R., E. Ishizaka-Beda, Y. Seto, and S. Nagata. 1990. Expression cloning of a receptor for murine granulocyte colony-stimulating factor. Cell. 61:341-350.

109. Fuller, F., J. G. Porter, A. E. Arfsten, J. Miller, J. S. Schilling, R. M. Scarr, R. W. Brown, and T. F. Osborne. 1988. Multiplegenes encode nuclear factor 1-like proteins that bind to the promoter for 3-hydroxy-3-methylglutaryl-coenzyme A reductase. Proc. Natl. Acad. Sci. USA. 85:8963-8967.

110. Fuller, F., J. G. Porter, A. E. Arfsten, J. Miller, J. S. Schilling, R. M. Scarr, R. W. Brown, and T. F. Osborne. 1988. Multiple genes encode nuclear factor 1-like proteins that bind to the promoter for 3-hydroxy-3-methylglutaryl-coenzyme A reductase. Proc. Natl. Acad. Sci. USA. 85:8963-8967.

111. Gerard, N. P., R. L. Eddy, Jr., T. B. Shows, and C. Gerard. 1990. The human neurokinin A receptor cDNA. Proc. Natl. Acad. Sci. USA. 87:5983-5987.

112. Gerland, G., V. Schmieden, P. R. Schofield, P. H. Seeburg, T. Siderovski, and G. Meier. 1990. Beta chain gene: generation of three receptor forms by cloned human α and β chain cDNAs. Science (Wash. DC). 244:551-556.

113. Haire, J., E. S. Gutoff, J. E. Krause, and T. J. Ley. 1990. Gavras. 1991. Cloning and expression of arat brain a2B-adrenergic receptor. J. Biol. Chem. 266:7300-7305.

114. Hagger, J. A., L. Lewicki, and D. B. Schenk. 1988. Atrial natriuretic peptide cDNA. Proc. Natl. Acad. Sci. USA. 85:10822-10826.

115. Hagiwara, K., G. Stenman, H. Honda, P. Sahlin, A. Andersson, K. Miyagawa, K. Naito, H. Sakamoto, O. Kato, T. Yoshida, T. Sugimura, and M. Terada. 1990. K-sam, an amplified gene in stomach cancer, is a member of the heparin-binding growth factor receptor gene family. Proc. Natl. Acad. Sci. USA. 87:5893-5897.

116. Hannon, R., T. Evans, M. J. Taylor, and H. Gould. 1991. Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. Proc. Natl. Acad. Sci. USA. 88:3004-3008.

117. Hao, Q.-L., N. Heisterkamp, and J. Groffen. 1989. Isolation and sequence analysis of a novel human tissue keratin gene. Mol. Cell. Biol. 9:1587-1593.

118. Harirahana, N. D. E., and K. R. Perry. 1989. Equine mouse ribosomal protein promoters have a similar architecture that includes internal sequence elements. Genes & Dev. 3:1789-1800.

119. Hartig, E., I. F. Loncarevic, M. Bütcher, P. Herrlich, and H. J. Rahmsdorf. 1991. A new cAMP response element in the transcribed region of the human c-fos gene. Nucleic Acids Res. 19:4153-4159.

120. Hatakeyama, M., M. Tsudo, S. Minamoto, T. Kono, T. Doi, T. Miyata, M. Miyasaka, and T. Taniguchi. 1989. Interleukin-2 receptor β chain gene: generation of three receptor forms by cloned human α and β chain cDNAs. Science (Wash. DC). 244:551-556.

121. Haub, O., B. Drucker, and M. Goldfarb. 1990. Expression of the murine fibroblast growth factor gene in the adult central nervous system. Proc. Natl. Acad. Sci. USA. 87:8022-8026.

122. Hillarp, A., and B. Dahlberg. 1990. Cloning of cDNA coding for the β chain of human complement component C4β-binding protein: sequence homology with the α chain. Proc. Natl. Acad. Sci. USA. 87:1183-1187.

123. Hirai, H., Y. Muru, K. Hagiwara, J. Nishida, and F. Takeda. 1987. Identification, cloning, and expression of a cytosolic megakaryocyte ribosomal protein 4.1. Proc. Natl. Acad. Sci. USA. 84:8497-8501.

124. Hof, M. A., D. H. Leininger, M. A. Holtzman, J. M. Medoff, and J. M. Leiden. 1991. Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor α gene. Proc. Natl. Acad. Sci. USA. 88:10118-10122.

125. Hoffman, E. K., S. P. Trusko, N. Freeman, and D. L. George. 1987. Structural and functional characterization of the promoter region of the mouse K-sam receptor gene. Mol. Cell. Biol. 7:2592-2596.

126. Holmsen, C. L., G. P. Lomonosoff, D. Evans, and J. A. Maule. 1989. Identification of the leader codons for translation of cowpea mosaics virus middle component RNA using site-directed mutagenesis and an infectious cDNA clone. Virol. 172:311-320.

127. Holmstam, D. A., W. D. Cook, and R. Dunn. 1987. Isolation and sequence of a cDNA corresponding to a related gene expressed in murine hemopoietic cells. Proc. Natl. Acad. Sci. USA. 84:8325-8329.

128. Holzer, K. P., W. Liu, and G. G. Hammes. 1989. Identification oftheinitiationcodonsfortranslationofcowpeamosaicvirus middle component RNA using site-directed mutagenesis of an infectious cDNA clone. Virol. 172:311-320.

129. Honkawa, H., W. Masahashi, S. Hashimoto, and T. Hashimoto-Gotoh. 1989. Identification of the initiation codons for translation of cowpea mosaics virus middle component RNA using site-directed mutagenesis of an infectious cDNA clone. Virol. 172:311-320.

130. Honken, T., W. Miyashita, K. Kojoh, J. H. Hassell, and Y. Yamada. 1987. Identification of a phenotype-specific enhancer in the first intron of the rat collagen II gene. Proc. Natl. Acad. Sci. USA. 84:8864-8868.

131. Howard, T. E., S. Y. Shaw, K. G. Langford, B. M. Martin, and K. E. Bern-
stein. 1990. Transcription of tissuerespective antisense-converting enzyme (ACE) is initiated within the 12th intron of the somatic ACE gene. Mol. Cell. Biol. 10:373-383.

151a. Thomas, C. A. A. Kozak, A. Chakraborti, R. N. Yewdell, G. G. Bennett, and A. S. Marchal. 1989. Circumstances and mechanism of inhibition of translation by eukaryotic ribosomes. In vitro. J. Biol. Chem. 265:131-135.

151. Thomas, C. A. A. Kozak, A. Chakraborti, R. N. Yewdell, G. G. Bennett, and A. S. Marchal. 1989. Circumstances and mechanism of inhibition of translation by eukaryotic ribosomes. J. Biol. Chem. 264:19100-19105.

151. Thomas, C. A. A. Kozak, A. Chakraborti, R. N. Yewdell, G. G. Bennett, and A. S. Marchal. 1989. Circumstances and mechanism of inhibition of translation by eukaryotic ribosomes. J. Biol. Chem. 264:19100-19105.

151. Thomas, C. A. A. Kozak, A. Chakraborti, R. N. Yewdell, G. G. Bennett, and A. S. Marchal. 1989. Circumstances and mechanism of inhibition of translation by eukaryotic ribosomes. J. Biol. Chem. 264:19100-19105.
203. Krust, A., S. Green, P. Argos, V. Kumar, P. Walter, J-M. Bornert, and P. Chambon. 1986. The chicken oestrogen receptor sequence: homology with the human receptor and the human oestrogen and glucocorticoid receptor. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:891–897.

204. Krust, A., P. Kastner, M. Petkovich, A. Zelent, and P. Chambon. 1989. A third human retinoid acid receptor, hRAR-y. *Proc. Natl. Acad. Sci. USA.* 86:5310–5314.

205. Lavoie, F., J. S. Taylor, D. M. W. Sporn, S. Gimpel, R. J. Jerzy, L. Brunton, K. H. Hjerrild, D. Cosman, and G. R. Martin. 1988. Four classes of mRNA are expressed from two widely separated promoter regions. *Cell.* 54:1015–1022.

206. Lavoie, F., J. S. Taylor, D. M. W. Sporn, S. Gimpel, R. J. Jerzy, L. Brunton, K. H. Hjerrild, D. Cosman, and G. R. Martin. 1988. Four classes of mRNA are expressed from two widely separated promoter regions. *Cell.* 54:1015–1022.

207. Lavoie, F., J. S. Taylor, D. M. W. Sporn, S. Gimpel, R. J. Jerzy, L. Brunton, K. H. Hjerrild, D. Cosman, and G. R. Martin. 1988. Four classes of mRNA are expressed from two widely separated promoter regions. *Cell.* 54:1015–1022.

208. Lavoie, F., J. S. Taylor, D. M. W. Sporn, S. Gimpel, R. J. Jerzy, L. Brunton, K. H. Hjerrild, D. Cosman, and G. R. Martin. 1988. Four classes of mRNA are expressed from two widely separated promoter regions. *Cell.* 54:1015–1022.

209. Lavoie, F., J. S. Taylor, D. M. W. Sporn, S. Gimpel, R. J. Jerzy, L. Brunton, K. H. Hjerrild, D. Cosman, and G. R. Martin. 1988. Four classes of mRNA are expressed from two widely separated promoter regions. *Cell.* 54:1015–1022.

210. Lafyatis, R., R. Lechleider, A. B. Roberts, and M. B. Sporn. 1990. Regulation of rat ornithine decarboxylase by utilization of alternative protein coding sequences of the cDNA. *Proc. Natl. Acad. Sci. USA.* 87:8034–8038.

211. Lazzaro, M. A., R. A. Houdin, D. S. Darling, and W. W. Chin. 1989. Heterogeneity at the 5' end of rat acetyl-CoA carboxylase mRNA. *J. Biol. Chem.* 263:1711–1718.

212. Leveille, G. J., T. E. Costello, M. J. Emerman, S. R. Coates, and J. S. Waters. 1986. Nucleotide sequence of the cDNA for a spliceless RNA transcript encoding human retinoic acid receptor. *Cell.* 43:899–908.

213. Levin, A., J. J. Wojcierowski, S. Wu, J. Pyrc, A. Ross, B. Dietz-Schold, D. Speicher, and H. Kopolowski. 1989. Sequence investigation of the major gastrointestinal tumour-associated antigen gene family, GAT373. *Proc. Natl. Acad. Sci. USA.* 86:27–31.

214. Liu, F., and B. Roizman. 1991. The promoter, transcriptional unit, and coding sequence of herpes simplex virus 1 family 35 proteins are contained within and in-frame with the U26 ORF of *J. Virol.* 65:206–212.

215. Liu, F.-T., K. Albrandt, and M. W. Robertson. 1988. cDNA heterogeneity suggests structural variants related to the high-affinity IgE receptor. *Proc. Natl. Acad. Sci. USA.* 85:5690–5694.

216. Livsnyanski, G., P. Koznetz, M. McHale, L. Cieslinski, G. Sabel, D. Taylor, R. Davis, T. Torphy, and J. B. Macbean. 1990. Cloning and expression of cDNAs for a human low-Km, rolipram-sensitive cyclic AMP phosphodiesterase. *J. Biol. Chem.* 265:486–496.

217. Lippert, P., D. Nash, and S. Korndorf. 1988. Cloning and sequence analysis of the cation-independent mannose 6-phosphate receptor. *J. Biol. Chem.* 263:2563–2570.

218. Lock, P. E., A. L. Holtzman, and A. R. Dunn. 1990. Functional analysis and nucleotide sequence of the promoter region of the mouse inositol kinase gene. *Cell Biol.* 10:4603–4611.

219. Lock, P., S. Ralph, E. Stanley, I. Boulet, R. Ramsay, and A. R. Dunn. 1991. Two isoforms of murine inositol kinase generated by utilization of alternative initiation codons, exhibit different patterns of subcellular localization. *Cell Biol.* 11:4363–4370.

220. Locke, S. J., L. Coteochea, W. Lorenz, W. Leung, D. Schwinn, T. Yang-Peng, M. Brownstein, R. Leffkowitz, and M. G. Caron. 1990. Molecular cloning and expression of the cDNA for the ala-adrenergic receptor. *J. Biol. Chem.* 265:6336–6369.

221. Leung, D. W., S. A. Spencer, G. Cachianes, R. G. Hammonds, C. Colvin, and D. C. Goodwin. 1989. Characterization of the human and murine IL-7 genes. *J. Immunol.* 144:3592–3601.

222. L¨uken, M. H., and D. S. Kirkland. 1989. Two N-myc polypeptides with distinct amino termini encoded by the second and third exons of the gene. *Cell.* 9:1545–1552.

223. Livi, G. P., P. Koznetz, M. McHale, L. Cieslinski, G. Sabel, D. Taylor, R. Davis, T. Torphy, and J. B. Macbean. 1990. Cloning and expression of cDNAs for a human low-Km, rolipram-sensitive cyclic AMP phosphodiesterase. *J. Biol. Chem.* 265:486–496.

224. Maldonado, F., and S. K. Hanks. 1988. A cDNA clone encoding human c-AMP-dependent protein kinase catalytic subunit cCA. *Nucleic Acids Res.* 16:8189.

225. Mali, M., P. Jakkola, A-M. Arvilommi, and M. Jalkanen. 1990. Sequence of human syndecan indicates a novel family of integral membrane proteoglycans. *J. Biol. Chem.* 265:6884–6889.

226. Marquedant, R., B. N. Harvey-Thomson, M. O. Marquedant, and K. H. Hjerrild. 1989. Epitope-directed generation of mouse thymidine kinase with altered specificities generated by complement components. *J. Immunol.* 144:3592–3601.

227. Mansen, J. M., and P. Blackshear. 1990. Regulation of rat ornithine decarboxylase mRNA translation by its 5' untranslated region. *J. Biol. Chem.* 265:11817–11822.

228. Martin, M. D., B. Mitchell, G. W. Both, and M. K. Estes. 1991. Expression of renin genes in the rat and functional characterization of the promoter of the ET2, the human c-ets-2 gene. *Proc. Natl. Acad. Sci. USA.* 88:1047–1051.

229. McCarty, T. J., and S. L. Hunkapiller. 1985. The preferential translation of Drosophila hsp70 mRNA requires sequences in the untranslated leader. *Cell.* 42:901–912.

---

Kozak An Analysis of Vertebrate mRNA Sequences
Messer, G., U. Spengler, M. C. Jung, G. Honold, K. Blomer, G. R. Pape, N. Meerovitch, K., J. Pelletier, and N. Sonenberg. 1989. A cellular protein that binds to the 5' noncoding region of poliovirus RNA: Implications for internal translation initiation. *Genes & Dev.* 3:1026–1034.

Mejier, D., A. Graus, R. Kraay, A. Langeveld, M. P. Mulder, and G. Grosveld. 1990. The octamer binding factor Oct: cDNA cloning and expression in early embryonic nuclei. *Nucl. Acids Res.* 18:7373–7385.

Mellentin, J., D. S. D. Smith, and M. L. Cleary. 1989. Ifl-1, a novel gene encoded by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. *Cell.* 58:77–83.

Mendel, D. R., P. Hansen, M. C. Jung, A. van Loon, and G. R. Crabtree. 1991. HNF-1α and HNF-1β (HNF-1) share dimerization and homoeodomain but not activation domains. *Genes & Dev.* 5:1042–1056.

Messer, G., U. Spengler, M. C. Jung, G. Honold, K. Blomer, G. R. Pape, G. Riehemuller, and E. H. Weiss. 1991. Polymorphic structure of the TNF locus. *Exp. Med.* 173:209–219.

Meyer, R., E. Hatada, H. P. Hofmann, M. Haiker, C. Bartsch, U. Rothlisberger, H. W. Lahr, E. Schlegel, A. Van Loon, and C. Scheidereit. 1991. Cloning of the DNA-binding subunit of human UBF. *Proc. Natl. Acad. Sci. USA.* 88:966–970.

Mills, J. S., A. W. McLaren, and C. R. Wolf. 1989. Alternative splicing in the human cytochrome P450B6 gene generates a high level of aberrant messages. *Nucleic Acids Res.* 17:8241–8255.

Miller, R. T., L. Couillon, G. Pages, R. P. Lifton, C. Sardet, and J. Pouyssegur. 1991. Structure of the 5' flanking regulatory region and gene for the liver cell factor-activatable NuH exchanger. *J. Biol. Chem.* 266:10813–10819.

Mirkov, T., E. M. Mathews, D. H. Du Plessis, and J. A. Dodds. 1989. Nucleotide sequence and translation of satellite tobacco mosaic virus RNA. *Virology.* 170:134–146.

Misrahi, M., H. Loosfelt, M. Atger, C. Meriel, V. Zerah, P. Dessen, and A. van Loon. 1991. Cloning of the DNA-binding subunit of human nu-line. *Biochem. Biophys. Res. Commun.* 178:751–759.

Miyajima, N., K. E. Bernstein. 1991. Isolation of a cDNA encoding the vascular type-1 angiotensin II receptor. *Nature (Land.).* 351:233–236.
adrenergic receptor gene. *Nucleic Acids Res.* 18:4591.

357. Simmons, D. M., J. Voss, H. Hingham, J. Holloway, R. Broide, M. Rosenfeld, and L. W. Swanson. 1990. Pituitary cell phenotypes involve cell-specific mRNA translation and processed interactions with other classes of transcription factors. *Genes Dev.* 4:695–711.

358. Sims, J. C., M. March, D. Cosman, M. Widmer, H. R. MacDonald, C. McMahon, C. Grabin, J. Wignall, J. Jackson, S. Call, D. Friend, A. Alpert, S. O'Hara, D. Urdal, and S. K. Dower. 1988. cDNA expression cloning of the IL-1 receptor, a member of the immunoglobulin superfamily. *Science (Wash. DC).* 241:585–588.

359. Sioni, H., H. Shida, S. H. Nam, T. Nosaka, M. Maki, and M. Hatazaka. 1988. Sequence requirements for nuclear localization of HTLV Type I pX protein, which regulates viral RNA processing. *Cell.* 55:197–209.

360. Slent, D. E., R. Hull, P. C. Turner, and T. M. A. Wilson. 1988. Studies on the mechanism of translational enhancement by the 5′-leader sequence of bovine viral RNA. *EMBO J.* 7:357–364.

361. Snyder, L. C., S. Trusko, N. Freeman, J. R. Eschleman, S. S. Fakharzadeh, and D. L. George. 1988. A gene amplified in a transformed mouse cell line undergoes complex transcriptional processing and encodes a nuclear protein. *J. Biol. Chem.* 263:17150–17158.

362. Sobieszczuk, P. W., T. J. Gonda, and A. R. Dunn. 1989. Structure and biological activity of the translational initiation sequences of the murine c-myb oncogene. *Nucleic Acids Res.* 17:9593–9611.

363. Soldati, T., B. W. Schafer, and J.-C. Perriard. 1990. Alternative ribosomal initiation gives rise to chicken brain-type creatine kinase isoproteins with heterogeneous amino termini. *J. Biol. Chem.* 265:4698–4506.

364. Sopa, M., Z. F. Barton, and J. Greenblatt. 1989. Structure and associated DNA-helicate activity of a general transcription initiation factor that binds to RNA polymerase II. *Nature (London).* 341:410–414.

365. Stetler-Stevenson, W. G., P. D. Brown, M. Onisto, A. Levy, and L. A. Sternberg, E. A., G. Spizz, W. M. Perry, D. Vizard, T. Weil, and E. N. Olson. 1990. The 5′-flanking region of the c-myc gene encodes a cell surface sialyltransferase. *J. Exp. Med.* 164:61–643.

366. Stanton, L. W., and J. M. Bishop. 1987. Alternative processing of RNA transcripts from the MYC promoter. *Mol. Cell Biol.* 7:4266–4272.

367. Stephens, R. M., D. Derse, and N. R. Rice. 1990. Cloning and characterization of cDNAs encoding equine infectious anemia virus tat and putative rev proteins. *J. Virol.* 64:3716–3722.

368. Stenmark, E. A., G. Spizz, W. M. Perry, D. Vizard, T. Weil, and E. N. Olson. 1988. Identification of upstream and intragenic regulatory elements that confer cell-type-restricted and differentiation-specific expression on the muscle creatine kinase gene. *Mol. Cell Biol.* 8:2896–2909.

369. Streuli, C. M., G. Onesto, A. Levy, and L. A. Liotta. 1990. Tissue inhibitor of metalloproteinases-2 (TIMP-2) mRNA expression in tumor cells and human tumor tissues. *J. Biol. Chem.* 265:13933–13938.

370. Stitzkaser, S. C., P. L. Whitfield, D. Chrislie, A. Bellamy, and G. W. Both. 1987. Processing of retroviruses glycoprotein VP7: implications for the retention of the protein in the endoplasmic reticulum. *J. Cell Biol.* 105:2897–2903.

371. Storzberg, E., A. G. Spizz, W. M. Perry, D. Vizard, T. Weil, and E. N. Olson. 1988. Identification of upstream and intragenic regulatory elements that confer cell-type-restricted and differentiation-specific expression on the muscle creatine kinase gene. *Mol. Cell Biol.* 8:2896–2909.

372. Streuli, C. M., G. Onesto, A. Levy, and L. A. Liotta. 1990. Tissue inhibitor of metalloproteinases-2 (TIMP-2) mRNA expression in tumor cells and human tumor tissues. *J. Biol. Chem.* 265:13933–13938.

373. Strickler, S. C., P. L. Whitfield, D. Chrislie, A. Bellamy, and G. W. Both. 1987. Processing of retrovirus glycoprotein VP7: implications for the retention of the protein in the endoplasmic reticulum. *J. Cell Biol.* 105:2897–2903.

374. Streuli, C. M., G. Onesto, A. Levy, and L. A. Liotta. 1990. Tissue inhibitor of metalloproteinases-2 (TIMP-2) mRNA expression in tumor cells and human tumor tissues. *J. Biol. Chem.* 265:13933–13938.

375. Strickler, S. C., P. L. Whitfield, D. Chrislie, A. Bellamy, and G. W. Both. 1987. Processing of retrovirus glycoprotein VP7: implications for the retention of the protein in the endoplasmic reticulum. *J. Cell Biol.* 105:2897–2903.

376. Stone, S. R., J. Hofstenge, and B. A. Hennings. 1987. Molecular cloning of cDNAs encoding two isoforms of the catalytic subunit of protein phosphatase 2A. *Biochemistry.* 26:7215–7220.

377. Straaten, V. F., R. Muller, T. Curran, C. Van Beveren, and I. M. Verma. 1983. Charon and nucleotide sequence of a human c-onc gene: deduced amino acid sequence of the human c-fos protein. *Proc. Natl. Acad. Sci. USA.* 80:3183–3187.

378. Stratki, A. H., I. L. Schwartz, A. Perez-Castillo, and J. H. Oppenheimer. 1989. Relationship of c-erbAmRNA content to tissue triiodothyronine. *Proc. Natl. Acad. Sci. USA.* 80:3183–3187.

379. Suzuki, S., M. Froehlich, M. Scollon, Y. Kasai, Z. Steplewski, H. Givol. 1987. Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional initiation. *Nature (Lond.).* 325:37–57.

380. Suzuki, S., M. Froehlich, M. Scollon, Y. Kasai, Z. Steplewski, H. Givol. 1987. Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional initiation. *Nature (Lond.).* 325:37–57.
408. Watanabe, T., H. Yonehura, K. Terazono, H. Yamamoto, and H. Wang, W-P., K. Lehtoma, M. L. Varban, I. Krishnan, and I-M. Chiu. 1989. Cloning of the gene coding for human class I heparin-binding growth factor and its expression in fetal tissues. Mol. Cell Biol. 9:2387-2395.

409. Waterhouse, P., R. Krajcsi, A. E. Tollefson, L. R. Gooding, K. O. C. Thompson, E. S. Ong, R. Lebo, D. J. Gruol, and J. M. Evans. 1989. Developmentally regulated use of alternative promoters creates a novel PDGF receptor transcript in mouse teratocarcinoma and embryonic stem cells. Mol. Cell Biol. 9:4563-4567.

410. Xie, W., J. G. Chipman, D. L. Robertson, R. L. Erikson, and D. L. Simon. 1991. Expression of a mitogen-responsive gene encoding protaglandin synthase is regulated by mRNA splicing. Proc. Natl. Acad. Sci. USA. 88:2692-2696.

411. Yan, C., and L. Tamm. 1991. Molecular cloning and characterization of interleukin-1β response element binding factors of the murine (2'-5') oligoadenylate synthetase ME-12 gene. Proc. Natl. Acad. Sci. USA. 88:144-148.

412. Young, J. C., M. L. Gishizky, and O. N. Witte. 1991. Hyperepression of interleukin-7 is not necessary or sufficient for transformation of a pre-B lymphoid cell line. Mol. Cell Biol. 11:854-863.

413. Zakany, J., C. K. Tuggle, M. Patel, and M. C. Nguyen-Huu. 1988. Spatial regulation of homeobox gene fusions in the embryonic central nervous system of transgenic mice. Neuron. 1:679-691.

414. Zelent, A., C. Mendelsohn, P. Kastner, A. Krust, J.-M. Garnier, F. Ruffenach, P. Leroy, and P. Chambon. 1991. Differentially expressed isoforms of the mouse retinoic acid receptor β are generated by usage of two promoters and alternative splicing. EMBO (Eur. Mol. Biol. Organ.) J. 10:71-81.

415. Zerial, M., L. Toschi, R.-P. Ryseck, M. Schuermann, R. Muller, and R. Bravo. 1989. The product of a novel growth factor activated gene, fos B, interacts with JUN proteins enhancing their DNA binding activity. EMBO (Eur. Mol. Biol. Organ.) J. 8:805-813.

416. Zon, I. I., S-F. Tsai, S. Burgess, P. Matsudaira, G. Bruns, and S. H. Orkin. 1990. The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. Proc. Natl. Acad. Sci. USA. 87:668-672.

417. Zot, A. S., and D. M. Fambrough. 1990. Structure of a gene for a lysosomal membrane glycoprotein (LEP100). J. Biol. Chem. 265:20988-20995.

418. Youn, J., T. A. panda, M. L. Raymond, H. Beug, S. H. Orkin, and J. D. Engel. 1990. Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. Genes Dev. 4:1650-1662.

419. Yamauchi, Y., S-I. Fukushima, K. Semba, J. Sukewaga, N. Miyajima, K-I. Matsubara, T. Yamamoto, and K. TToyoshima. 1987. The yes-related cellular gene lyn encodes a possible tyrosine kinase related to p561. Mol. Cell Biol. 7:237-243.

420. Yamashita, K., T. Taga, Y. Hirata, H. Yawata, Y. Kawanishi, B. Seed, T. Taniguchi, T. Hirano, and T. Kishimoto. 1988. Cloning and expression of a human interleukin-6 (BSF-2/IFNβ2) receptor. Science (Wash. DC). 241:825-828.

421. Weil, D., S. Brosset, and F. Dautry. 1990. RNA processing is a limiting step for murine tumor necrosis factor β expression in response to interleukin-2. Mol. Cell Biol. 10:5865-5875.

422. Weinberger, A. D., and S. L. Swain. 1990. IL-2 receptor (Tac antigen) protein expression is down-regulated by the 5′ untranslated region of the mRNA. J. Immunol. 144:4712-4720.

423. Weinberger, C., C. C. Thompson, E. S. Ong, R. Lebo, D. J. Gruol, and R. M. Evans. 1986. The c-erh-A gene encodes a thyroid hormone receptor. Nature (Lond.). 324:641-646.

424. Williams, M. A., and R. A. Lamb. 1989. Effect of mutations and deletions in a bicistronic mRNA on the synthesis of influenza B virus NB and NA glycoproteins. J. Virol. 63:28-35.

425. Williams, K., T. A. panda, M. L. Raymond, H. Beug, S. H. Orkin, and J. D. Engel. 1990. Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. Genes Dev. 4:1650-1662.