Volume 3 · Number 2 · 1993
GENE EXPRESSION

The two functionally distinct amino termini of chicken c-*ets-1* products arise from alternative promoter usage

Pascale Crepieux, Dominique Leprince, Anne Flourens, Olivier Albagli, Elisabeth Ferreira, and Dominique Stéhelin

Unité d'Oncologie Moleculaire, CNRS URA 1160, Institut Pasteur, Lille, France

The chicken c-*ets-1* locus gives rise to two distinct transcription factors differing by structurally and functionally unrelated N-termini. $p54^{c\text{-}ets\text{-}1}$ shows a striking phylogenetic conservation from Xenopus to humans, while $p68^{c\text{-}ets\text{-}1}$, the cellular counterpart of the E26-derived v-*ets* oncogene, is apparently restricted to avian and reptilian species. In the chick embryo, both mRNAs are expressed in a wide array of tissues of mesodermal origin; however, in the embryo and after hatching, $p68^{c\text{-}ets\text{-}1}$ is excluded from lymphoid cells where $p54^{c\text{-}ets\text{-}1}$ accumulates.

In this report, we define the basis of the differential expression of the chicken c-*ets-1* products to assess their different potentials as transcription factors. We demonstrate that the two distinct N-termini arise from alternative promoter usage within the chicken c-*ets-1* locus. Examination of both promoters reveals that transcription initiates from multiple sites, consistent with the absence of TATA and CAAT elements. Of these two regulatory regions, only the one that initiates the $p54^{c\text{-}ets\text{-}1}$ mRNA synthesis is of the G+C-rich type, and its organization is conserved in humans. The avian-specific $p68^{c\text{-}ets\text{-}1}$ promoter activity was enhanced by its own product. In addition, we identify numerous potential binding sites for lymphoid-specific transcription factors that might contribute to a tight repressor effect in lymphoid tissues.

The *ets* gene family groups a number of related transcription factors that are conserved from Drosophila to humans (Lautenberger et al., 1992 ; MacLeod et al., 1992; Laudet et al., 1993). This family is defined by the presence of a new type of DNA-binding domain, the ETS domain (Karim et al., 1990), most often located at the carboxy-terminus of the protein with some exceptions, including the elk-1 (Rao et al., 1989), elf-1 (Thompson et al., 1992), and SAP-1 (Dalton and Treisman, 1992) proteins. The *ets* family members exert their function as transcription factors by interacting in a sequence-specific manner with purine-rich motifs they recognize with variable affinities (Wang et al., 1992) in the promoters and enhancers of several viral and cellular genes (Gunther et al., 1990; Ho et al., 1990; Wasylyk et al., 1990; Virbasius and Scarpulla, 1991; Wasylyk et al., 1991). Nevertheless, a dual function for *ets* family members has recently come to light from the observation that some of them can form ternary complexes, in association with unrelated transcriptional factors such as the SRF (Hipskind et al., 1991; Dalton and Treisman, 1992).

The roles of ets-related protein regions out-

Received May 7, 1993; revision accepted July 7, 1993.

Correspondence: Dominique Leprince, CNRS URA 1160, Institut Pasteur, 1 rue Calmette, 59019 Lille, France Tel 33-20-87-78-79 Fax 33-20-87-79-08

 1052-2166/93/0302/215-11$2.00

side of the DNA-binding domain still remain largely unclear, though they were partly elicited by recent studies on the avian proteins produced by the c-*ets-1* gene (Schneikert et al., 1992). The chicken c-*ets-1* locus, identified as the cellular counterpart of the E26 v-*ets* oncogene (Leprince et al., 1983; Nunn et al., 1983), was the first characterized member of the *ets* gene family. This locus gives rise to two different transcription factors, $p54^{c\text{-}ets\text{-}1}$ and $p68^{c\text{-}ets\text{-}1}$, which differ only by unrelated N-termini, respectively encoded by a single exon absent from v-*ets*, I^{54}, or by the α and β exons, homologous to the 5′ part of v-*ets* (Leprince et al., 1988). These alternative exons are fused to a common set of 3′ exons named a to F. The common exons encode an N-terminal regulatory domain, a transactivating domain, and the C-terminal DNA-binding domain. While the highly hydrophobic α- and β-encoded amino acids consist of an additional transactivating domain, the function of the hydrophilic I^{54}-encoded amino acids remains undefined (Schneikert et al., 1992). We can therefore speculate that each N-terminus interacts with different transcription regulators to fulfill its function, possibly in a cell-specific manner. Consistent with this hypothesis is the fact that $p54^{c\text{-}ets\text{-}1}$ and $p68^{c\text{-}ets\text{-}1}$ display differences in their expression pattern: in chicken, $p54^{c\text{-}ets\text{-}1}$ is widely expressed, with moderate levels in most tissues, but high levels in lymphoid cells (Ghysdael et al., 1986). In contrast, $p68^{c\text{-}ets\text{-}1}$ expression is restricted to a blood vessel-containing fraction of the spleen (Leprince et al., 1990) and other mesodermal tissues such as the embryonic dermis at E6, but remains undetectable in lymphoid tissues (Quéva et al., 1993).

A better understanding of the positive and negative controls that underlie a cell-specific pattern of gene expression requires careful examination of the mechanism regulating the expression of the regulators themselves. In this report, we have initiated such studies by describing the molecular mechanism governing the expression of the chicken c-*ets-1* locus. To unravel the respective regulation of $p54^{c\text{-}ets\text{-}1}$ and $p68^{c\text{-}ets\text{-}1}$ expression, we asked if they arise from alternative splicing of a precursor mRNA initiated at a single promoter, or whether a differential promoter usage would account for the divergences in their expression patterns. The study reported here ascertains the second mechanism.

Materials and methods

Molecular cloning

Isolation of a new chicken c-*ets-1* genomic clone. A recombinant DNA library in the EMBL4 vector constructed with a partial Sau 3A digest of total chicken embryo DNA was used to isolate the promoter region of $p68^{c\text{-}ets\text{-}1}$ mRNA. It was screened first with a 5.4 kbp Hind III genomic probe including the α exon (Gégonne et al., 1987). The positive clones were then counterselected with a labeled oligonucleotide corresponding to the 5′ end of the $p68^{c\text{-}ets\text{-}1}$ cDNA (primer 1: 5′ ACA AGT GTG GGG AGC CGT GGA GGA 3′). We obtained a 14.0 kbp long genomic insert, which is referred to as clone 68.1ETS. The previously described lambda c-*ets* B recombinant phage (Gégonne et al., 1987) containing sequences localized within the 40.0 kbp of genomic DNA including the I^{54} exon was subjected to promoter analysis after restriction enzymes mapping.

Reporter and expression plasmids used in the promoters functional assays. The c-*ets-1* promoter regions and the $p68^{c\text{-}ets\text{-}1}$ promoter deletion mutants were subcloned into the pLUCDSS luciferase reporter plasmid, generously provided by Dr. F. Gouilleux (Gouilleux et al., 1991), who derived it from pSPLUC2 (DeWet et al., 1987). The fragments were inserted either into the Pvu II–Hind III ($p68^{c\text{-}ets\text{-}1}$), or Pvu II ($p54^{c\text{-}ets\text{-}1}$) restriction sites. In order to allow in-frame synthesis of the luciferase protein during the functional assay, the $p54^{c\text{-}ets\text{-}1}$ promoter fragment subcloned into the M13 mp19 vector was mutated in the translation initiation codon (ATG→AAG) according to standard methods (Sayers et al., 1988) using the Amersham oligonucleotide in vitro mutagenesis kit. The ptkFLUC vector (DeWet et al., 1987; Gouilleux et al., 1991) including the thymidine kinase minimal promoter was used as a positive control. The chicken $p68^{c\text{-}ets\text{-}1}$ cDNA (Leprince et al., 1988) was cloned into the EcoR I restriction site of the pSG5 expression vector (Wasylyk et al., 1990).

Sequencing analysis. The c-*ets-1* regulatory regions were sequenced on both strands by the dideoxy termination method (Sanger et al., 1977), after cloning into the polylinker restriction sites of M13mp18 and mp19 derivatives.

Mapping of the *c-ets-1* mRNA 5′ ends

RNase protection analysis was performed as previously described (Leprince et al., 1988) with the use of a Promega kit. A 400 bp α^{32}P-CTP labeled $p68^{c\text{-}ets\text{-}1}$ RNA probe was transcribed from a corresponding BamH I–Hind III DNA fragment cloned into the pSP64 vector (Fig. 4). The probe was hybridized at 50°C to 20 μg of total RNA from chicken spleen, thymus, or embryonic dermis extracted using the CsCl/guanidinium isothiocyanate method (Maniatis et al., 1982). The RNA–RNA hybrids were digested by a mixture of T1 (Gibco BRL) and A (Boehringer Mannheim) RNases, and the resulting products were electrophoresed on a 6% sequencing acrylamide gel. Their size was determined by comparison with an M13 sequence ladder.

Primer extension experiments were completed as described (Maniatis et al., 1982). The following 5′ end-labeled oligonucleotide, complementary to the $p54^{c\text{-}ets\text{-}1}$ mRNA, was used as a primer:

primer 2 : 5′ TCC TCT TCC TCC TCC TTC CTT CCT CGC TCT 3′

This primer was annealed at 60°C, and retrotranscription was completed using MMLV RTase (Gibco BRL). The elongation products were run on a 6% sequencing gel.

DNA transfection

Chicken embryonic fibroblasts (CEF) were plated at 10^6 cells per 60 mm plate the day before transfection. Transfections were achieved using DOTAP (Boehringer Mannheim). Forty-eight hours after transfection with 10 μg of DNA, the cells were washed three times and scraped into PBS. The cells were pelleted and resuspended in a lysis buffer composed of 25 mM Tris phosphate (pH 7.8), 8 mM $MgCl_2$, 1 mM DTT, Triton X-100 (1% v/v), and glycerol (15% v/v). After 3 freeze-thawing cycles, the cellular debris was pelleted by centrifugation, and the supernatants were recovered for the luciferase assay.

Luciferase activity assay

In the standard assay, 25 μl of 1 mM luciferin was added to a cell lysate containing 20 μg of proteins in lysis buffer. Ten μl of 10 mM ATP were injected, thus triggering the light output of the tube over a 30 second interval. The Bio-Orbit 1251 Luminometer background was typically in the range of 20 light units. The values obtained were normalized to the level of human growth hormone (hGH) released in the media of transfected cells resulting from cotransfection of 1.5 μg of ptkGH expression plasmid (Selden et al., 1986).

Results

Conservation of the organization of the $p54^{c\text{-}ets\text{-}1}$ promoter between humans and chicken

The avian *c-ets-1* gene gives rise to two different types of mRNAs (Duterque-Coquillaud et al., 1988; Leprince et al., 1988), as summarized in Figure 1. The $p54^{c\text{-}ets\text{-}1}$ mRNA joins the I^{54} exon to the common set of 3′ exons, namely a to F, while the $p68^{c\text{-}ets\text{-}1}$ mRNA results from the fusion of the Y (see below), α, and β exons to the a to F exons. Given that both types of mRNA display differential expression (Leprince et al., 1988; Quéva et al., 1993), we suspected that they might be subject to independent regulation, possibly via tissue-specific alternative splicing and/or differentially regulated specific promoters.

Our first goal was to compare the transcription regulation region of the $p54^{c\text{-}ets\text{-}1}$ mRNA between humans and chicken. We carried out primer extension analysis on total chicken RNAs extracted from spleen, thymus, and the RP9 B-lymphoma cell line. Primer 2 (see Materials and Methods and Fig. 3) initiates retrotranscription 121 bp 5′ of the ATG codon located in the I^{54} exon. This oligonucleotide is complementary to region +121/+150 in our sequence (Fig. 3), close to the 5′ end of the $p54^{c\text{-}ets\text{-}1}$ cDNA (Duterque-Coquillaud et al., 1988). In each tissue tested, three major bands of 150, 116, and 110 bp were detected (Fig. 2), thus localizing the major start sites 201 to 241 bp 5′ from the translation start codon found in the I^{54} exon (Fig. 3). Minor bands were also observed, indicative of additional minor cap sites.

To determine whether these 5′ ends were derived from I^{54} or from one or more upstream exon(s), we next performed RNase protection analysis. A 14 kbp genomic DNA fragment containing the I^{54} exon has been previously isolated (Gégonne et al., 1987; Duterque-Coquillaud et al., 1988). A 732 bp Hind III–Cla I fragment derived from this genomic DNA and including the 5′ half of I^{54} was transcribed into a cRNA probe suitable for hybridization with total RNAs from spleen and thymus. The sizes of the pro-

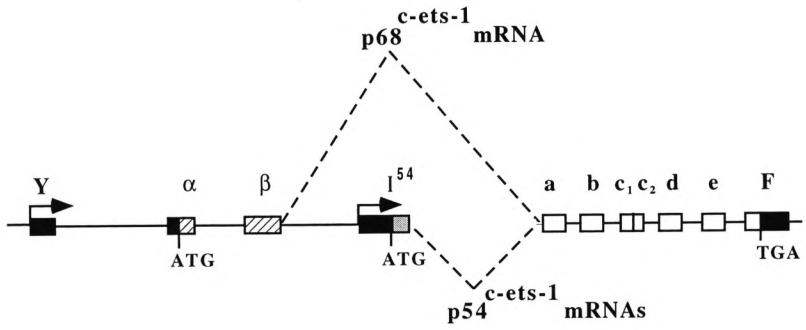

Figure 1. Genomic organization of the chicken *c-ets-1* locus. Open boxes mark the common exons, namely a to F; the alternative exons are indicated by striped (α and β) or dashed (I^{54}) boxes, and the non-coding sequences are boxed in black. The two types of mRNA, which differ in the 5′ part of the gene, are shown. The Y exon and the localization of the promoters are described in this paper.

tected bands obtained confirmed the multiplicity of the start sites, along with their positions, and identified the I^{54} exon as the first exon of the $p54^{c\text{-}ets\text{-}1}$ mRNAs (data not shown). Clustered initiation start sites have also been described in the case of the human I^{54} exon, although there are some discrepancies about their positions (Jorcyk et al., 1991; Oka et al., 1991; Majérus et al., 1992).

The promoter activity of the 732 bp Hind III–Cla I genomic sequence was assayed by transfection into CEF. For this purpose, the DNA fragment was subcloned after mutagenesis of the $p54^{c\text{-}ets\text{-}1}$ initiation codon (see Fig. 3) in both orientations into the pLUCDSS vector, in front of the coding sequence for the firefly luciferase cDNA (Table 1). LUCDSS and ptkFLUC plasmids were also transfected as controls. The RLU (relative light units) value obtained when the 732 bp Hind III–Cla I fragment is assayed in

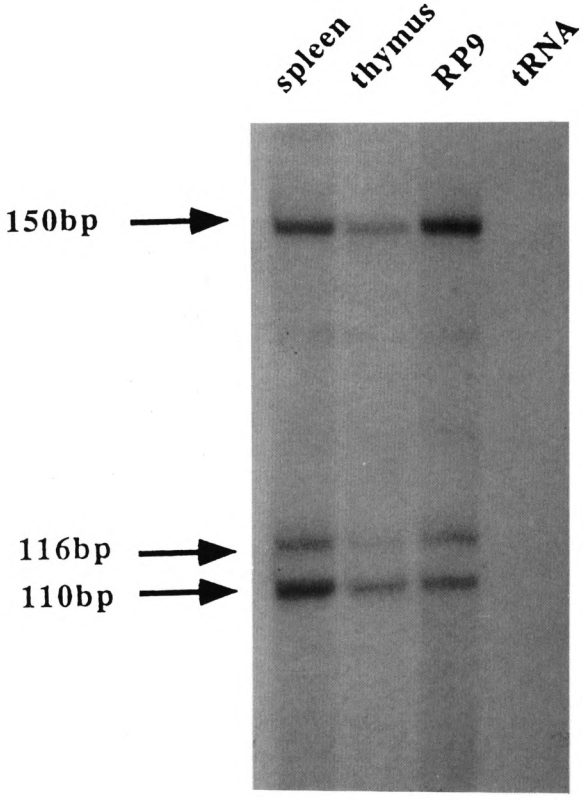

Figure 2. Mapping of the $p54^{c\text{-}ets\text{-}1}$ transcription start sites by primer extension analysis on total RNAs extracted from chicken spleen, thymus, or the RP9 B cell line. No signal is detected in the yeast tRNA control lane. The three major start sites map 241, 207, and 201 bp upstream from the I^{54} initiation codon.

Table 1. Functional activity of the two avian *c-ets-1* promoters in CEF (chicken embryo fibroblasts).

The DNA fragments using the nomenclature defined in Figures 3 and 5B were cloned in the sense orientation unless the AS letters are added. The pLUCDSS vector has no intrinsic promoter activity. The ptkFLUC vector is used as a positive control. Results are normalized for transfection efficiency and for protein amount. They represent mean values obtained from several independent experiments and are expressed as a percentage of the thymidine kinase promoter activity. Standard deviations are indicated.

Construct	*Activity in CEF*
$p54^{c\text{-}ets\text{-}1}$	
−732	258 ± 56
−732AS	6.2 ± 3
$p68^{c\text{-}ets\text{-}1}$	
−1123	60 ± 13
−533	62 ± 17
−322	83 ± 24
−90	47 ± 20
−28	6 ± 2
+12	12 ± 7
LUCDSS	10 ± 2.5
ptkFLUC	100

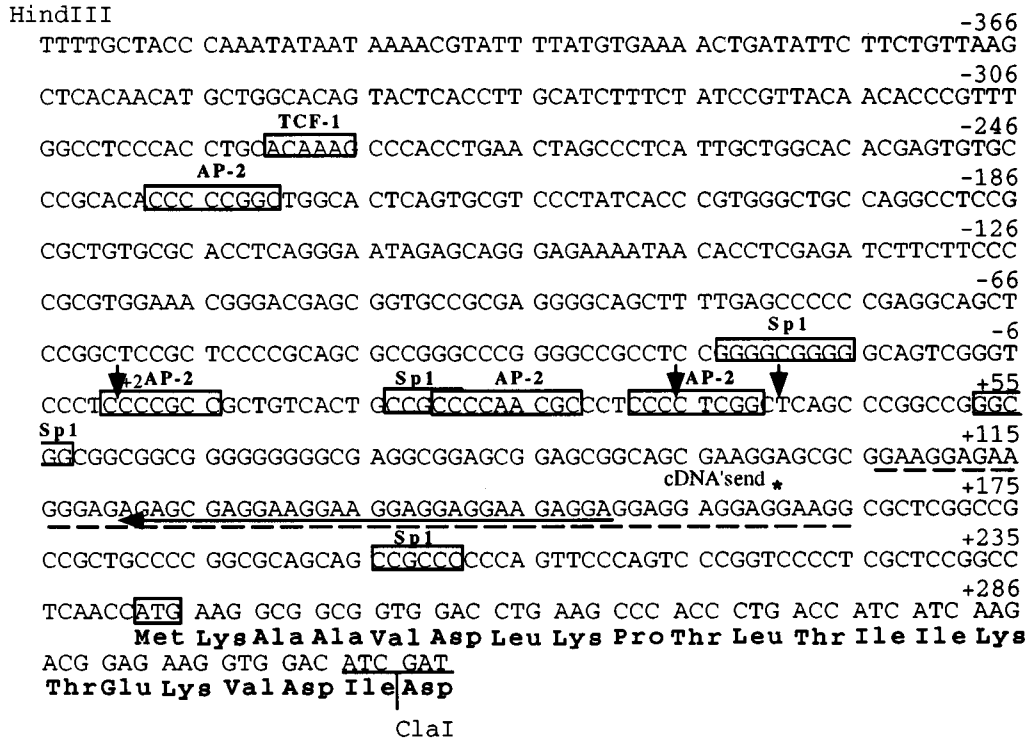

Figure 3. Nucleotide sequence of the 732 bp Hind III–Cla I DNA fragment. The mRNA major cap sites are indicated by vertical arrows, while the $p54^{c\text{-}ets\text{-}1}$ cDNA 5′ end is noted with an asterisk. The G+A stretch is underlined with dashes, and the translation initiation codon of $p54^{c\text{-}ets\text{-}1}$ is boxed. The position of primer 2 is indicated by a horizontal arrow. Putative recognition sequences for transcription factors are boxed. The bases are numbered with respect to the most upstream start site.

the sense orientation is equivalent to more than twice the thymidine kinase minimal promoter activity, thus confirming that we have indeed cloned a promoter. Neither the insert in an antisense orientation nor the LUCDSS vector displays significant promoter activity in this assay.

This 732 bp Hind III–Cla I fragment was completely sequenced (Fig. 3). Similarly to its human homologue, the avian promoter lacks a TATA box or a CAAT box; rather, it has a very high G+C base pair content (more than 65% among the 430 bp located 5′ from the ATG codon). As a consequence of the G+C abundance, four Sp1 (GGCG/TG/AG/A; Evans et al., 1988) and four AP-2 (CCCNCNG/CG/CG/C; Imagawa et al., 1987) cognate motifs are found in the vicinity of the cap sites, at positions −24, +17, +53, +196 and −238, −1, +20, +32, respectively. Interestingly, AP-2 has been shown to activate the human promoter in a transient transfection assay performed in the Daudi cell line (Oka et al., 1991). Another potential binding site seems to be found exclusively in chicken: a recognition sequence at position −291 for TCF-1, an HMG-box protein specifically expressed in T-cells (Waterman et al., 1991). Within the +106/+165 region, the sequence is composed exclusively of G+A base pairs, similarly to the region 1604/1723 bp described in the human promoter (Jorcyk et al., 1991). Purine stretches have been proposed to form triple helix DNA structures (Hanvey et al., 1988).

The $p54^{c\text{-}ets\text{-}1}$ mRNA is initiated about 10.0 kbp downstream of the $p68^{c\text{-}ets\text{-}1}$-specific α and β exons group (Fig. 1). Turning to the complete organization of the chicken *c-ets-1* locus, we could then conclude that the mRNA encoding $p68^{c\text{-}ets\text{-}1}$ initiates its synthesis from a second promoter just upstream from the already identified α and β exons, or one or more 5′ noncoding exon(s). Our study then focused on characterizing this putative second promoter region.

Mapping of initiation sites of the p68$^{c\text{-}ets\text{-}1}$ transcript and molecular cloning of the 5′-flanking region encompassing the p68$^{c\text{-}ets\text{-}1}$ first exon

Cloning of the ca. 1.5 kbp p68$^{c\text{-}ets\text{-}1}$ cDNA, isolated from a chicken spleen cell library, has been previously reported (Leprince et al., 1988). It includes 59 nucleotides upstream from the 5′ boundary of the α exon, probably corresponding to at least one unidentified noncoding exon. This putative exon was named Y. Primer extension analysis performed from an oligonucleotide overlapping this putative exon and the α exon 5′ boundary suggested that the cDNA previously isolated was full-length in 5′, since about 70 bp were extended by this method (data not shown). This result further suggested that sequences just upstream of the Y exon might include the putative p68$^{c\text{-}ets\text{-}1}$ mRNA regulatory region. By differential screening of a total chicken embryo DNA library with a 5.4 kbp Hind III genomic probe encompassing the α exon, and with an oligonucleotide colinear to the 5′ end of the cDNA, we isolated genomic DNA including the Y exon. Restriction map analysis showed its location about 10.0 kbp upstream from the α exon (Figs. 1 and 4). Its 3′ boundary was sequenced and displayed typical features of a true splice donor site that should be fused to the splice acceptor site of the α exon to yield the p68$^{c\text{-}ets\text{-}1}$ cDNA (Fig. 4).

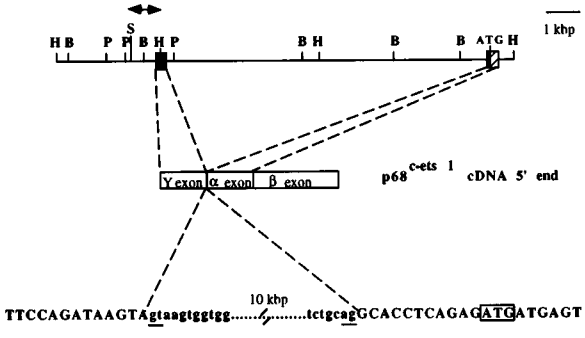

Figure 4. Genomic organization of the 5′ part of the chicken *c-ets-1* locus. Relevant restriction sites are indicated as follows: B = BamH I; H = Hind III; P = Pst I; S = Ssp I. The Y and α exons are marked by boxes; the coding part of the α exon is striped. A double-headed arrow covers the 1204 bp DNA fragment, which was cloned into the pLUCDSS vector for testing promoter activity (fragment −1123). The three most 5′ exons included in the p68$^{c\text{-}ets\text{-}1}$ cDNA are shown below. At the bottom, the sequence of the Y and α exons splice junctions is shown. The intron 5′ and 3′ boundaries are noted in lowercase letters. The conserved gt and ag nucleotides are underlined, and the initiation codon is boxed.

To confirm the transcription start sites of the p68$^{c\text{-}ets\text{-}1}$ mRNA and to ensure that the Y exon is the unique 5′ noncoding exon, we performed RNase protection analysis on total RNA extracted from chicken spleen, thymus, and embryonic dermis. A ^{32}P-labeled RNA probe complementary to the coding strand of a genomic 400 bp BamH I–Hind III DNA fragment comprising the 5′ part of the Y exon was used (see Figure 4). The result of a representative experiment is shown in Figure 5A. Protected fragments were detected only in the spleen and in embryonic dermis, where p68$^{c\text{-}ets\text{-}1}$ is expressed, but not in the thymus, where p68$^{c\text{-}ets\text{-}1}$ mRNA is not detected, consistent with tissue distribution previously described (Leprince et al., 1990; Quéva et al., 1993). We obtained multiple protected bands—one major band of 86 bp and 4 minor bands of 89, 85, 83, and 77 bp respectively—suggestive of multiple transcription start sites rather than multiple exons, given the small differences in the size of the bands observed. This result indicates that the Y exon is between 105 and 117 bp in size. It should be borne in mind that the size of cDNAs as well as that of primer extension or tailed PCR products can be determined by strong stops during the retrotranscription process. Despite extensive efforts, we could never correlate primer extension with RNase protection data, which probably signifies a premature retrotranscription arrest during the synthesis of the cDNA strand. This observation could also explain why primer extension results (data not shown) matched the size of the longest p68$^{c\text{-}ets\text{-}1}$ cDNA isolated (see above).

Given the small size of their 5′ noncoding regions, we demonstrate in addition the presence of a common unexpectedly long 3′ noncoding region, ca. 6.0 kbp, for both the p54$^{c\text{-}ets\text{-}1}$ and p68$^{c\text{-}ets\text{-}1}$ mRNAs. These long 3′ noncoding regions might allow additional regulation at the posttranscriptional level.

Functional activity of the p68$^{c\text{-}ets\text{-}1}$ mRNA promoter region in fibroblasts

In order to demonstrate that we have isolated an efficient promoter region for the transcription of p68$^{c\text{-}ets\text{-}1}$ mRNA, we inserted a 1204 bp Ssp I–Hind III restriction fragment into the

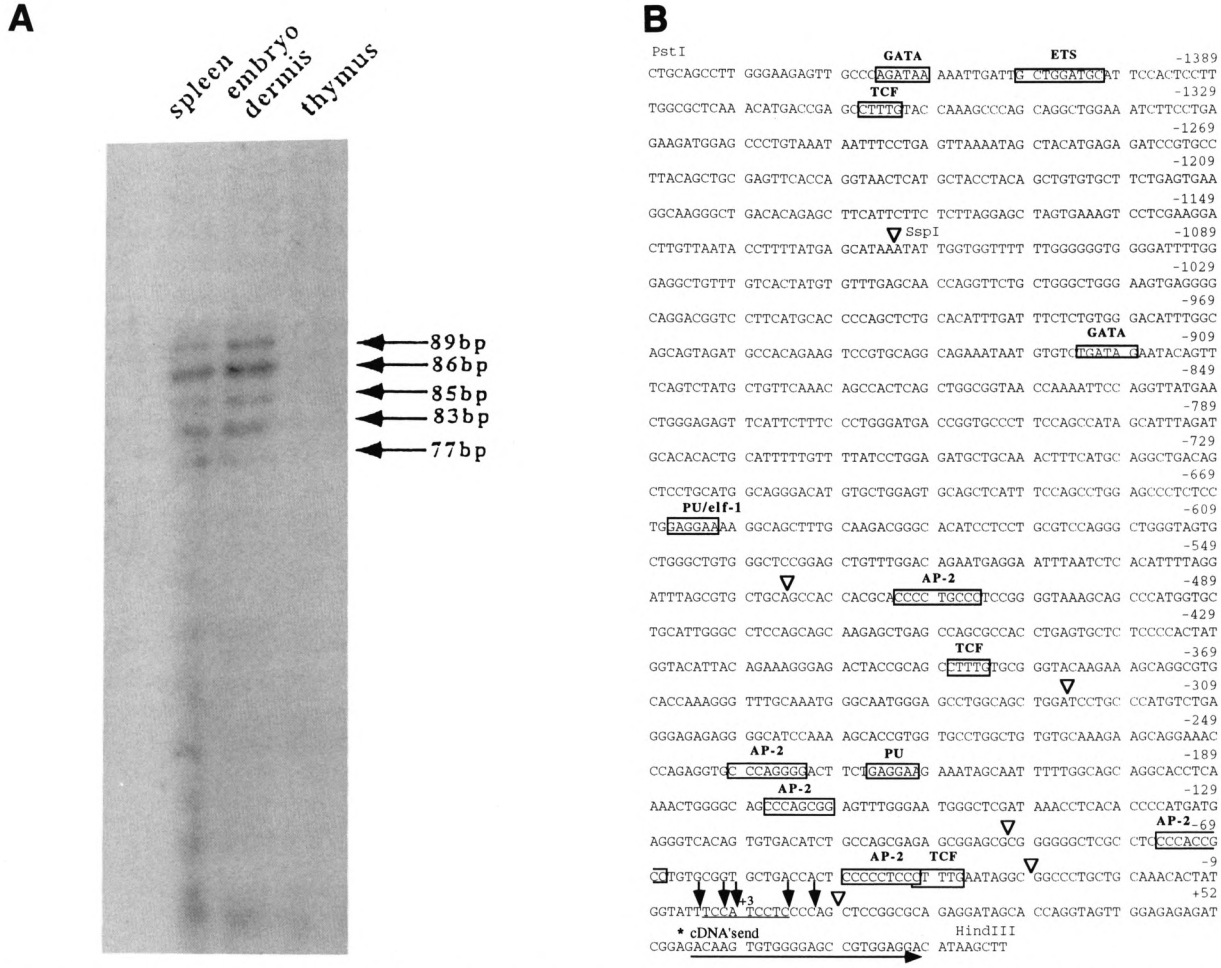

Figure 5. A. Mapping of the p68$^{c\text{-}ets\text{-}1}$ transcripts start sites by RNase protection analysis on total RNA extracted from chicken spleen, E6 embryonic dermis, or thymus. Five start sites extend the Y exon beyond the Hind III restriction site of 77 to 89 bp, as indicated by arrows. **B.** Nucleotide sequence of the 1.5 kbp chicken genomic DNA spanning the first exon and the 5′ upstream region of the p68$^{c\text{-}ets\text{-}1}$ mRNA. It is numbered with respect to the major start site. The nomenclature is the same as in Figure 3, except that the horizontal arrow marks primer 1. The Inr region is underlined. The open triangles indicate the 5′ endpoints of the deletions tested in the functional assay (see Table 1).

pLUCDSS vector. Its 3′ border is the Hind III restriction site located in the Y exon (Fig. 4), and this fragment is referred to as fragment −1123 on the basis of the position of the most upstream cap site. This reporter plasmid was transfected into CEF, parallel to the pLUCDSS vector or the ptkFLUC plasmid as controls. We also assayed the promoter activity of different deletion constructs, in order to define crucial regions for transcriptional regulation.

Mean values of several experiments are indicated in Table 1. These results show that the −1123 DNA fragment display weak (60% of the activity of the minimal tk promoter on average) but significant promoter activity, when compared to that of the pLUCDSS vector (6-fold activation). The levels of the −1123 and −533 fragments' activity are identical, while activity is slightly increased upon deletion of another 211 bp. In contrast, promoter activity of the −90 construct is slightly reduced, and the −28 deletion mutant activity decreases dramatically (10-fold inhibition compared to the full-length promoter), indicative of a positively regulated region located between positions −90 and −28. The low values obtained with the +12 construct corroborate the predicted position of the cap sites.

Nucleotidic sequence of the p68$^{c\text{-}ets\text{-}1}$ promoter

Sequencing of the 1.5 kbp genomic DNA fragment upstream from the Hind III restriction site located within the Y exon was performed. Examination of the sequence surrounding the predicted cap sites reveals no canonical TATA or CAAT boxes, as had been previously observed with the human and chicken p54$^{c\text{-}ets\text{-}1}$ promoters. However, in striking contrast to these two promoters, a high G+C content was not seen in the p68$^{c\text{-}ets\text{-}1}$ promoter (Fig. 5B). Rather, a pyrimidine-rich initiator-like sequence (Inr region), previously described as a potential recognition site for the transcription initiation complex (Smale and Baltimore, 1989), overlaps the position of the start sites. Preliminary DNase I footprinting experiments indicate a protection of this region (data not shown).

Other features of interest are three putative binding sites for ets-related factors at positions −1409, −666, and −225. These sites are centered around a GGAA/T core sequence (Woods et al., 1992). They are suggestive of autoregulation, or of cross-regulation by different members of the *ets* family. Another possible aspect of the regulation of the p68$^{c\text{-}ets\text{-}1}$ expression is suggested by the presence of several consensus binding sites for lymphotropic transcription factors. We found three putative binding sites for the lymphoid-specific TCF-1 (−1366, −397, −40), a PU box (Klemsz et al., 1990) included in a binding site for the elf-1 protein (Thompson et al., 1992; Wang et al., 1992) at position −666, and two GATA sites (Evans and Felsenfeld, 1991; Ko et al., 1991) at positions −1424 and −923. These DNA motifs might account for the repression of p68$^{c\text{-}ets\text{-}1}$ mRNA in lymphoid cells. Some binding sites for more ubiquitously expressed transcription factors were also found. Although the overall base composition of the p68$^{c\text{-}ets\text{-}1}$ mRNA promoter is not of the G+C-rich type, its sequence reveals multiple AP-2 binding sites (Imagawa et al., 1987) in positions −522, −239, −176, −75, and −48. Therefore, some of them are in the vicinity of the predicted cap sites.

The p68$^{c\text{-}ets\text{-}1}$ protein enhances its own expression

The presence of several binding sites for ets-related transcription factors prompted us to investigate the possibility of autoregulation of the p68$^{c\text{-}ets\text{-}1}$ promoter, as previously demonstrated for a number of other transcription factors. Cotransfections in CEF of the −1123 construct with a pSG5 expression vector producing the p68$^{c\text{-}ets\text{-}1}$ protein resulted in a slight but reproducible (twofold ± 0.3 SD with 1 μg of expression plasmid) and dose-dependent activation (Fig. 6). Values obtained in lanes 4 and 5 indicate a lack of effect of up to 4 μg of pSG5 p68$^{c\text{-}ets\text{-}1}$ on the pLUCDSS vector. Such twofold activation has been reported as autoregulation in the case of other promoters, among which is the human *c-ets-1* gene (Oka et al., 1991). Such weak activation may be due to the low level of *c-ets-1* proteins exogenously expressed in fibroblasts (our unpublished data). We are currently attempting to establish whether this effect is direct and to identify the DNA sequence involved.

Discussion

Previous studies have demonstrated that the expression of the chicken *c-ets-1* locus leads to the

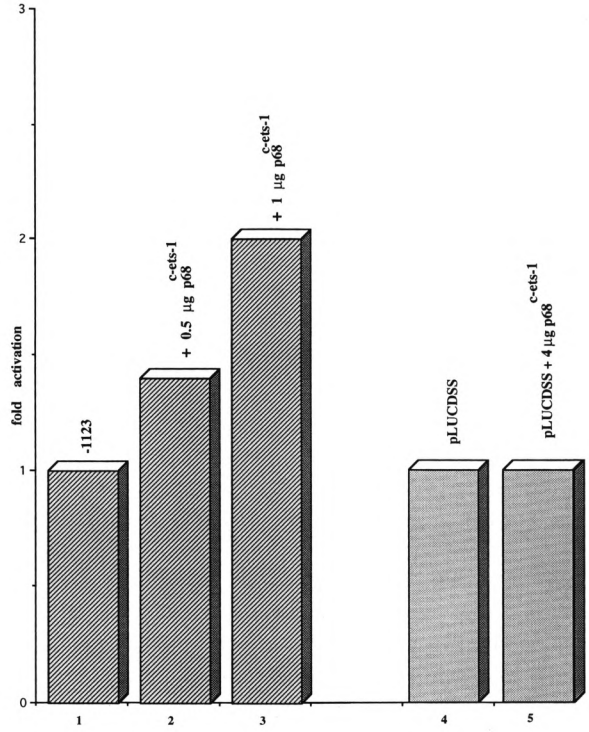

Figure 6. Effect of the p68$^{c\text{-}ets\text{-}1}$ protein on its own promoter assayed in CEF. The −1123 deletion mutant subcloned into the pLUCDSS was transfected alone (lane 1), or with 0.5 or 1 μg of pSG5p68$^{c\text{-}ets\text{-}1}$ expression vector (lanes 2 and 3, respectively). Lanes 4 and 5 show the lack of effect of the expression vector on pLUCDSS activity.

synthesis of two proteins differing in their N-termini, $p54^{c\text{-}ets\text{-}1}$ and $p68^{c\text{-}ets\text{-}1}$ (Leprince et al., 1988). Initially, their respective expression, described by immunoprecipitation and RNase protection analysis, suggested marked differences in their level of expression and in their tissue distribution, with preferential localization of $p54^{c\text{-}ets\text{-}1}$ in lymphoid cells (Ghysdael et al., 1986) and restriction of $p68^{c\text{-}ets\text{-}1}$ to blood vessels of the spleen (Leprince et al., 1990). But the highly sensitive in situ hybridization method reveals a more subtle chicken *c-ets-1* mRNA expression pattern in the embryo (Quéva et al., 1993). In summary, during embryonic life, the tissue distribution of $p54^{c\text{-}ets\text{-}1}$ and $p68^{c\text{-}ets\text{-}1}$ largely overlaps. Strikingly, however, $p68^{c\text{-}ets\text{-}1}$ is completely excluded from chick thymocytes as well as from lymphocytes, while $p54^{c\text{-}ets\text{-}1}$ is highly expressed in these cells.

To investigate whether transcriptional regulation could account for this differential expression, we were interested in defining the promoter regions involved. In this study, we show that the two major transcripts produced from the chicken *c-ets-1* locus are driven by two distinct promoter regions. The promoter regions directing the respective transcription of $p68^{c\text{-}ets\text{-}1}$ and of $54^{c\text{-}ets\text{-}1}$ mRNAs both exhibit multiple start sites, possibly due to the absence of canonical TATA and CAAT motifs. Nevertheless, they differ in several aspects, among which is the G+C-rich composition of the $p54^{c\text{-}ets\text{-}1}$ promoter, in contrast to the C+T-rich $p68^{c\text{-}ets\text{-}1}$ promoter region. In addition, the putative consensus DNA motifs for transcription factors differ somewhat. All of these divergences might account for the differential expression of both *c-ets-1* major transcripts in avian tissues.

We demonstrate that the $p54^{c\text{-}ets\text{-}1}$ mRNA 5′-flanking region displays a G+C-rich sequence comparable to the human *c-ets-1* and *c-ets-2* promoters previously reported (Mavrothalassitis et al., 1990; Jorcyk et al., 1991; Oka et al., 1991). This type of promoter was initially thought to be typical of housekeeping genes, although it now appears that G+C-rich regions also characterize numerous tissue-specific promoters, such as the hematopoietic *pim-1* gene promoter (Meeker et al., 1990). $p54^{c\text{-}ets\text{-}1}$ has been identified as a protein remarkably conserved throughout evolution (Watson et al., 1988). We show here that the conservation extends to the promoter region, which appears strikingly similar in sequence and organization from chicken to human (Jorcyk et al., 1991; Oka et al., 1991). Thus, a 360 bp region (from nucleotides −60 to the ATG codon) in the avian promoter is 65% homologous on average to a colinear region in the human promoter. As for some other proto-oncogene promoters, several of the transcription factor consensus motifs are conserved, but their position varies in the overall sequence. Nevertheless, these observations suggest similar transcriptional regulation between the two species. This regulation may also extend to autoregulation, supported by the observation that the murine $p63^{c\text{-}ets\text{-}1}$ promoter is positively regulated by the avian $p54^{c\text{-}ets\text{-}1}$ protein (Seth and Papas, 1990), which is reminiscent of autoregulation of the human promoter (Oka et al., 1991; Majérus et al., 1992).

In contrast to the well-conserved I^{54} exon, the α and β exons, specific for $p68^{c\text{-}ets\text{-}1}$, have not been identified in any mammalian species until now (Albagli et al., 1992). The $p68^{c\text{-}ets\text{-}1}$ mRNA 5′-flanking region reported here lacks characteristic structural features of usual promoters. Nevertheless, examples of transcription regulation sequences lacking a TATA box or a high G+C content have been well documented for a number of developmentally regulated genes, such as Drosophila homeotic genes (Biggin and Tjian, 1988) or lymphocyte differentiation genes (Anderson et al., 1988), whose transcripts are driven from a C+T-rich initiator sequence (the so-called Inr).

We identified several putative binding sites for known transcription factors in this promoter, the functionality of which remains speculative at this time. Because the most striking divergence in the patterns of expression of both *c-ets-1* mRNA is illustrated in lymphoid organs, experiments are now under way to test the functionality of lymphoid-specific transcription factors possibly involved in this differential regulation. Among the candidates, the T-cell factor TCF-1 (Waterman et al., 1991; van de Wetering et al., 1991; Allen et al., 1992), GATA-3 (Ko et al., 1991), elf-1 (Thompson et al., 1992), or $p54^{c\text{-}ets\text{-}1}$ proteins might be effectors that inhibit $p68^{c\text{-}ets\text{-}1}$ expression or activate $p54^{c\text{-}ets\text{-}1}$ in lymphoid cells. The important role of the $p54^{c\text{-}ets\text{-}1}$ protein in lymphoid organs is suggested by the discovery of target genes representative of the onset of lymphocyte-specific biological functions during T cell ontogeny (Prosser et al., 1992;

Ho et al., 1990; Leiden, 1992). The lymphoid-specific repression of $p68^{c\text{-}ets\text{-}1}$ appears puzzling, since $p54^{c\text{-}ets\text{-}1}$ and $p68^{c\text{-}ets\text{-}1}$ are structurally related transcription factors that bind their DNA cognate motif with analogous specificity and affinity. They differ only in their N-termini. The α and β exons are translated into a hydrophobic peptide, which possesses autonomous transactivating properties when fused to a heterologous DNA-binding domain, while the amino acids encoded by I^{54} are mostly hydrophilic and are devoid of such properties (Schneikert et al., 1992). Activation domains interact with the general transcription apparatus either directly or through bridging co-activators. The differences between the two *c-ets-1* proteins may reside in the mechanism by which they interact with other transcriptional activators or co-activators, and thus alter the spectrum of regulated genes. Thus, one might speculate that the strict inhibition of $p68^{c\text{-}ets\text{-}1}$ expression in lymphoid cells avoids competition for common target genes that both proteins would regulate differently.

The $p68^{c\text{-}ets\text{-}1}$ mRNA is initiated from a weak promoter. Other genes have been described that show reduced promoter activities highly strengthened by a distant control element. In fact, we did not expect either a potent promoter or enhancers, since previous experiments aimed at measuring the $p68^{c\text{-}ets\text{-}1}$ mRNA level in tissues always indicated very low levels of expression. In this respect, this promoter is akin to the interleukin-1 receptor gene (Ye et al., 1993). Given the restricted pattern of expression of $p68^{c\text{-}ets\text{-}1}$, we could not test its mRNA promoter in a cell type in which the endogenous protein is expressed. We assayed CEF as an easily transfectable and commonly used experimental model, although $p68^{c\text{-}ets\text{-}1}$ is not expressed in fibroblasts (our unpublished observations). Promoter activities of other tissue-specific genes, such as GATA-1 (Hannon et al., 1991) or *c-rel* (Hanninck et al., 1990), have been already observed in CEF, although no explanation for these results is available at the moment. In fact, when we assayed a DNA fragment extending 3.0 kbp upstream of the Y exon, the promoter activity was dramatically reduced (data not shown), possibly reflecting the lack of function of the endogenous promoter in these cells.

The various binding sites for ets-related transcription factors are possibly involved in the mediation of the slight positive autoregulation of the $p68^{c\text{-}ets\text{-}1}$ promoter we report. Experiments are now in progress to find a cell in which this autoregulation is enhanced, in case an accessory protein, absent in the CEF, helps $p68^{c\text{-}ets\text{-}1}$ to autoregulate its expression positively. Because in situ hybridization experiments show localized and high-level expression of $p68^{c\text{-}ets\text{-}1}$ in a small subset of cells in the embryonic dermis (Quéva et al., 1993), we are currently trying to derive an in vitro system from this tissue. This model should allow us to address several interesting questions, among which is the status of both promoters in the few cells that synthetize both messengers. Dermis cell culture should also provide a system that would enhance the $p68^{c\text{-}ets\text{-}1}$ promoter activity, thus showing the regions crucial for transcriptional regulation.

Ackowledgments

We would like to thank members of the GEPHO for their constant interest and support of this work, and especially S. Ansieau, J. Coll, C. Dozier, and M. Duterque-Coquillaud. We are also grateful to R. Lafyatis, V. Fafeur, V. Laudet, and C. Quéva for numerous helpful discussions and critical reading of the manuscript. C. Lagrou is warmly thanked for his efficient help in cell cultures.

The financial support of this work was provided by the Centre National de la Recherche Scientifique, by the Association pour la Recherche contre le Cancer, and by the Institut Pasteur de Lille.

The costs of publishing this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC Section 1734 solely to indicate this fact.

References

O. Albagli, A. Flourens, P. Crépieux, A. Bègue, D. Stéhelin, and D. Leprince (1992), Oncogene 7, 1435–1439.

J. M. Allen, K. A. Forbush, and R. M. Perlmutter (1992), Mol Cell Biol 12, 2758–2768.

S. J. Anderson, H. S. Chou, and D. Y. Loh (1988), Proc Natl Acad Sci USA 85, 3551–3554.

M. D. Biggin and R. Tjian (1988), Cell 53, 699–711.

S. Dalton and R. Treisman (1992), Cell 68, 597–612.

J. R. DeWet, K. V. Wood, M. Deluca, D. R. Helinski, and S. Subramant (1987), Mol Cell Biol 7, 725–737.

M. Duterque-Coquillaud, D. Leprince, A. Flourens, C. Henry, J. Ghysdael, B. Debuire, and D. Stéhelin (1988), Oncogene Res 2, 335–344.

T. Evans, T. DeChiara, and A. Efstratiadis (1988), J Mol Biol 199, 61–81.
T. Evans and G. Felsenfeld (1991), Mol Cell Biol 11, 843–853.
A. Gégonne, D. Leprince, M. Duterque-Coquillaud, B. Vandenbunder, A. Flourens, J. Ghysdael, B. Debuire, and D. Stéhelin (1987), Mol Cell Biol 7, 806–812.
J. Ghysdael, A. Gégonne, P. Pognonec, D. Dernis, D. Leprince, and D. Stéhelin (1986), Proc Natl Acad Sci USA 83, 1714–1718.
F. Gouilleux, B. Sola, B. Couette, and H. Richard-Foy (1991), Nucleic Acids Res 19, 1563–1569.
C. V. Gunther, J. A. Nye, R. S. Bryner, and B. J. Graves (1990), Genes Dev 4, 667–679.
M. Hanninck and H. Temin (1990), Oncogene 5, 1843–1850.
R. Hannon, T. Evans, G. Felsenfeld, and R. Gould (1991), Proc Natl Acad Sci USA 88, 3004–3008.
J. C. Hanvey, M. Shimizu, and R. D. Wells (1988), Proc Natl Acad Sci USA 85, 6292–6296.
R. A. Hipskind, V. N. Rao, C. G. F. Mueller, E. S. P. Reddy, and A. Nordheim (1991), Nature 354, 531–534.
I. C. Ho, N. K. Bhat, L. R. Gottschall, T. Lindsten, C. B. Thompson, T. S. Papas, and J. M. Leiden (1990), Science 250, 814–818.
M. Imagawa, R. Chiu, and M. Karin (1987), Cell 51, 251–260.
C. L. Jorcyk, D. K. Watson, G. J. Mavrothalassitis, and T. S. Papas (1991), Oncogene 6, 523–532.
F. D. Karim, L. D. Urness, C. S. Thummel, M. J. Klemsz, S. R. McKircher, A. Celada, C. Van Beveren, R. A. Maki, C. V. Gunther, J. A. Nye, and B. J. Graves (1990), Genes Dev 4, 1451–1453.
M. J. Klemsz, S. R. McKercher, A. Celada, C. Van Beveren, and R. Maki (1990), Cell 61, 113–124.
L. J. Ko, M. Yamamoto, M. W. Leonard, K. M. George, P. Ting, and J. D. Engel (1991), Mol Cell Biol 11, 2778–2784.
V. Laudet, C. Niel, M. Duterque-Coquillaud, D. Leprince, and D. Stéhelin (1992), Biochem Biophys Res Commun 190, 8–14.
J. A. Lautenberger, L. A. Burdett, M. A. Gunnel, S. Qi, D. K. Watson, S. J. O'Brien, and T. S. Papas (1992), Oncogene 7, 1713–1719.
J. M. Leiden (1992), Immunol Today 13, 22–30.
D. Leprince, A. Gégonne, J. Coll, C. de Taisne, A. Schneeberger, C. Lagrou, and D. Stéhelin (1983), Nature 306, 395–397.
D. Leprince, M. Duterque-Coquillaud, R. P. Li, C. Henry, A. Flourens, B. Debuire, and D. Stéhelin (1988), J Virol 62, 3233–3241.
D. Leprince, J. C. Gesquière, and D. Stéhelin (1990), Oncogene Res 5, 255–265.
K. MacLeod, D. Leprince, and D. Stéhelin (1992), Trends Biochem Sci 17, 251–256.
M. A. Majérus, F. Bibollet-Ruche, J. B. Telliez, B. Wasylyk, and B. Bailleul (1992), Nucleic Acids Res 20, 2699–2703.
T. Maniatis, E. F. Fritsch, and J. Sambrook (1982), Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
G. J. Mavrothalassitis, D. K. Watson, and T. S. Papas (1990), Proc Natl Acad Sci USA 87, 1047–1051.
T. C. Meeker, J. Loeb, M. Ayres, and W. Sellers (1990), Mol Cell Biol 10, 1680–1688.
M. Nunn, P. H. Seeburg, C. Moscovici, and P. H. Duesberg (1983), Nature 306, 391–395.
T. Oka, A. Rairkar, and J. H. Chen (1991), Oncogene 6, 2077–2083.
H. M. Prosser, D. Wotton, A. Gégonne, J. Ghysdael, S. Wang, N. A. Speck, and M. J. Owen (1992), Proc Natl Acad Sci USA 89, 9934–9938.
C. Quéva, D. Leprince, D. Stéhelin, and B. Vandenbunder (1993), Oncogene, in press.
V. N. Rao, K. Huebner, M. Isobe, A. Ar-Rushdi, C. M. Croce, and E. S. P. Reddy (1989), Science 244, 66–70.
F. S. Sanger, S. Nicklen, and A. R. Coulson (1977), Proc Natl Acad Sci USA 74, 5463–5467.
J. R. Sayers, W. Schmidt, and F. Eckstein (1988), Nucleic Acids Res 16, 791–802.
J. Schneikert, Y. Lutz, and B. Wasylyk (1992), Oncogene 7, 249–256.
R. F. Selden, K. B. Howie, M. E. Rowe, H. M. Goodman, and D. D. More (1986), Mol Cell Biol 6, 3173–3179.
A. Seth and T. S. Papas (1990), Oncogene 5, 1761–1767.
S. T. Smale and D. Baltimore (1989), Cell 57, 103–113.
C. B. Thompson, C. Wang, I. Ho, P. R. Bohjanen, B. Petryniak, C. H. June, S. Miesfeldt, L. Zhang, G. J. Nabel, B. Karpinski, and J. M. Leiden (1992), Mol Cell Biol 12, 1043–1053.
M. van de Wetering, M. Oosterwegel, D. Dooijes, and H. Clevers (1991), EMBO J 10, 123–132.
J. V. Virbasius and R. C. Scarpulla (1991), Mol Cell Biol 11, 5631–5638.
C. Y. Wang, B. Petryniak, I. C. Ho, C. B. Thompson, and J. M. Leiden (1992), J Exp Med 175, 1391–1399.
B. Wasylyk, C. Wasylyk, P. Flores, A. Bègue, D. Leprince, and D. Stéhelin (1990), Nature 346, 191–193.
C. Wasylyk, A. Gutman, R. Nicholson, and B. Wasylyk (1991), EMBO J, 5, 1127–1134.
M. L. Waterman, W. H. Fisher, and K. A. Jones (1991), Genes Dev 5, 656–669.
D. K. Watson, M. J. McWilliams, P. Lapis, J. A. Lautenberger, C. W. Schweinfest, and T. S. Papas (1988), Proc Natl Acad Sci USA 85, 7862–7866.
D. B. Woods, J. Ghysdael, and M. Owen (1992), Nucleic Acids Res 20, 699–704.
K. Ye, C. A. Dinarello, and B. D. Clark (1993), Proc Natl Acad Sci USA 90, 2295–2299.