

# Translation start site multiplicity of the CCAAT/enhancer binding protein $\alpha$ mRNA is dictated by a small 5' open reading frame

Cor F.Calkhoven, Peter R.J.Bouwman, Lenie Snippe and Geert AB\*

Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received October 18, 1994; Revised and Accepted November 14, 1994

EMBL accession no. X66844

## ABSTRACT

**The CCAAT/enhancer binding proteins (C/EBP)  $\alpha$  and  $\beta$  of the bZIP family of transcription factors each occur as multiple forms due to translation initiation at different in-frame AUG codons from the same messenger RNA. The C/EBP $\alpha$  mRNAs of chicken, rat and *Xenopus* all contain a small 5' open reading frame (5'ORF) whose size (18 nucleotides) and distance (seven nucleotides) to the C/EBP $\alpha$  cistron has been conserved in vertebrate evolution. The present studies shows that the small 5'ORF is crucial to the leaky scanning mechanism of ribosomes causing a fraction of them to ignore the first C/EBP $\alpha$  AUG codon and to start at internal AUGs. Our data challenge the view that translational start site multiplicity is mainly governed by the sequence context of the potential initiation codons. Western analysis showed that the two major chicken C/EBP $\alpha$  translation products, the full-length cC/EBP $\alpha$ -42 which acts a *trans*-activator in liver and the N-terminally truncated cC/EBP $\alpha$ -29 which lacks transcription activation potential, occur in a fixed ratio which is similar in different expressing tissues, like liver, lung and small intestine. The presence of a similar, thusfar unnoticed, small ORF 5' to the major initiation codon of C/EBP $\beta$  mRNA suggests that start site multiplicity from this mRNA may be governed by the same mechanism.**

## INTRODUCTION

CCAAT/enhancer binding proteins (C/EBPs) belong to the bZIP class of transcription factors (1). The positively charged DNA binding domain and the leucine zipper dimerisation domain are located in the C-terminal section of the C/EBP molecule (2). Several C/EBPs, each encoded by a different member of the family of C/EBP genes, have been identified (3–17). Of these, the liver-enriched C/EBP $\alpha$  and C/EBP $\beta$  have been most extensively studied (3,11,18,19). They are involved in the transcription of several liver-specific genes, such as the serum albumin gene (20).

C/EBP $\alpha$  and C/EBP $\beta$  share the common feature that, besides the full-length product, N-terminally truncated polypeptides are translated from the same mRNA by use of internal, in-frame AUG codons (9,21). For rat C/EBP $\alpha$ , it has been shown that the smaller translation product lacking the N-terminal 117 amino acids is devoid of its transcription activation potential in liver (21). Studies with the mouse adipocyte cell line 3T3-L1 have shown that the full-length product inhibits cell proliferation, whereas the smaller translation product is not antimitotic (22). The C/EBP $\beta$  gene also issues two proteins, the liver-enriched transcriptional activator (LAP) and inhibitory (LIP) proteins. It is believed that the shorter products have a physiological function as antagonists of their *trans*-activating counterparts. The formation of multiple polypeptides from the rat C/EBP messenger RNAs is best explained by a ribosome scanning mechanism in which a fraction of ribosomes ignores the first AUG codon and start at internal AUG codons (9,21).

We have cloned a member of the C/EBP gene family of chicken (12) to study its role in the liver-specific expression of the estrogen-regulated apoVLDL II gene (23). The encoded protein, called cC/EBP, most strongly resembles C/EBP $\alpha$  of the rodent C/EBP family. The chicken C/EBP $\beta$  gene homologue has been cloned by Katz *et al.* (17) and encodes the myeloid-specific transcription factor, NF-M.

Alignment of the rat and chicken C/EBP $\alpha$  sequences has delineated three conserved regions in the N-terminal moiety of the polypeptide chain in addition to the C-terminal bZIP domain (12). Comparison of the chicken and rat C/EBP $\alpha$  shows that the methionines which are used for the internal starts are located at analogous positions within the variable sequences linking the conserved regions I and II, and between region III and the bZIP domain. Interestingly, the leader sequences of rat and chicken C/EBP $\alpha$  mRNA were found to contain a small open reading frame at seven nucleotides from the C/EBP coding region (12). The most important finding of the present studies is that this small 5'ORF, which appears to be conserved in vertebrate evolution, is the primary mRNA feature responsible for C/EBP $\alpha$  translation start site multiplicity. Moreover, sequence comparison of C/EBP $\beta$  mRNAs has revealed a, thusfar unnoticed, small ORF in front

\*To whom correspondence should be addressed

of the major C/EBP $\beta$  translation initiation codon. This feature, together with the similar distribution and sequence context of internal start sites, suggests that C/EBP $\alpha$  and  $\beta$  may follow the same strategy for generating multiple translation products from the same mRNA.

## MATERIALS AND METHODS

### Construction of cC/EBP $\alpha$ expression plasmids

A 1.38-kb *Ava*II/*Hind*II restriction fragment, containing the chicken C/EBP $\alpha$  coding sequence (EMBL accession no. X66844), 225-bp of 5' flanking and 183-bp of 3' flanking sequences, was blunted and cloned in the *Sma*I site of pGEM7Zf(+) (Promega). For transfection experiments, the *Eco*RI/*Bam*HI fragment from cC/EBP-pGEM7Zf(+), was inserted into pGS5 (24) yielding cC/EBPwt-pSG5. Several cC/EBP $\alpha$  mutants were generated by site-directed mutagenesis of the cC/EBPwt.pSG5 plasmid following the method of Kunkel *et al.* (25); Ausubel *et al.* (26). Because of the high GC-content of the cC/EBP sequence, annealing of the mutagenic oligonucleotide (50 ng) to the uracil-containing DNA template (1  $\mu$ g) was by a temperature shift from 90 to 45°C, immediately followed by primer extension in the presence of 5  $\mu$ g T4 gene 34 protein (Boehringer). Oligonucleotide primers used for generating the various mutants were:

For *cC/EBP:mut5'ORF*, oligo 5'-CCGGAGCCCTTCAGGATC CCGGCAGGCTG-3' (mismatches underlined) to mutate the ATG of the small 5'ORF and to introduce a diagnostic *Bam*HI site; for *cC/EBP:KozakATG1*, oligo 5'-GCAGGCTGTAGG-ACCACCATGGAGCAAGCC-3', to create diagnostic *Ava*II and *Nco*I sites; for *cC/EBP:mutATG3*, oligo 5'-TTCCACGGGATC CACGGGGCC-3' to mutate Met3 (codon 102) into Ile and to introduce a diagnostic *Bam*HI site; for *cC/EBP: $\Delta$ ldr*, oligo 5'-AGGCTGTAGGATCCCCATGGAGCAAGC-3' to create a *Bam*HI just upstream of the first ATG which was used for the excision and cloning of the downstream sequences; for *cC/EBP: $\Delta$ dom1*, oligo 5'-GGGGATTCGAATTCCACGGC-ATG-3' to introduce an *Eco*RI site just upstream of the third AUG codon (codon 102) which allowed the deletion of upstream sequences. *cC/EBP:bZIP* was constructed by cloning the *Sma*I/*Bam*HI fragment containing the sequences downstream of codon 220.

For the transfection of LMH cells, the reporter plasmid (D)9-tkCAT containing nine copies of the mouse albumin C/EBP binding site 'D' in front of TK minimal promoter and CAT gene was used (5,27). All constructs were checked by sequence or restriction analysis.

### Transfection

COS-1 cells were transfected with 10  $\mu$ g of pSG5-based expression plasmid as described by Ausubel *et al.*, 1994, and cultured in DMEM/F12/5% fetal calf serum (Gibco). LMH cells (28) were maintained in Waymouth's MB 752/1 medium supplemented with 10% fetal calf serum, 50  $\mu$ g/ml streptomycin and 50 U/ml penicillin (Gibco). The cells were cultured in a 10% CO<sub>2</sub>/air mixture at 37°C. DNA transfections were carried out with the calcium phosphate method (29). The day before transfection, the cells were plated at 40% confluence on 6 cm  $\phi$  dishes. The DNA precipitates contained 2  $\mu$ g (D)<sub>9</sub>-tkCAT reporter plasmid (5) and 2.5 or 5.0  $\mu$ g pSG5-based expression

vector supplemented with pGEM7Zf(+) (Promega) as carrier to total concentration of 20  $\mu$ g DNA/500  $\mu$ l. After 4 h incubation, the cells were shocked for 1 min with 15% glycerol, washed and maintained in serum-free medium. Transfected cells were harvested 40–44 h after transfection. CAT expression was by measured by non-TLC CAT assay (30) using equal amounts of cell protein per assay. To account for transfection efficiency, all transfections were in duplicate and performed at least three times.

### Nuclear proteins

Nuclei were isolated from liver, lung and small intestine as described by Lichtsteiner *et al.* (24) and from cell cultures as described by Hoppe-Seyler *et al.* (42). Liver nuclear ammonium sulfate extracts were prepared as described earlier (23).

Total nuclear lysates from liver were prepared using a protocol kindly provided by Lavery and Schibler [NUN extracts (32)]. All solutions contained several proteinase inhibitors: 125  $\mu$ M PMSF, 0.7  $\mu$ g/ml leupeptin, 0.7  $\mu$ g/ml pepstatin A, 1% Trasylol.

### cC/EBP $\alpha$ antiserum

Antiserum was generated by immunization of rabbits with recombinant full-length cC/EBP $\alpha$ , purified over SDS/PAA gel (33). For overexpression of the recombinant cC/EBP $\alpha$ , the *Bam*HI fragment of the cC/EBP: $\Delta$ Ldr mutant was cloned in the prokaryotic expression vector pET3c (44) and introduced in the *Escherichia coli* strain BL21DE3.pLysE. Cultures were grown at 28°C to OD<sub>660</sub>=0.6 and induced with 1 mM IPTG. After 4 h at 28°C, cells were harvested, washed and resuspended in 1/100 volume solution of 20% glycerol, 100 mM KCl, 12.5 mM MgCl<sub>2</sub> and 1 mM DTT in 25 mM HEPES (pH 7.8). Total protein extracts were made by sonification, and removal of precipitated material by 30 min centrifugation in a micro-centrifuge.

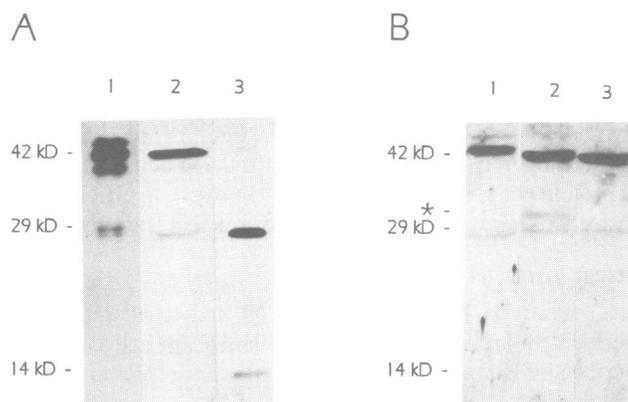
### Western analysis

Nuclear proteins and COS-1 cells were lysed in boiling, 2 $\times$ SDS-sample buffer. The proteins were separated on a 12.5% SDS/PAA gel before electroblotting on PVDF (Immobilon-P, Millipore) or nitrocellulose (Hybond-ECL, Amersham) membrane. After blocking for 1 h in TBST [50 mM Tris-HCl (pH 7.6), 150 mM NaCl, 0.02% Tween-20] supplemented with 5% dry milk, the membranes were incubated with primary antiserum for 2 h at room temperature (dilution 1/1500 in TBST-5% milk). Subsequently, the membranes were washed three times with TBST-5% milk and incubated for 1 h with HRP-anti rabbit antibody dilution of 1/5000 or 1/50 000 in TBST-5% milk. Membranes were washed three times with TBST-5% milk and luminicent-detection took place as described by the manufacturer (Amersham ECL system).

## RESULTS

### Tissue distribution of cC/EBP $\alpha$

Western analysis of chicken liver nuclear extract using an antiserum raised against recombinant chicken C/EBP $\alpha$  revealed several bands (Fig. 1A, lane 2). As will be proven by transient expression experiments of several cC/EBP $\alpha$  mutants (see below), these bands represent the full-length cC/EBP $\alpha$  polypeptide (42 kD) and truncated polypeptides formed by translational starts at internal, in-frame AUG codons from the same messenger RNA (40,29 and 14 kD). Earlier, Ossipow *et al.* (21) and Lin *et al.*



**Figure 1.** Presence of chicken C/EBP $\alpha$  isoforms in different cell type. (A) Total cell extract of COS-1 cells transfected with chicken C/EBPwt-pSG5 expression plasmid (lane 1), total liver nuclear lysates (NUN) (50  $\mu$ g protein) (lane 2) and liver nuclear ammonium sulfate extracts (50  $\mu$ g protein) (lane 3) were analyzed by Western blotting with anti-C/EBP antiserum raised against recombinant chicken C/EBP $\alpha$ . (B) total nuclear lysates (NUN) from chicken liver (lane 1), lung (lane 2) and small intestine (lane 3) were analyzed similarly. The origin of the additional band in lung (marked \*) is unknown.

(22) showed that rat C/EBP $\alpha$  mRNA is translated into isoforms of 42, 30 and 20 kD. Judged from Northern analysis, the chicken C/EBP $\alpha$  gene is not only transcribed in liver, but also in some other tissues, most abundantly in lung and small intestine (data not shown). When nuclear lysates from these three tissues were compared by Western blotting, we found similar ratios of the 42-kD and 29-kD cC/EBP $\alpha$  isoforms (Fig. 1B). So, the ratio in which the isoforms are present appears not to be subject to tissue-specific regulation. During the course of these studies, we noticed that the 42-kD cC/EBP $\alpha$  is partially lost in the ammonium sulfate nuclear extraction procedure (23,35,36). In the obtained extracts, the 29-kD protein is the most abundant isoform (Fig. 1A, lane 3). To avoid the fractionation of cC/EBP $\alpha$  isoforms, we changed to use complete nuclear lysates prepared following a recently described protocol by Lavery and Schibler (32). From these experiments we conclude that of the cC/EBP $\alpha$  proteins, the 42-kD full-length peptide is the prevalent form in nuclei of liver, lung and small intestine and that the 29-kD peptide is present in considerable lower amounts. The 14-kD product was detectable after prolonged exposure of immunoblots only (Fig. 1A, lane 3).

#### Identification of cC/EBP $\alpha$ translation products

Comparative sequence analysis showed that chicken and rat C/EBP $\alpha$  have highly similar primary structures (12). Four conserved regions, I, II, III and bZIP, with over 70% positional identity of amino acids can be distinguished (Fig. 2A). The variable regions, preceding and linking the conserved domains, harbour methionines that are possible internal start sites because their cognate AUG codons are in a favourable context (37). Specifically, in all sequences either the purine at position -3 or the guanine at position +4 both characteristic for the optimal Kozak sequence (A/GCCAUGG) is present (Table 1).

To identify the various cC/EBP $\alpha$ -immunoreactive bands observed in chicken liver nuclear extract, we examined the

expression of wild type and modified cC/EBP $\alpha$  genes (Fig. 3B) by transfecting them into COS-1 cells using the SV40-early-promoter-containing expression plasmid pSG5 as vector. After 48 h in culture, cells were harvested and total cell protein was analyzed for cC/EBP $\alpha$  polypeptides by Western blotting (Fig. 3B). COS-1 cells transfected with the wild type construct, containing the entire cC/EBP $\alpha$  coding sequence including the 5'-untranslated leader, yielded a pattern of bands resembling that of liver nuclei (Fig. 1A, lanes 1 and 2). A cluster of three bands migrates around the expected position of the full-length cC/EBP $\alpha$  polypeptide. Of these bands, the (major) 42-kD band and (minor) 40-kD bands most likely correspond to translation products initiated at the first two AUG codons of the cC/EBP $\alpha$  coding sequence, codon position 1 and 15, respectively. This can be concluded from the transfection with a construct from which the leader sequence had been deleted (cC/EBP: $\Delta$ ldr). Deletion of the leader did not affect the 42-kD and 40-kD bands, but caused disappearance of the higher molecular weight band suggesting that the latter band (and possibly a still larger band as well) originates from a non-AUG start site within the leader (Fig. 2B; Fig. 3, cC/EBPwt and  $\Delta$ ldr) (see Discussion section).

Another notable effect of deleting the leader sequence is the disappearance of the 29-kD band. This indicates that the 29-kD protein does not result from proteolysis, but rather from leaky scanning of ribosomes across the first two AUGs, confirming a similar conclusion by Ossipow *et al.* (21) and Lin *et al.*, (22) for rodent C/EBP $\alpha$ . The 29-kD band most likely represents internal starts from the third in-frame AUG codon (codon 102), because all the immunoreactive material is shifted to this particular position when the N-terminal conserved sequence I (cC/EBP: $\Delta$ domI) is deleted (Fig. 3). Moreover, mutating the third Met codon in the wild type context (cC/EBP:mutATG3) apparently causes a shift of initiation from this particular codon to the next in-frame AUG codon 87 nt downstream. Finally, the 14-kD band is probably initiated from the ATG codon at position 231 because it corresponds to the main band generated from a construct (cC/EBP:bZIP) which has retained this particular codon as the first Met codon (Fig. 3).

#### Structural features of cC/EBP $\alpha$ mRNA causing leaky scanning

Comparison of the chicken, rat and *Xenopus* C/EBP $\alpha$  mRNA leader sequences (Fig. 2B) (12,16) revealed that all three species harbour a small 5'ORF which has been conserved with respect to length (six codons), distance to the main C/EBP $\alpha$  coding region (seven nucleotides) and strength of the AUG-context sequence (Table 1). Prompted by the studies of Kozak (38) on the effects of a small 5'ORF in a bi-cistronic preinsulin gene construct on translation reinitiation, we investigated the possible role of the C/EBP $\alpha$  5'ORF in the translational start site multiplicity of C/EBP $\alpha$  mRNA. Besides the small 5'ORF, we examined the sequence around the first cC/EBP $\alpha$  AUG as a possible feature influencing the phenomenon. To distinguish between the importance of both features, two additional expression constructs were transfected into COS-1 cells, one (cC/EBP:mut5'ORF) in which the AUG codon of the 5'ORF was mutated, eliminating the small 5'ORF, and another (cC/EBP:KozakATG1) in which the sequence around the first ATG was replaced by the optimal Kozak sequence CCACCAATGG. Judged from the intensity of the 29-kD band compared to that in the cC/EBP $\alpha$  wild-type transformant, leaky scanning of ribosomes across the first two



**Table 1.** Sequence context of potential translation initiation sites in C/EBP $\alpha$  and - $\beta$  mRNAs

Gene	5' ORF start	Major start	Internal start	Ref.
chicken C/EBP $\alpha$	<u>A</u> CCAUGC	uuCAUG <u>G</u>	<u>G</u> gCAUGc	12
rat C/EBP $\alpha$	<u>G</u> CCAUGC	cCCAUG <u>G</u>	<u>G</u> tCAUGu	3
<i>Xenopus</i> C/EBP $\alpha$	<u>A</u> CCAUGC	tCCAUG <u>G</u>	tgtAUG <u>G</u>	16
chicken C/EBP $\beta$	<u>A</u> gCAUGc	uCCAUG <u>G</u>	<u>G</u> gCAUGu	17
rat C/EBP $\beta$	<u>A</u> gCAUGc	cCCAUG <u>G</u>	<u>G</u> CCAUG <u>G</u>	19
human C/EBP $\beta$	<u>A</u> gCAUGt	tCCAUG <u>G</u>	<u>G</u> gCAUG <u>G</u>	7

Genes are indicated following the C/EBP  $\alpha$  and  $\beta$  classification. Chicken C/EBP $\beta$  is also known as NF-M (17) and human C/EBP $\beta$  and NF-IL6 (7). The critical purine at position -3 and the guanine at position +4 are underlined. Mismatches with the optimal Kozak sequence A/GCCAUGG are in lower case.

that of a control transfection with the empty pSG5-vector (Fig. 4A). This is in agreement with the findings by Ossipow *et al.* (21) obtained with the full-length 42-kD and truncated 30-kD rat C/EBP $\alpha$  products. Since the truncated proteins have retained the specific DNA binding activity (data not shown), the results imply that the N-terminal sequence 1–102, including the conserved sequence region I, harbours the *trans*-activation function. Region I coincides with the potent N-terminal *trans*-activating sequence defined in rat C/EBP $\alpha$  (39).

The different *trans*-activating potentials of the full-length and N-terminally truncated products allowed us to use the CAT activity as a parameter for the ratio of both products expressed from the various constructs. Figure 4B shows that the wild type cC/EBP $\alpha$  and the cC/EBP:mutATG3 construct gives only modest stimulation of reporter gene expression, 1.6- to 3-fold induction compared to the empty pSG5-vector. The cC/EBP:KozakATG1 construct stimulates transcription 4- to 6.6-fold and the cC/EBP:mut5'ORF 7.6-fold. The magnitudes of these stimulations are in line with the amount of 29-kD product compared to the relatively constant amount of full-length cC/EBP $\alpha$ , observed in the Western analysis of transfected COS-1 cells (Fig. 3A). Given the strong inhibition, it appears that the 29-kD product can act as a powerful antagonist of 42-kD cC/EBP $\alpha$  dependent transcription activation.

## DISCUSSION

Our results with chicken C/EBP $\alpha$  confirm the findings with rodent C/EBP $\alpha$  by Ossipow *et al.* (21) and Lin *et al.* (22) showing that these mRNAs are translated into multiple proteins with different transcription activation potential due to leaky scanning of a fraction of ribosomes across the first AUG codon and reinitiation at internal AUGs. Transfection of the chicken hepatoma cell line LMH with constructs producing either the 42-kD, or the 29-kD, or the 14-kD protein, clearly shows that the N-terminal part of the protein (region I) harbours the sequences responsible for *trans*-activation in liver cells. For an understanding of the mechanism of the translation multiplicity the following points are relevant. First, the AUG codon of the small 5'ORF is in a strong Kozak sequence (ACCAUGC) with an A at position -3. Therefore, it is expected that ribosomes will efficiently initiate at this AUG. The presence of a strong

AUG within a leader sequence followed shortly by a terminator codon, as is the case in C/EBP $\alpha$  mRNA, is known to reduce but not abolish initiations from a downstream AUG (40,41). It is believed that after ribosomes have translated the minicistron, the 40-S ribosomal subunits remain bound to the messenger RNA, resume scanning and may reinitiate at another AUG downstream. Reinitiation occurs only when the 5'ORF is short, which might have to do with the kinetics of release of the initiation factors (41). The distance of the 5'ORF terminator codon to the next AUG codon appears to be important for reinitiation, since it has been found that expanding the distance increases the efficiency of reinitiation (38). The distance between both reading frames in the C/EBP $\alpha$  mRNA is only short (7 nt), so initiation at the first AUG codon would be at a disadvantage against internal initiations. The Kozak sequences of these particular AUGs will certainly influence reinitiation efficiency. In this respect, it should be noted that the AUG at C/EBP $\alpha$  codon positions 1 is in a productive, though non-optimal Kozak sequence (present results), the AUG codon at position 102 in a more favourable Kozak sequence, and the AUG at codon position 15 in a bad Kozak sequence. Our experiments point to the 5' small ORF as the main mRNA feature responsible for leaky scanning. Removing the small 5'ORF by mutation of its start site decreases the translation of the transcriptionally inactive 29-kD product, probably by reducing the number of ribosomes scanning over the first cC/EBP start site. This is also confirmed by the finding that this mutant construct gives the strongest elevation of transactivation in transfected LMH cells.

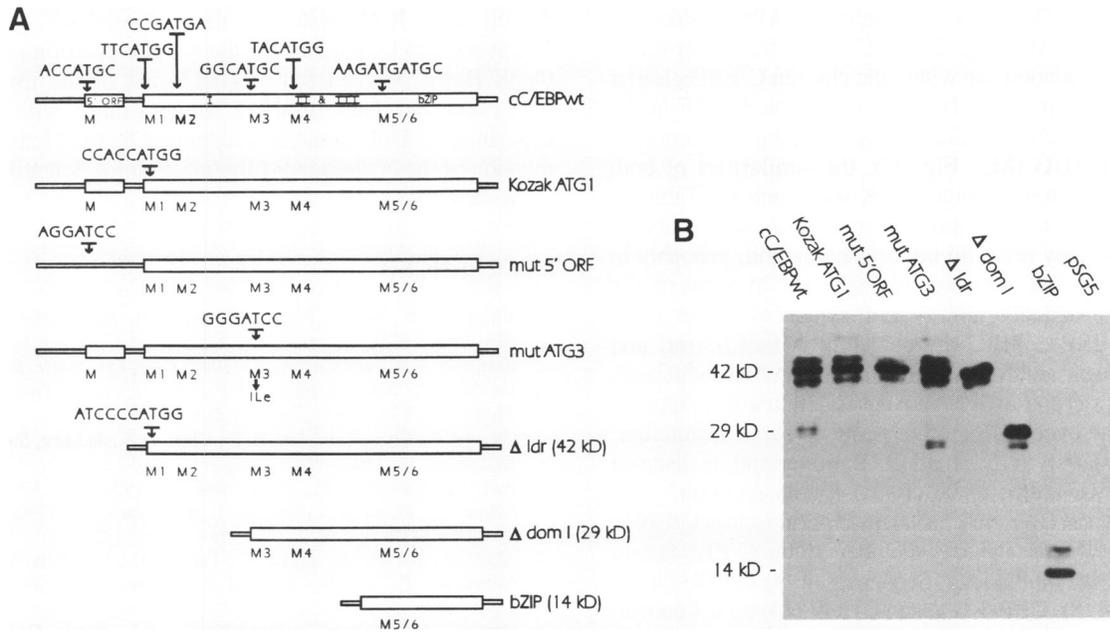
We consider it unlikely that the effect of the mutations on translation is exerted via a change in the secondary structure of the mRNA leader. The initiation codons of the 5'ORF and the cC/EBP $\alpha$  cistron appear not to participate in putative hairpin structures of the very GC-rich leader sequence predicted by the RNAFOLD program (IntelliGenetics). Neither the 3-nt mutation of the 5'ORF start (mut 5'ORF) nor the 4-nt mutation of the C/EBP Kozak sequences (Kozak ATG1) affected the predicted secondary structure of the mRNA.

The conservation of mRNA features between chicken, rodents and *Xenopus*, including the length of the 5' ORF, its spacing with the C/EBP $\alpha$  cistron, the degree of conformance of initiation sequences to the Kozak consensus sequence and the position of the internal AUG start sites, suggests that it may serve a functional purpose. Small open reading frames are present in the 5' untranslated sequences of many mRNAs, but usually position and length are not conserved between species. The results of Ossipow *et al.* (21) and Lin *et al.* (22), and our present data indicate that the C/EBP $\alpha$  mRNAs have evolved to allow the generation of multiple translation products with different transcription activation potential from the same mRNA. Data described on C/EBP $\alpha$ -42/C/EBP $\alpha$ -30 ratios in liver during development of the liver are conflicting. Whereas Ossipow *et al.* (21) finds no changes, Lin *et al.* (22) report to have observed a 2-fold drop in the C/EBP $\alpha$ -42/C/EBP $\alpha$ -30 ratio from birth to 14 days post-natal. Moreover, Lin *et al.* (22) report changing ratios during differentiation of 3T3-L1 preadipocytes. We find that similar ratios of cC/EBP $\alpha$ -42/cC/EBP $\alpha$ -29 are found when different tissues are compared. However, in view of the potent antagonistic action of small amounts of 29-kD product, possibly caused by its enhanced DNA-binding affinity compared with 42-kD product (9; own observation), small variations in the amount of the 29-kD product that may have remained unnoticed

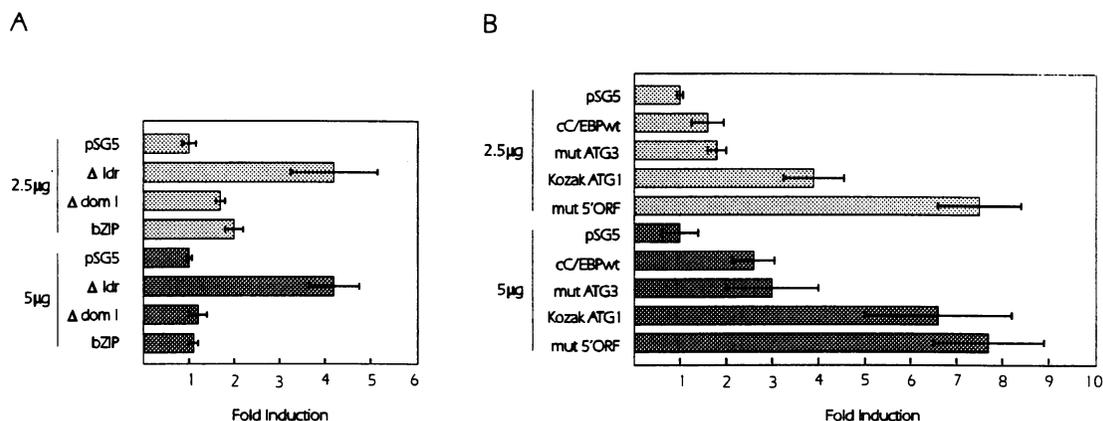
would have dramatic effects on transcription. Because of the conflicting results, the question whether the translation from multiple start sites is regulated, has not been satisfactory answered yet.

Generation of a truncated translation product (LIP) with antagonistic activity in addition to a full-length polypeptide (LAP)

with transcription activation potential has also been observed for C/EBP $\beta$  (9). The mechanism of their initiation has been reported to depend on the strength of the respective initiation codons. However, start site multiplicity of C/EBP $\beta$  may depend on structural features similar to those described here for C/EBP $\alpha$ , specifically on a small 5'ORF. Interestingly, the C/EBP $\beta$  mRNAs



**Figure 3.** Structure and transient expression of chicken C/EBP $\alpha$  gene constructs. (A) Schematic representation of the cC/EBPwt and altered gene constructs with the small 5'ORF and C/EBP $\alpha$  peptide coding regions, including the evolutionary conserved sequences I, II, III and bZIP, indicated. Methionines are nominated M for the 5'ORF and M1 through M6 for the C/EBP coding sequence in order of their occurrence. The sequence context of the potential initiation codons is depicted. For each construct, the essential mutations and deletions of the wt structure are indicated. (B) Immunodetection of chicken C/EBP peptides expressed from different C/EBP constructs after transfection into COS-1 cells (0.085 OD<sub>260nm</sub> each).



**Figure 4.** Transactivation potential of the cC/EBP $\alpha$  wild type and mutant constructs. CAT activities in LMH cells transfected with 2.5  $\mu$ g (D)<sub>9</sub>-tkCAT reporter plasmid and the indicated amount of effector plasmid are expressed relative to a pSG5-transfection with the (D)<sub>9</sub>-tkCAT reporter plasmid. This basal transcription level, which is probably caused by endogenous C/EBP in LMH cells, was set to 1. (A) LMH cells transfected with plasmids yielding no cC/EBP $\alpha$  (pSG5), or exclusively full-length cC/EBP $\alpha$ -42 ( $\Delta$ ldr), cC/EBP $\alpha$ -29 ( $\Delta$ domI) or cC/EBP $\alpha$ -14 (bZIP). (B) LMH cells transfected with plasmids yielding no cC/EBP $\alpha$  mRNA (pSG5), cC/EBP $\alpha$  mRNA with the wt sequence (cC/EBPwt), the third cC/EBP $\alpha$  ATG mutated (mutATG3), ATG1 in an optimal Kozak sequence (KozakATG1) or the 5'ORF abolished (mut5'ORF).

from chicken (= NF-M (17), rat (3) and human (= NF-IL6 (7) all contain a small 5'ORF at 4 nt from the major C/EBP $\beta$  initiator codon and potential initiator codons in similar Kozak sequences, features which to our knowledge have not been noticed earlier (Fig. 2C; Table 1). The similar structural organisation of the C/EBP $\alpha$  and C/EBP $\beta$  leader sequences becomes apparent when the major initiator codons are placed in register. This results in optimal positional identity of six out of seven N-terminal amino acids (Fig. 2A and B). Following this alignment, the upstream in frame AUG codon (M1, Fig. 2C) yielding LAP\* (9) does not correspond to Met 1 (M1, Fig. 2B) of C/EBP $\alpha$  but rather to the non-AUG (CUG) initiation site within the chicken C/EBP $\alpha$  leader (Fig. 2B). Although it has to be proven that the 5'ORF in the C/EBP $\beta$  leader indeed determines leaky scanning of ribosomes across the major AUG (M2, Fig 2C), the similarities of both C/EBP genes, including strength of Kozak context (Table 1) strongly suggests that these features have been conserved during evolution because they serve an important function, probably in generating translational start site multiplicity. In view of the differences in the deduced amino acid sequences of the small 5'ORFs [for C/EBP $\alpha$ ; MPGR(L) (chicken), MPGE(L) (rat) and MLAW(L) (*Xenopus*), and for C/EBP $\beta$ ; MPPHSAAR(L) (chicken), MPPAAARR(L) (rat) and MSPPAAAAAC(L) (human)] it appears not likely that the encoded putative peptides serve a function. More likely translation of the small ORF proper and its distance to the C/EBP ORF seems to be crucial for its function.

The mRNAs of the liver enriched transcription factors C/EBP $\alpha$  and C/EBP $\beta$  appear to be the only members of the C/EBP family designed for start site multiplicity. Sequences of the C/EBP-family members C/EBP $\delta$  (8), CRP-1 (11) and CHOP (14) do not contain small 5'ORFs in their leader sequences.

The presence of a larger than 42-kD full-length product in chicken liver and transfected COS cells requires an explanation. Post-translational modification such as phosphorylation can be refuted as an explanation, because the band heterogeneity is not observed with the cC/EBP $\alpha$ -42 peptide generated from the construct lacking the 5' leader sequence. More likely, the data suggests that the larger product represents initiation from a non-AUG codon within the leader sequence. Taking the apparent size difference of 15 amino acids with the full-length product into account, the larger band could possibly have been initiated at the CUG codons 45 nucleotides upstream of the cC/EBP $\alpha$  initiator codon. CUG codon like these, which are in the favourable translation sequence ACCCUGGGAU, recently have been shown to act as initiation sites in other messenger RNAs (42,43), for example of c-Myc (44). The functional significance of such products is not clear. In the case of chicken C/EBP $\alpha$  it is doubtful that it serves any function because they are not generated from the rat C/EBP $\alpha$  mRNA which lacks the in frame CUG.

In our present study we show the importance of a small ORF located just upstream the C/EBP $\alpha$  ORF in generating translation start site multiplicity, causing the formation of transcription factor isoforms with different transactivation potentials in hepatoma cells. Comparison of C/EBP $\alpha$  and C/EBP $\beta$  mRNA primary structures from different vertebrates reveals that the features necessary for start site multiplicity are conserved between these transcription factor family members. Earlier Descombes and Schibler (9) showed that the ratio of LAP and LIP, the two major C/EBP $\beta$  isoforms, changes during liver differentiation in rat. Whether the ratio between the produced C/EBP $\alpha$  isoforms is or

is not subject to regulation during development, has not been settled yet (21,22). Our studies show no obvious differences in cC/EBP $\alpha$  isoform ratios between liver, lung and small intestine. Further studies are necessary to gain better insight in the possible regulation and the physiological role of the C/EBP $\alpha$  isoforms.

## ACKNOWLEDGEMENTS

We are grateful to A.O.Brinkmann for providing COS-1 cells, V.Poli and R.Cortese for the (D)<sub>9</sub>-tkCAT plasmid and D.Lavery and U.Schibler for the NUN extraction protocol. We thank Harm Nijveen for making some of the mutant DNA-constructs and Jaap Brouwer for assistance with the animal experiments. This research is supported by the Medical Sciences division of the Netherlands Organization for Scientific Research (MW-NWO).

## REFERENCES

- Vinson, C.R., Sigler, P.B., and McKnight, S.L. (1989) *Science*. 246: 911–916.
- Kerppola, T.K., and Curran, T. (1991) *Curr. Op. Struct. Biol.* 1: 71–79.
- Landschulz, W.H., Johnson, P.F., Adashi, E.Y., Graves, B.J., and McKnight, S.L. (1988) *Genes & Dev.* 2: 786–800.
- Chang C.J., Chen, T.T., Lei, H.Y., Chen, D.S., and Lee, S.C. (1990) *Mol. Cell. Biol.* 10: 6642–6653.
- Poli, V., Mancini, F.P., and Cortese, R. (1990) *Cell*. 63: 643–653.
- Roman, C., Platero, J.S., Shuman, J., and Calame, K. (1990) *Genes & Dev.* 4: 1404–1415.
- Akira, S., Ishihiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T., and Kishimoto, T. (1990) *EMBO J.* 9: 1897–1906.
- Cao, Z., Umek, R.M., and McKnight, S.L. (1991) *Genes & Dev.* 5: 1538–1552.
- Descombes, P., and Schibler, U. (1991) *Cell*. 67: 569–579.
- Kageyama, R., Sasai, Y., and Nakanishi, S. (1991) *J. Biol. Chem.* 266: 15525–15531.
- Williams, S.C., Cantwell, C.A., and Johnson, P.F. (1991) *Genes & Dev.* 5: 1553–1567.
- Kalkhoven, C.F., AB, G., and Wijnholds, J. (1992) *Nucleic Acids Res.* 20: 4093.
- Montell, D.J., Rorth, P., and Spradling, A.C. (1992) *Cell*. 71: 51–62.
- Ron, D., and Habener, F. (1992) *Genes & Dev.* 6: 439–453.
- Thomassin, H., Hamel, D., Bernier, D., Guertin, M., and Belanger, L. (1992) *Nucleic Acids Res.* 20: 3091–3098.
- Xu, Q., and Tata, J.R. (1992) *Mech. Dev.* 38: 69–81.
- Katz, K., Kowentz-Leutz, E., Müller, C., Meese, K., Ness, S.A., and Leutz, A. (1993) *EMBO J.* 12: 1321–1332.
- Birkenmeier, E.H., Gwynn, B., Howard, S., Jerry, J., Gordon, J.I., Landschulz, W.H., and McKnight, S.L. (1989) *Genes & Dev.* 3: 1146–1156.
- Descombes, P., Chojkier, M., Lichtsteiner, S., Falvey, E., and Schibler, U. (1990) *Genes & Dev.* 4: 1541–1551.
- Friedman, A.D., Landschulz, W.H., and McKnight, S.L. (1989) *Genes & Dev.* 3: 1314–1322.
- Ossipow, V., Descombes, P., and Schibler, U. (1993) *Proc. Natl. Acad. Sci. USA.* 90: 8219–8223.
- Lin, F.T., MacDougald, O.A., Diehl, A.M., and Lane, M.D. (1993) *Proc. Natl. Acad. Sci. USA.* 90: 9606–9610.
- Wijnholds, J., Muller, E., and AB, G. (1991) *Nucleic Acids Res.* 19: 33–41.
- Green, S., Issemann, I., and Sheer, E. (1988) *Nucleic Acids Res.* 16: 369.
- Kunkel, T.A., Roberts, J.D., and Zakour, R.A. (1987) *Meth. Enzymol.* 154: 367–382.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1994) *Current protocols in molecular biology*. Green publishing associates, Inc. and John Wiley & Sons, Inc. Brooklyn, New York.
- Maire, P., Wuarin, J., and Schibler, U. (1989) *Science* 244: 343–346.
- Kawaguchi, T., Nomura K., Hirayama, Y and Kitagawa, T. (1987) *Cancer Res.* 47: 4460–4464.
- van der Eb, A.J., and Graham, F.L. (1980) *Meth. Enzymol.* 65: 826–838.

30. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A laboratory manual*. Cold Spring Harbor Press. Plainview. New York.
31. Hoppe-Seyler, F., Butz, K., Rittmüller, C., and von Knebel Doeberitz, M. (1991) *Nucleic Acids Res.* 19: 5080.
32. Lavery, D.J., and Schibler, U. (1993) *Genes & Dev.* 7: 1871–1884.
33. Harlow, E., and Lane, D. (1988) *Antibodies: A laboratory manual*. Cold Spring Harbor Press. New York.
34. Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) *Meth. Enzymol.* 185: 60–89.
35. Lichtsteiner, S., Wuarin, J., and Schibler, U. (1987) *Cell.* 51: 963–973.
36. Gorski, K., Carneiro, M., and Schibler, U. (1986) *Cell.* 47: 767–776.
37. Kozak, M. (1989) *J. Cell Biol.* 108: 229–241.
38. Kozak, M. (1987) *Mol. Cell. Biol.* 7: 3438–3445.
39. Friedman, A.D., and McKnight, S.L. (1990) *Genes & Dev.* 4: 1416–1426.
40. Kozak, M. (1991) *J. Cell Biol.* 115: 887–903.
41. Kozak, M. (1991) *J. Biol. Chem.* 266: 19867–19870.
42. Boeck R., and Kolakofsky, D. (1994) *EMBO J.* 13: 3608–3617.
43. Grünert S., and Jackson, R.J. (1994) *EMBO J.* 13: 3618–3630.
44. Hann, S.R., King, M.W., Bentley, D.L., Anderson, C.W., and Eisenman, R.N. (1988) *Cell.* 52: 185–195.