Volume 16 Number 16 1988 Nucleic Acids Research

# Multiple elements are required for expression of an intermediate filament gene

Christina M.Sax, Francis X.Farrell[1], Janet A.Tobian[2] and Zendra E.Zehner[1]*

Department of Human Genetics and [1]Department of Biochemistry and Molecular Biophysics, The Massey Cancer Center, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA 23298 and [2]Human Genetics Branch, NICHD, The National Institutes of Health, Bethesda, MD 20205, USA

Received February 22, 1988; Revised and Accepted July 18, 1988

## ABSTRACT

The expression of vimentin is unique within the intermediate filament multigene family. It is the only member which deviates from its usual tissue-specific expression pattern and whose 5'-flanking region contains multiple GC boxes, the binding site for Sp1. The activity of vimentin 5'-end:CAT fusions has been compared in cells where vimentin is highly expressed (mouse L cells) or not expressed at all (MH1C1). In addition, CAT activity has been examined by microinjection into Xenopus oocytes. Both in vivo expression and in vitro binding studies implicate Sp1 as a general regulatory factor in vimentin gene expression. Increased expression of 5'-end:CAT fusions in mouse L cells suggests that a fibroblast-specific enhancer element resides in the region -321 to -160. Low transcriptional activity in MH1C1 cells may be due to either the lack of this positive transcription factor(s) or the presence of a repressor element. Here, we demonstrate that the unique and complex pattern of vimentin gene expression is controlled by multiple cis-acting elements.

## INTRODUCTION

The proteins of the eukaryotic cytoskeleton provide an elegant system to address the mechanisms of differential gene expression. Components of the three filamentous networks - the microtubules, intermediate filaments, and microfilaments - exhibit both developmental and tissue-specific expression. The intermediate filament protein (IFP) multigene family is composed of five differentially expressed gene classes: the cytokeratins are expressed in epithelial cells, desmin in muscle cells, glial fibrillary acidic protein (GFA) in glial cells, the neurofilaments in neurons, and vimentin in cells of mesenchymal origin (1). Among the IFP family members, only vimentin deviates from this specific expression pattern, because it is often expressed with another IFP member in early stages of cellular differentiation. For example, vimentin is expressed with the cytokeratins in early Xenopus and murine embryonic development (2,3); with the neurofilaments in developing neurons (4,5); with GFA in developing astrocytes (6-8); and with desmin during differentiation from the myoblast to the myotube (9). Moreover, in many established

cell lines, despite their embryonic origin, vimentin either replaces or is co-expressed with the normal cellular counterpart (10-12). The unique ability of vimentin to deviate from its usual tissue-specific expression pattern has prompted us to characterize the structure and regulation of the vimentin gene.

Vimentin is encoded by a single copy gene spanning 8.5 kb of the chicken genome (13-15). Examination of the 5'-flanking region of this gene (13) reveals several interesting features (Fig. 1). The chicken vimentin gene lacks the conventional TATAAATA consensus sequence but does contain a possible variant of this sequence CATAAGAG (16,17). In addition, it contains an inversely oriented CCAAT box on the noncoding strand, and five GC boxes. The GC box [5'-(G/T)GGG(C/A)GGPuPuPy-3'] (18,19) is the recognition sequence for the cellular transcription factors Sp1 (20) and LSF (21), and has been implicated in the regulation of several viral and cellular genes (see 22 for review). In addition, the 5'-flanking sequence of chicken vimentin displays identity with the corresponding regions of the hamster and human gene (13,23,24,47), as well as with known SV40 enhancer elements (13,25).

We have begun analyzing *in vitro* and *in vivo* the elements governing vimentin gene regulation. The presence of multiple GC boxes in the 5'-flanking region of chicken vimentin, their relative scarcity in the remaining 8.5 kb of the gene, and their presence in the flanking region of the hamster (23,24) and human (47) vimentin genes, intimates a transcriptional role for Sp1 in vimentin expression. Our results indicate that Sp1-enriched protein fractions are capable of specifically interacting with at least the three most proximal GC boxes of chicken vimentin. In addition, *in vivo* analysis of the transcriptional capabilities of vimentin 5'-flanking segments indicates the existence of upstream sequence elements which confer tissue-specific regulation on vimentin.

## MATERIALS AND METHODS

### S1 Analysis of the 5'-end of Vimentin mRNA

S1 analysis of the 5'-end of vimentin mRNA was carried out as described previously (15). A 316 bp BanI fragment (-179 to +137) was labeled on the 5'-end by first removing the phosphate moiety with calf intestional phosphatase and then labelling with polynucleotide kinase. The labelled DNA was heated at 90°C in 40% DMSO plus 1.3 mM EDTA and strands separated by gel electrophoresis. Both strands were heteroduplexed overnight at 55°C with chicken total RNA or poly A+ mRNA extracted from breast muscle or fibroblasts or calf liver tRNA as a control (15). S1 (300 U) was added and following digestion (at 37°C for 1 hr) (15), RNA-protected DNA fragments were sized on a 6% sequencing gel.

Preparation of Sp1-Enriched Protein Fractions

Sp1-enriched fractions were prepared from HeLa cells (5-10 liters) grown in suspension culture. An initial crude extract was prepared according to the method of Manley, et al (26). This extract was passed over a heparin agarose column and the 0.25 to 0.4 M KCl fraction collected (27). The Sp1 content of this fraction was assessed by its ability to footprint the previously described GC boxes of SV40 (Fig. 4C) (50) and is referred to as the 0.4 M heparin agarose fraction. Later, extracts were enriched by a single pass over a DNA affinity column using either a 57 bp HpaII DNA fragment containing GC box 3 or a synthetic oligonucleotide prepared as described by Kadonaga and Tjian (28). For the experiments reported here these two affinity column protein fractions were comparable. Gel analyses indicate that these protein fractions do not contain only Sp1, therefore, they are referred to as Sp1-enriched.

Plasmids, DNA Fragments, and Radiolabeling

The radiolabeled fragment used in either gel mobility shift or footprint assays was generated from a subcloned chicken vimentin 5'-flanking fragment. A 231 bp fragment (-160 to +71) was end-labelled (radio-specific activity > $1X10^6$ cpm/ug of DNA) at a pUC18-derived site using the Klenow fragment of DNA polymerase I. The control fragment was a 249 bp fragment generated by AvaII digestion of pBR322 and end-labelled as above.

Fragments derived from the 5'-flanking region of the chicken vimentin gene were cloned into the multi-cloning site of p8CAT. The p8CAT vector is a derivative of pEMBL8 (29) in which the F1 origin of replication has been removed and the bacterial chloramphenicol acetyltransferase gene (CAT) inserted. The cloned fragments include a 57 bp SmaI fragment (pcV/GC3) spanning -162 to -76, and a 334 bp SmaI fragment spanning -496 to -162 (pcV-496/-162). pcV-321F was constructed by first cloning a 394 bp AluI fragment (-321 to +73) into the SmaI site of puc18. The construct was linearized at a BamHI, puc18-derived site, and then digested with Bal31 nuclease to remove the ATG of chicken vimentin. The resulting fragment (-321 to +1) was inserted into the SmaI site of p8CAT. pcV-321R was constructed by digesting pcV-321F with EcoRI and HindIII at p8CAT-derived sites, filling-in with the Klenow fragment of DNA polymerase I, and recloning in the opposite orientation. pcV-160 is derived from pcV-321F, by digestion at a vimentin-derived AvaII site (position -160). The competitive plasmid (pSV40GC) used in microinjections was constructed by cloning the HindIII/PvuII fragment of pSV2-CAT into SmaI/HindIII digested pUC18.

Transfection, Microinjection, and CAT Assay

Intranuclear injection of Xenopus laevis oocytes was performed as described

(30). Twenty ng of each 5'-end:CAT fusion was injected into the germinal vesicle of 10-20 oocytes. Oocytes were incubated at room temperature for 15-20 hrs. Transfections were carried out as previously described (31). Mouse L-cells (fibroblasts) and a rat hepatoma cell line, $MH_1C_1$ (ATCC #CCL144), were plated at $5x10^5$ cells per 100-mm dish one day before transfection. Each plate was transfected with 20 ug of $CaPO_4$-precipitated DNA. Cultures were harvested 48 hrs post-transfection, and CAT assays performed as described (32). Transfected cells were disrupted by repeated freezing and thawing in 0.25 M Tris pH7.8, whereas microinjected oocytes were homogenized in 25 mM Tris pH7.5, 10 mM $MgCl_2$, 1 mM EDTA pH 7.5.

Gel Mobility Shift Analysis

DNA-binding reactions were carried out in 25 mM Tris pH7.9, 6.25 mM $MgCl_2$, 0.5 mM EDTA, 50 mM KCl, 0.5 mM DTT, 2% polyvinylalcohol, and 10% glycerol. The reactions contained a radiolabeled 231 bp chicken vimentin 5'-fragment, 7 fmoles poly(dI-dC) poly(dI-dC), and the affinity column fraction. In competition assays, a 128 bp BglI/SphI DNA fragment of pSV2-CAT containing SV40's 6 GC boxes or the 249 bp pBR322 fragment was added to the binding reaction (33). Reaction mixtures were incubated on ice for 15 min followed by 5 min at room temperature. DNA-protein complexes were electrophoresed and visualized by the adapted method of Strauss and Varshavsky (34). Samples were loaded onto a 5% acrylamide gel prepared in 1 mM EDTA, 3.3 mM NaAcetate, and 6.7 mM Tris pH7.5, and electrophoresed at 130 volts for 4 hrs at 4 C.

Footprint Analysis

Footprinting reactions were carried out as described by Jones, et al (52). Two ng of end-labelled fragment was combined with variable amounts of extract in a 50 ul volume containing the following final buffer concentration: 25 mM Tris-HCl pH 8, 6.25 mM $MgCl_2$, 0.5 mM EDTA, 50 mM KCl, 0.5 mM DTT, 2% polyvinylalcohol, and 10% glycerol. Reactions were incubated on ice for 15 min followed by 2 min at room temperature. Fifty ul of a 5 mM $CaCl_2$ plus 10 mM $MgCl_2$ solution was added, followed by 2.5 ul of a freshly diluted solution (5ug/ml) of DNase I (Pharmacia), and digestion allowed to proceed for 60 sec. The reactions were terminated by a 100 ul solution containing 200 mM NaCl, 20 mM EDTA, 1% SDS, and 250 ug/ml tRNA, extracted with phenol/chloroform, ethanol precipitated, rinsed, and dried before loading on a 6% acrylamide, 8 M urea gel.

Southwestern Analysis

Southwestern analysis was preformed via the modified method of Singh, et al. (41). Fifty ug of Sp1-enriched HeLa nuclear extract was diluted 1:1 (vol/vol) with sample buffer (2% SDS, 100 mM Tris pH 7.5, 280 mM beta-mercaptoethanol,

20% glycerol, and 0.002% phenol-red) and boiled for 2 min. Proteins were size-fractionated on a 10% acrylamide, 0.1% SDS gel in a buffer of 50 mM Tris, 400 mM glycine, and 0.1% SDS. The gel was soaked for 30 min in 25 mM Tris, 190 mM glycine, 20% methanol, and protein transferred to nitrocellulose by electroblotting in the same buffer. Immediately following transfer, the filters were subjected to pre-binding at room temperature for 1 hr in 20 mls of 5% Carnation non-fat dry milk, 50 mM Tris pH 7.5, 50 mM NaCl, 0.1 mM EDTA, and 1 mM DTT. The filters were rinsed in TNED (10 mM Tris pH 7.5, 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT) at room temperature for 2 min. Binding was carried out in 20 mls TNED containing 2.5 ug/ml poly(dI-dC)•poly(dI-dC) and $10^6$cpm/ml of one of two $^{32}$P-end labelled, annealed, double-stranded (DS) oligonucleoties. The DS oligonucleotides used were GC (strands 5'-GATCGGGGCGGGGC-3' and 3'-CCCCGCCCCGCTAG-5') and NS (5'-ATGCGCATCCAAGTATCGCGTA-3' and 3'-TACGCGTAGGTTCATAGCGCAT-5'). Binding was performed at room temperature for 60 min, filters washed twice in TNED at room temperature for 10 min, and exposed to XAR-5 film.

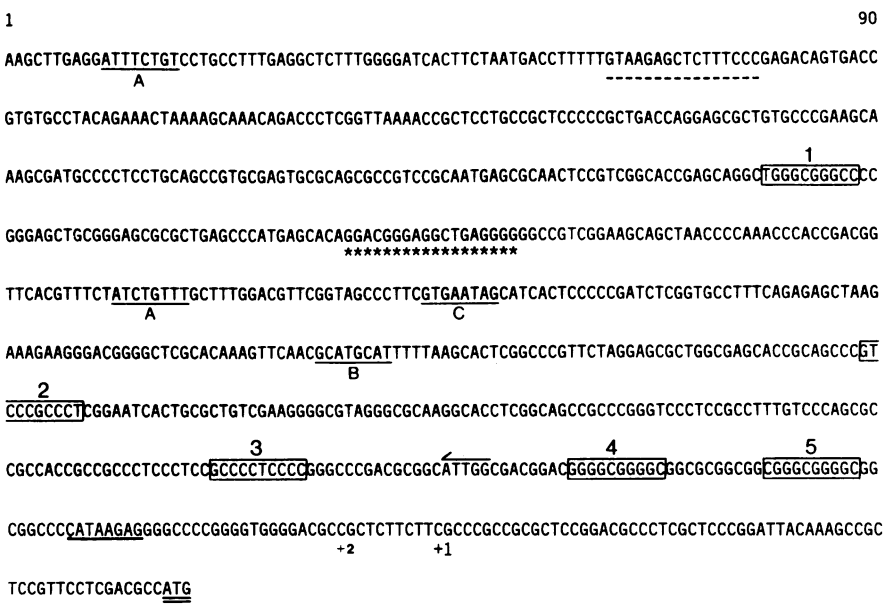



Figure 1. Sequence of the Chicken Vimentin 5'-Flanking Region.
The 5'-flanking sequence of the chicken vimentin gene (13) is shown. Sequence elements are as follows: the 5 GC boxes, numbered 1-5 (▭); the uniquely oriented CCAAT box (←——); an 18 bp (**) and 16 bp (--) identity to the hamster vimentin 5'-flanking region; and areas similar to the A, B, and C enhancer elements of the SV40 genome (underlined and denoted by A, B, or C) (25). +1 and +2 indicate the major and minor start points for transcription downstream of the TATA box variant, CATAAGAG (16,17). The 'ATG' of exon 1 is noted (═).

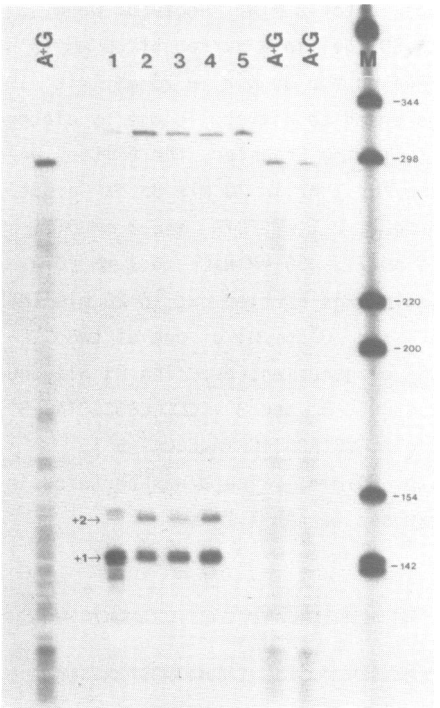



Figure 2. S1 Mapping of Vimentin mRNA. Various RNA's were heteroduplexed with a $^{32}$-P labelled BanI fragment (-179 to +137) spanning the 5'-end of the vimentin gene. The following RNA's (chicken unless noted otherwise) were used: lane 1, 30 ug of total RNA extracted from 14-day embryonic breast muscle; lane 2, 2 ug poly A$^+$ mRNA from breast muscle; lane 3, 5 ug total RNA from embryonic fibroblasts; lane 4, 32 ug from 16- day breast muscle; and lane 5, 30 ug of calf liver tRNA. A+G is a sequencing ladder of the same, asymmetrically labeled, BanI fragment (51). The position of $^{32}$-P labeled double-stranded molecular weight markers is noted.

## RESULTS

### Vimentin 5'-Flanking Sequence Elements

Initially, 800 bp of 5'-flanking region were sequenced (13) and analyzed for the presence of previously described functional regulatory elements (Fig. 1). S1 analysis (Fig. 2) shows that transcription initiates at two sites [a major (+1) and minor (+2) site] 35 and 25 bp downstream, respectively of the putative TATA box variant, CATAAGAG (16,17). In all RNA's assayed the minor initiation site representing 8 to 26% of overall initiation was visible. We feel these two sites are not an artifact of S1 digestion because site 2 is always proportional to site 1 in a variety of RNA's which contain different amounts of vimentin mRNA. Also, in a primer extension experiment (data not shown) two distinct products of comparable size difference (10 bps) were obtained. The chicken vimentin gene is the first IFP gene to exhibit 5'-heterogeneity which may be attributed to its poor TATA box. Earlier, we have shown that this gene produces multiple functional mRNA's via differential utilization of multiple polyadenylation sites (15,16) which has recently been shown for the human vimentin gene (55). The function of this 5'-, 3'- heterogeneity is unknown.

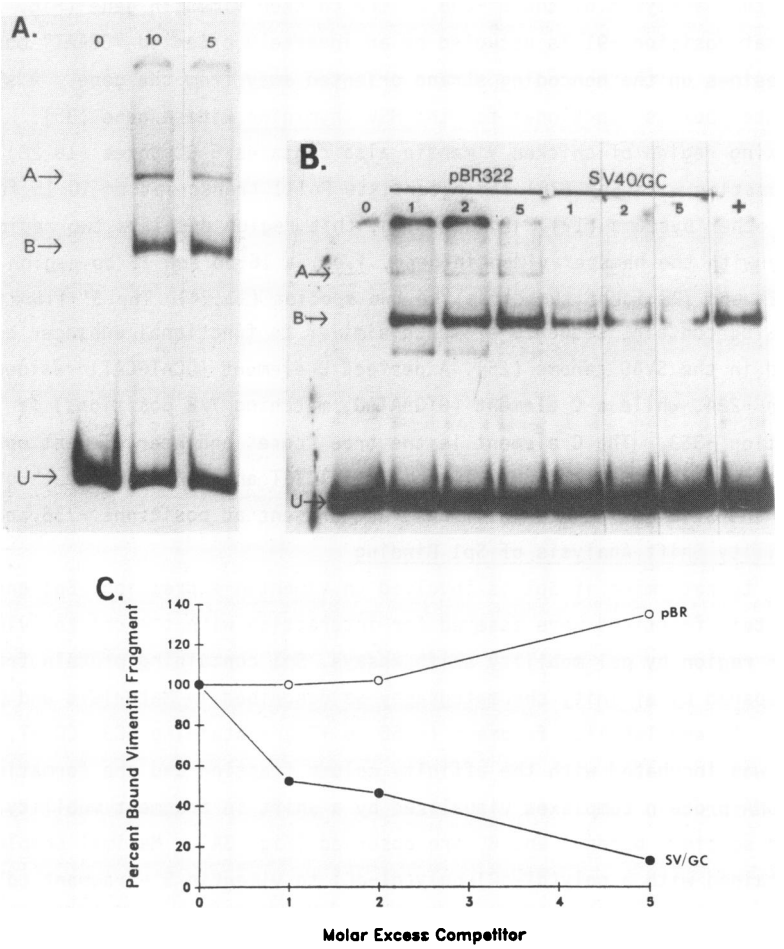



Figure 3. Gel Mobility Shift Analysis of Sp1 Binding.
A.) 14 fmoles of a radiolabeled 231 bp chicken vimentin 5' fragment (+71 to -160, containing GC3-CCAAT-GC4-GC5) was incubated with 7 fmoles of poly(dI-dC)·poly(dI-dC) and 0, 10, and 5 ug of the affinity column protein fraction. U represents the fragment's native electrophoretic mobility whereas A and B denote two protein:DNA complexed species.
B.) 14 fmoles of the fragment described in Fig. 3A, 7 fmoles poly(dI-dC)·poly (dI-dC) and 0 (lane 0), 14 fmoles (lane 1), 28 fmoles (lane 2) or 70 fmoles (lane 5) of competitor pBR322 or SV40/GC fragment were incubated with 0 ug (lane 0) or 5 ug (all other lanes) of affinity column fraction and subjected to electrophoresis. The competitor DNA fragment contains the 6 GC box motifs of SV40 whereas the comparable size pBR322 fragment contains no GC box sequences.
C.) The autoradiographs of Fig. 3B were scanned by densitometry. Since band A represents a minor species and rapidly disappears upon addition of SV40 competitor DNA, only band B is considered here. The percent of vimentin fragment bound (band B) in lanes 1, 2, and 5 pBR322 or SV40/GC is expressed as the percentage of the control, lane +, with no competitor DNA.

Further analysis of the 5'-end of the chicken vimentin gene (Fig. 1) shows that position -91 is occupied by an inversely oriented "CCAAT" box, which resides on the noncoding strand oriented away from the gene. A similarly oriented box is functional for the HSV thymidine kinase gene (35). The 5'-flanking region of chicken vimentin also contains 5 GC boxes (18,28) which in conjunction with Sp1 (20) can potentiate PolII transcription 10-15 fold in several other systems (19). In addition, this region displays two regions of identity with the hamster vimentin gene, i.e., a 16-bp and 18 bp region exhibiting 62% and 72% identity across the two species (13,24). The 5'-flanking region also contains sequence elements similar to functional enhancer elements observed in the SV40 genome (25). A perfect B element (GCATGCAT) resides at position -284, while a C element (GTGAATAG, matching 7/8 positions) is located at position -363. The C element is the true "core" enhancer element observed in many enhancers (36). Two A elements (ATTTCTGT and ATCTGTTT), each of which matches the SV40 box at 6/8 positions, are present at positions -756 and -395.

Gel Mobility Shift Analysis of Sp1 Binding

In order to determine if Sp1 is involved in vimentin expression, Sp1-enriched HeLa protein fractions were assayed for interaction with the chicken vimentin promoter region by gel mobility shift assays. Sp1 containing protein fractions were prepared by affinity chromatography as described in Materials and Methods (26-28). An end-labelled fragment (-160 to +71) containing GC3, CCAAT, GC4, and GC5 was incubated with the affinity column fraction and the formation of stable DNA-protein complexes visualized by a shift in fragment mobility. Two retarded species, bands A and B, are observed (Fig. 3A). Maximal complexes were obtained with a poly(dI-dC)•poly(dI-dC) to vimentin 5'-fragment (dI-dC/VF) molar ratio of 0.5. This ratio was calculated based upon an average dI-dC length of 2423 bp (Pharmacia). Molar ratios in excess of 0.5 block the appearance of a retarded vimentin species (data not shown). Under conditions which favor maximal specific, complex formation, only 3-5% of the complexed fragment resides in band A while 95% resides in band B.

In order to confirm that complex formation is dependent upon the GC box sequence, we have carried out a similar set of analyses using a pBR322-derived fragment (position 886-1135) which lacks a GC box and is similar in size to that of the vimentin 5'-fragment (data not shown). In no instance, does the pBR322 fragment display a shift in electrophoretic mobility.

To further assess the specificity of the observed DNA-protein interactions, we have tested the ability of two heterologous fragments, one containing GC boxes and the other lacking such sequences, to compete with vimentin

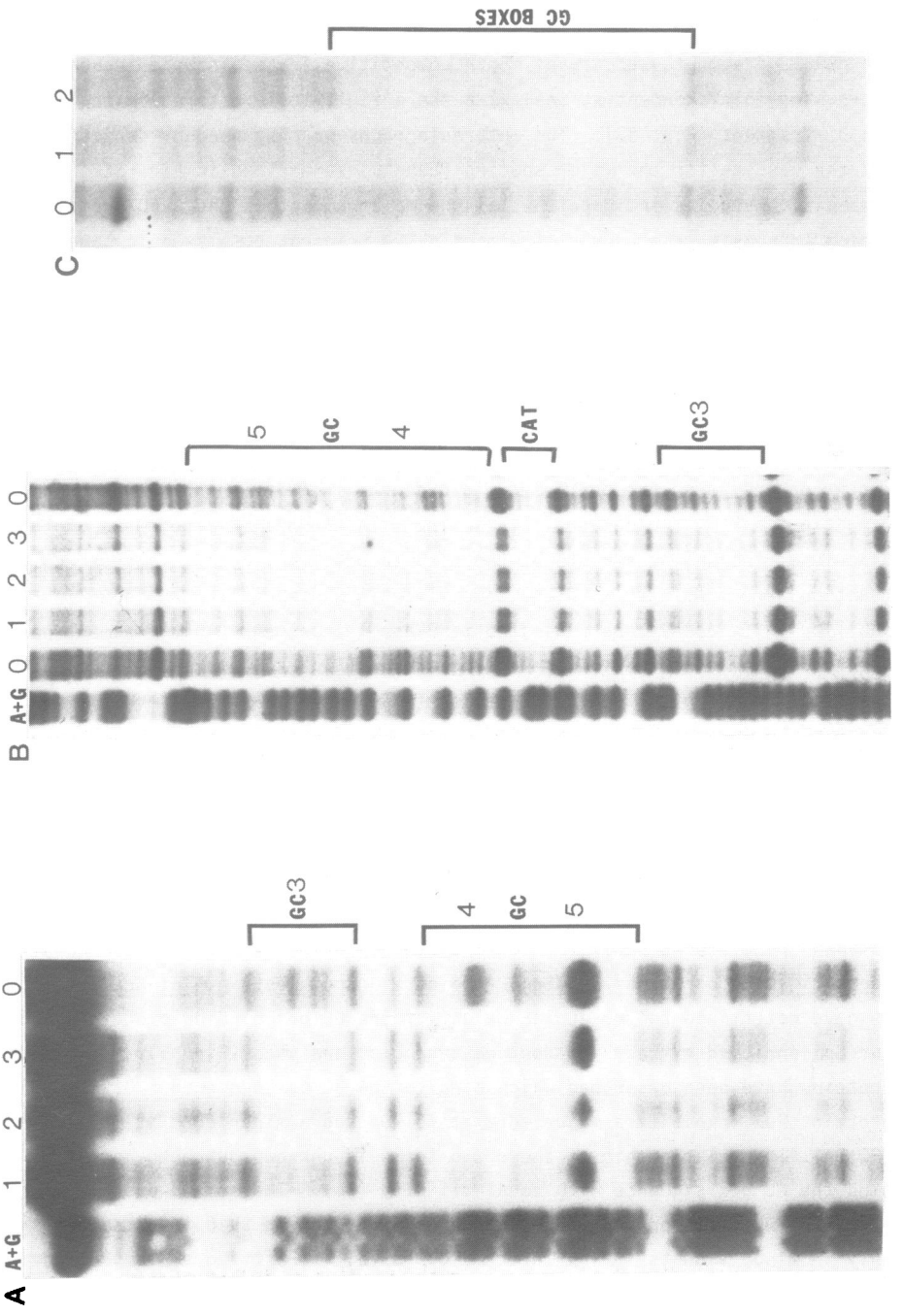
DNA in protein complex formation. The GC box fragment was a 128 bp BglI/SphI fragment containing the 6 GC boxes of SV40 whereas the nonGC box fragment was the pBR322 fragment used above. Both competitive fragments were incubated (at increasing molar concentrations) with the affinity column fraction and vimentin 5'-fragment (Fig. 3B). The autoradiographs were scanned by densitometry, and the results for the major species (band B) summarized in Fig. 3C. These competition studies indicate that as the molar amount of the SV40 promoter region increases, the amount of vimentin DNA fragment complexed in bands A and B decreases. A 1-fold molar excess of SV40 DNA inhibits the formation of band B by 50% and band A has virtually disappeared. A 5-fold excess substantially reduces the appearance of complexed protein vimentin fragment, allowing only 12% of that usually observed. No competition for complex formation was observed between vimentin DNA and the pBR322-derived fragment lacking GC boxes.

Footprint Analysis of Sp1 Binding

Footprint analyses were performed, in order to identify which sequences of the vimentin flanking region were interacting with the Sp1-enriched HeLa protein fractions. Increasing amounts of the 0.25 to 0.4 M KCl heparin agarose columnn fraction were complexed with the same vimentin 5'-fragment used above in gel mobility shift assays and subjected to DNase 1 treatment. This fragment has been end-labelled either on the coding strand (Fig. 4A) at a EcoR1/puc18 derived site or on the noncoding strand (Fig. 4B) at a natural AvaII site. Our results indicate that GC3, GC4, and GC5 are protected from DNaseI cleavage by an Sp1-enriched HeLa protein fraction. However, GC box 3 exhibits a poor footprint on the noncoding strand and a good footprint on the coding strand. For the most part, SV40's GC boxes do not show a strand preference (18) which may not be the case here. Fig. 4C shows the ability of this protein fraction to footprint the well-characterized GC boxes of SV40. The GC boxes of the vimentin gene are as amply protected by these protein fractions as SV40's.

Identification of GC box Binding Proteins via Southwestern Blot Analyses

We have determined the molecular weight of the GC box binding protein(s) in our Sp1-enriched extracts by Southwestern blot analyses (Fig. 5). To date two proteins have been described which bind to the GC box sequence (20,21). One protein, Sp1, exhibits a major form at 95 kDa and a minor form at 105 kDa (20,54). The relationship of these two forms is unknown but both appear to bind the GC box sequence with equal affinity (54). The second protein is LSF (230 kDa) which is present in higher concentrations in SV40 infected cells and only in minor amounts in non-infected cells (21). Fifty ug of our Sp1-enriched (affinity column) HeLa extract was size-fractionated on 10% acrylamide gels


GC BOXES
C
2
1
0
5
GC
4
CAT
GC3
B
A+G
0
1
2
3
0
GC3
4
GC
5
A
A+G
0
1
2
3

along with pre-stained molecular weight markers and transferred to nitrocellulose by electroblotting. Protein samples were run in duplicate. One set was hybridized with the same annealed, double-stranded, GC box oligonucleotide (GS) used in affinity column preparation (Fig. 5A) (see Materials and Methods for sequence). As a control, the other set was hybridized with an annealed, double-stranded, oligonucleotide (NS) which lacks a GC box sequence (Fig. 5B). GS binds predominantly to a protein of molecular weight 95 kDa which is the size of the major form of Sp1. Weaker binding is visible to two other polypeptides at 230 and 105 kDa which could be LSF and the minor form of Sp1 (20,21). Both of these GC box binding proteins are known to exist at lower cellular concentrations than the major form of Sp1 (20,21). Interestingly, we detect weak binding to a protein which corresponds to 130 kDa in size. We know of no GC box binding protein which corresponds to this molecular weight; therefore, this could represent a new GC box binding protein. Importantly, none of these GC box binding proteins interact with the NS oligonucleotide (Fig. 5B).

In Vivo Analysis of Vimentin Regulation

In order to determine which vimentin 5'-flanking sequences are required for tissue-specific expression, we have compared the transcriptional activity of appropriate 5'-end:CAT fusions in selected cell types which express vimentin in a manner consistent with that seen in bona fide tissues *in situ*. For example, in fibroblasts (mouse L cells) it has been shown that as much as 3% of the cellular protein is vimentin (37) whereas in certain hepatoma cell lines ($MH_1C_1$ rat hepatoma) immunological assays detect no vimentin (38). This is as expected since vimentin is not produced in liver tissue where keratin is the liver-specific IFP (1). Moreover, we detect no vimentin mRNA in $MH_1C_1$ cells using cRNA radioactive probes (data not shown). *Xenopus* oocytes were selected for expression studies because it is known that genes which employ Sp1 are expressed in these large cells (39). Although it has been shown that vimentin is expressed here, it is only at low levels since the keratins are the predominant IFP expressed in oocytes (2). Therefore, it seemed reasonable that if vimentin promoter elements function in the oocyte, it would most

Figure 4. DNase 1 Footprinting of the Vimentin 5'-Flanking Region.
The coding (A) and noncoding (B) strand of the chicken vimentin region (-160 to +71) and the 6 GC boxes of the SV40 promoter (C) were labelled as described in Materials and Methods. Two ng of DNA was incubated with 0 (lane 0), 1 ul (lane 1), 2 ul (lane 2) and 3 ul (lane 3) of a 0.4 M KCl heparin-agarose fraction (3.5 mg/ml). Protein:DNA complexes were digested with DNase 1 (5 ug/ml) and analyzed on a 6% polyacrylamide denaturing gel. The A+G sequencing ladder (51) plus the position of the GC and CAT boxes are shown.

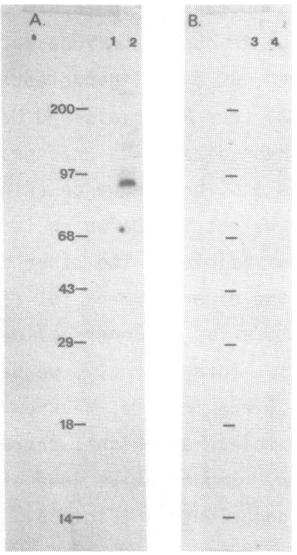



Figure 5. Southwestern Blot Analysis of Sp1-enriched Protein Fraction. Fifty ug of Sp1-enriched HeLa nuclear extract (lanes 2 and 4) and 15 ug of pre-stained protein molecular weight standards (BRL; lanes 1 and 3) were size-fractionated on a 10% acrylamide, 0.1% SDS gel and electroblotted onto nitrocellulose as described in Materials and Methods. Panel A, lanes 1 and 2, was probed with an annealed, double-stranded oligonucleotide "GC" which contains the consensus GC box sequence. Panel B, lanes 3 and 4, was probed with the annealed, double-stranded oligonucleotide "NS" which does not contain a GC box consensus sequence (see Materials and Methods for probe sequence). The migration of molecular weight standards is noted in kilodaltons (kDa).

likely be due to Sp1 elements driving transcription. Using the vector p8CAT, a variety of vimentin 5'-flanking fragments (Fig. 6) were fused to the promoter-less bacterial gene chloramphenicol acetyltransferase (CAT), and then assayed for their ability to express CAT upon microinjection into oocytes, and transfection into L-cells and $MH_1C_1$ cells (Table 1).

The CAT-fusion pcV-321F spans the region -321 to +1 and is fused to the CAT gene in its forward orientation. It contains the B box (-284), GC2 (-228), GC3 (-115), the CCAAT box (-91), GC4 (-78) and GC5(-58). This region is

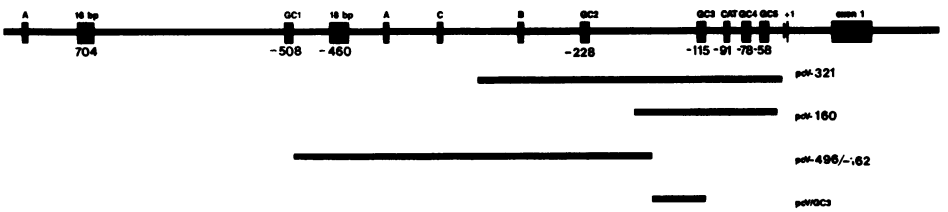



Figure 6. Vimentin 5'-Flanking Region and Promoter-CAT Fusions. A schematic representation of the chicken vimentin 5'-flanking region as shown in Fig. 1. GC1 (-508), GC2 (-228), GC3 (-115), GC4 (-78), and GC5 (-58) identify the five flanking GC boxes; CAT denotes the inversely oriented 5'-CCAAT-3' box (-91); A (-756) and (-395), B (-284), and C (-363) are regions similar to SV40 functional enhancer elements; and 18 bp (-460) and 16 bp box (-704) denote two regions with 72% and 62% identity to regions in the chicken and hamster genes. The position of various vimentin 5'-end fragments fused to the CAT gene and analyzed for promoter activity are shown below.

Table 1. Microinjection and Transfection of Promoter-CAT Fusions.

| | Xenopus oocyte | | Mouse L-cell | | Rat $MH_1C_1$ | |
|---|---|---|---|---|---|---|
| CAT-fusion | pmole/ug | Fold | pmole/ug | Fold | pmole/ug | Fold |
| p8CAT | 5.9 | 1.0 | 2.5 ± 1.4 | 1.0 | 0.026 ± 0.010 | 1.0 |
| pcV-321F | 67.8 ± 6.6 | 11.5 | 379 ± 37 | 151 | 0.044 | 1.7 |
| pcV-321R | 82.6 | 14.0 | 318 | 127 | 0.073 | 2.8 |
| pcV-160 | 41.0 ± 6.2 | 6.9 | 195 | 78 | 0.057 | 2.2 |
| pcV-496/-162 | 45.9 ± 7.5 | 7.8 | 255 ± 23 | 102 | 0.005 ± 0.001 | 0.2 |
| pcV/GC3 | 3.8 | 0.6 | 2.1 | 0.8 | 0.002 | 0.08 |
| pSV2-CAT | 22.2 | 3.8 | 973 | 389 | 0.28 | 10.8 |
| pcV-321F + pSV40GC | 12.1 | 2.1 | - | - | - | - |
| pcV-160 + pSV40GC | 17.0 | 2.9 | - | - | - | |

Fragments derived from the chicken vimentin 5'-flanking region were fused to the bacterial CAT gene and microinjected into Xenopus oocytes, or transiently transfected into mouse L-cells and a rat hepatoma cell line, $MH_1C_1$. F and R denote the forward and reverse orientation of the fragment fused to CAT. The activity of each CAT-fusion is expressed as the pmoles of $^{14}$C-chloramphenicol acetylated per ug of cellular protein (pmoles/ug). 'Fold' denotes the -fold induction of CAT activity for a vimentin 5'-end:CAT fusion relative to the promoterless p8CAT vector. In cases where the CAT-fusion has been transfected and assayed multiple times, the mean CAT activity and standard deviation are shown.

capable of promoting CAT activity 151-fold over that of p8CAT in L-cells and 11.5-fold in oocytes (Table 1). Representative CAT assays of mouse L cell activity are shown in Fig. 7A. Thus, region -321 to +1 of chicken vimentin contains sequence elements required for gene expression in cell systems where vimentin is usually expressed. In addition, this level of activity is comparable regardless of fragment orientation, i.e., pcV-321F is roughly equal to pcV-321R. When pcV-321F or pcV-160 is co-injected with equimolar amounts of the SV40 enhancer + GC boxes (pSV40GC) into oocytes, CAT activity is reduced by 75% and 88%, respectively. This reduction suggests that the SV40 enhancer and/or GC boxes can compete with vimentin DNA in vivo presumably by binding to common (limiting) transcription factors. In contrast, pcV-321F does not direct efficient CAT gene expression (1.7-fold) in $MH_1C_1$ cells (Fig. 7B) where pSV2-CAT is expressed (10.8-fold), but at levels lower than mouse L cells (389-fold) and higher than in oocytes (3.8-fold) (Table 1).

Subclones of the vimentin 5'-flanking region were constructed to assess the contribution of individual elements to overall transcriptional strength.

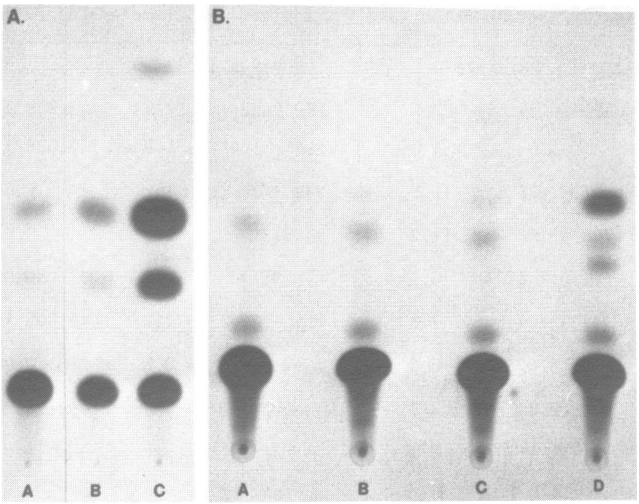



Figure 7. CAT Activity of Vimentin 5'-End:CAT Fusions In Vivo.
A.) L-cells were transiently transfected with p8CAT (A), pcV/GC3 (B), or pcV-321F (C) and the extracts assayed for CAT activity by TLC separation of the unacetylated (lowest species) and acetylated forms of $^{14}$C-chloramphenicol. B.) $MH_1C_1$ cells were transiently transfected with p8CAT (A), pcV/GC3 (B), pcV-321F (C), or pSV2-CAT (D) and CAT activity assayed as described in Fig. 7A.

pcV-496/-162 contains a vimentin fragment spanning the region -496 to -162. While this fragment lacks the proximal promoter region, it does contain considerable 5'-flanking sequences including GC2 (-228). pcV-496/-162 promotes CAT activity (102-fold) in L-cells, but only 7.8-fold over p8CAT in oocytes. Whereas, pcV-160 containing GC3-5, the "CCAAT" box and only 45 bp upstream of GC3 also yields CAT activity of 78-fold over p8CAT in mouse L cells. It would appear that removal of 3 out of 4 GC boxes reduces gene activity only 32% in L cells. In $MH_1C_1$ cells pcV-160 yields comparable activity to pcV-321F, but pcV-496/-162 does not promote transcription (0.2-fold). pcV/GC3 contains a 57 bp fragment spanning -162 to 105, and GC3 (-115) but no other known regulatory element. This fragment (GC3) does not act as an efficient promoter in any of the cell systems assayed (Table 1). The lack of significant CAT activity arising from pcV/GC3 in L-cells and $MH_1C_1$ cells is shown in Fig. 7B.

## DISCUSSION

The IFP multigene family is composed of a complex set of homologous protein molecules which display a highly tissue-specific and developmental pattern of gene expression (1). This family is believed to have arisen from a

single gene which duplicated and then diverged to meet the unique structural and cytoskeletal requirements of highly differentiated cells (40). Like other members of the IFP multigene family, vimentin exhibits a tissue-specific expression pattern (mesenchymal), but is unique in that during early stages of cellular differentiation, it is often expressed in addition to the constituent IFP, and is then down-regulated as differentiation proceeds (2-9). Also, unlike all other IF proteins, vimentin is expressed in most cells grown in culture, regardless of their embryonic origin (10-12). These unique properties suggest that vimentin represents a primordial member of the IFP family which probably falls under both tissue-specific and global mechanisms of control.

The functional role of the vimentin GC boxes was studied in vitro using HeLa protein fractions enriched for Sp1 and a vimentin fragment containing GC3, GC4, and GC5. All three of these boxes are expected to interact with Sp1, since they exhibit a high degree of identity with the consensus sequence (10/10 matches for GC3 and GC4, and 9/10 match for GC5) (13). Gel mobility shift assays using the affinity column fraction result in the appearance of two retarded species (bands A and B) which are present over a wide-range of protein (Fig. 3) and noncompetitive carrier DNA (data not shown). Three lines of evidence suggest that these species arise from specific interactions between vimentin's proximal GC boxes and the Sp1-enriched fraction. First, a control pBR322-derived fragment which does not contain a GC box sequence fails to exhibit a shift in electrophoretic mobility when incubated with the affinity column fraction. Secondly, the incorporation of a SV40 fragment containing 6 GC boxes effectively competes with vimentin DNA for protein interaction. At a 5-fold molar excess of SV40 DNA, both retarded species A and B virtually disappear. Third, a pBR322 derived fragment lacking a GC box sequence, fails to compete even at a 5-fold molar excess with vimentin DNA for protein binding.

The appearance of two bands in our gel mobility shift assays (Fig. 3) is interesting and has been noted for SV40 early promoter:Sp1 interaction (21). It may be explained by a number of possibilities. First, multiple GC binding sites in close proximity may preclude maximum factor binding and give rise to more than one conformer as trapped in the gel mobility shift assay. In this case, we might expect to see an increase in species A (largest molecular weight) as the protein concentration is increased, but at all protein concentrations assayed, species A is always considerably less than species B. DNA footprint analyses (Fig. 4) indicate that all 3 GC boxes are capable of binding protein. A second possibility is that our Sp1-enriched protein fractions contain other GC box binding proteins like LSF which has a substantially

greater molecular weight than Sp1 and exhibits a different gel mobility shift pattern (21). Indeed, Southwestern blots indicate that LSF is present in our affinity column extract at 5% of the Sp1 level. Interestingly, this is also the proportion of species A to B. In conclusion, we feel that the two retarded species obtained in Fig. 3 are due to multiple GC binding sites and cognate factors (Sp1, LSF) present in varying amounts in our protein extracts.

To further substantiate that the retarded species is indeed a specific interaction between GC boxes and Sp1, footprint analyses indicate that vimentin's most proximal GC boxes (GC3, GC4, and GC5) are protected from DNase 1 cleavage by protein binding. It is interesting to note that by their spacing GC3, GC4, GC5 and CAT are all within 115 bp of the initiation site and lie on the same face of the double helix. Whether there is a preference for initial binding at one site followed by cooperative binding to other sites is unknown. The position of the CCAAT box between GC3 and GC4, and binding of its cognate transcription factor might also play a role in GC box utilization in vivo.

Southwestern analyses provide good evidence that the major protein interacting with our double-stranded GC box sequence is Sp1 as suggested by its distinct molecular weight. Interestingly, three other minor species are detectable, two of which match known GC box binding proteins. None of these proteins interact with comparable DNA lacking a GC box sequence (Fig. 5B).

The functionality of vimentin's GC boxes has also been tested in vivo and it would appear that a single GC box acting alone does not constitute an efficient promoter since pcV/GC3 exhibits little activity. However, GC3 is protected in vitro from DNaseI cleavage (on the coding strand) by Sp1-enriched fractions; therefore, its inability to act as an efficient promoter is not due to an inability to bind Sp1. Moreover, only fragments containing multiple GC boxes are capable of promoting substantial CAT activity in oocytes and L-cells. Removal of the proximal promoter region (containing GC3-5) in pcV-496/-162 results in a 33% reduction of CAT activity in L-cells. In view of these results, we suggest that in order for Sp1 to act efficiently in chicken vimentin transcription, multiple GC boxes and additional regulatory elements are required. This may also be the case for other GC box genes such as the human metallothioneine IA gene, where a single GC box responsive to Sp1 is thought to act in concert with other multiple regulatory elements (44). Other examples include the Ha-ras gene where either 3 or 6 GC boxes direct efficient transcription in vivo, but 2 boxes support only 25% of the normal gene activity, and in the HSV-tk gene where mutations in one of two GC boxes results in a 5-15% drop in transcriptional capability of this promoter (43). In support

of this theory is the multiple nature of the GC box in several promoters (22).

Vimentin is the only member of the large IFP family known to harbor multiple GC boxes in its 5'-flanking region. Three other IFP genes have been reported to contain only a single GC box; the bovine cytokeratin no. 19 gene (45), the human 50-kDa cytokeratin gene (46), and the hamster desmin gene (24). All exhibit either an 8/10 or 9/10 match with the consensus GC box sequence, but our results would suggest that these single GC boxes are not efficient promoters. Rittling et al proposed (without data) that the multiple GC boxes of the vimentin genes were not important for expression because their relative positioning was not conserved across species (47). On the contrary, a direct comparison of the relevant 5'-flanking regions of these three genes (Fig. 8) shows considerable conservation in GC box orientation and placement (13,23,47). Moreover, a fragment of the chicken gene containing GC3-5 and only 40 bp upstream when fused to CAT yields considerable activity (78-fold) in mouse L cells. Our deletion analysis coupled with DNA:protein binding studies demonstrate that Sp1 and GC boxes are probably involved in vimentin gene expression. As noted by Kadonaga et al., it is no longer apparent what genes harboring functional GC boxes have in common (53).

While Sp1 probably does not account for the tissue- or developmental-specific expression of vimentin, it is clearly an important component of vimentin's diverse expression pattern. We think Sp1 promotes a basal level of vimentin expression which may be modulated in either a positive or negative fashion by additional, developmental or tissue-specific factors. The existence of such factors is suggested when one compares the promoter activity of various vimentin 5'-fragments in Xenopus oocytes, mouse L-cells, and a rat hepatoma cell line (Table 1). These three systems differ in the extent to which they

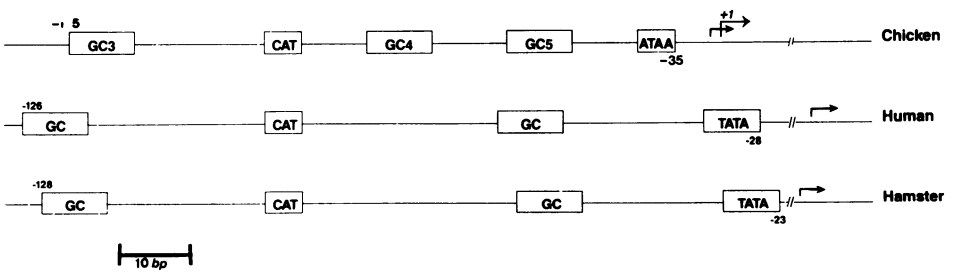



Figure 8. Comparison of the 5'-Flanking Region of Three Vimentin Genes. The 5'-flanking region of the chicken (13), human (47) and hamster (23) vimentin genes has been aligned relative to the common CAT box. GC and TATA boxes (ATAA for chicken) are shown. The arrows denote the start point for transcription of the hamster and human genes whereas +1 marks the major site for chicken.

express vimentin. While Xenopus oocytes express vimentin, they do so at a much lower level than L-cells (2,37). Most of this expression is probably due to GC boxes and Sp1 because co-injection with pSV40GC reduces gene activity considerably. $MH_1C_1$ cells, on the other hand, have not been observed to express vimentin (38). Both pcV-321F and pcV-496/-162 express CAT activity in oocytes and L-cells. However, this activity is substantially increased (10 to 15-fold) in L-cells where vimentin is the sole IFP (37). Although pcV-496/-162 lacks the proximal promoter region containing GC3-5, it is still capable of significant expression (100-fold) in L-cells and even in oocytes (7.8-fold). We do not feel that this level of activity could be due to its single GC box as we have already shown a single GC box does not suppport gene activity in vivo. Instead, our results suggest that vimentin gene expression is further modulated by regulatory elements in addition to GC boxes and their factors. The combined results of pcV-321F and pcV-496/-162 expression in L-cells (and perhaps oocytes) suggest the existence of an upstream enhancer element which can modulate vimentin expression in a tissue-specific manner. This enhancer like other such elements is functional regardless of orientation. L cells (and perhaps oocytes) must synthesize a postive factor which binds in this region, and this factor must be lacking in hepatocytes where vimentin is not expressed in situ (38).

A comparison of the relative CAT activity of pcV-321 to pSV2CAT in $MH_1C_1$ versus mouse L cells is roughly equivalent; therefore, GC box promoter elements are active in $MH_1C_1$ cells albeit at low levels. Moreover, removal of these GC and CAT boxes in $MH_1C_1$ cells (construct pcV-496/-162) completely abolishes gene activity, indicating that initial transcriptional activity must reside in these elements. The paradox is that in situ the natural vimentin gene is not expressed (38). One possibility is that a negative element exists in vivo which overrides GC box promoter activity in tissues where vimentin is not expressed. Alternatively, fluctuating levels of Sp1 could account for decreased gene expression, but this is unlikely because Sp1 has been found in every tissue analyzed to date (54). Therefore, we feel changing levels of Sp1 (or LSF) could not account for tissue-specific expression of vimentin. Instead, both we and Rittling et al (47) have identified an upstream region (We have delineated this region to 40 bp) which appears to totally silence gene activity (data not shown). Currently, experiments are in progress to define this element and determine its role in vimentin gene regulation.

To date, our findings indicate that expression of the chicken vimentin gene is controlled by multiple cis-acting elements, one of which is Sp1 and GC boxes. The results of our functional studies suggest additional upstream

regulatory elements which include a tissue-specific enhancer (-321 to -160) and possibly a repressor element within the 5'-flanking region of the gene. Identification and characterization of these additional sequence elements and their cognate protein factors is essential to understanding vimentin's unique expression pattern within the IFP multigene family.

ACKNOWLEDGMENTS

This work was supported by United States Public Health Service Grant AM33310 and a grant from the Muscular Dystrophy Association to Z.E.Z.

*To whom correspondence should be addressed

REFERENCES

1. Lazarides,E. (1980) Nature 283:249-256.
2. Godsave,S.F., Anderton,B.H., Heasman,J., and Wylie,C.C. (1984) J. Embryol. Exp. Morph. 83:169-187.
3. Lane,E.B., Hogan,B.L., Kurkinen,M., Garrels,J.I. (1983) Nature 303:701-704.
4. Bignami,A., Raju,T., Dahl,D. (1982) Develop. Biol. 91:286-295.
5. Drager,U.C. (1983) Nature 303:169-172.
6. Schnitzer,J., Franke,W.W., Schachner,M. (1981) J. Cell Biol. 90:435-447.
7. Tapscott,S.J., Bennett,G.S., Toyama,Y., Kleinbart,F., Holtzer,H. (1981) Dev. Biol. 86:40-54.
8. Yen,S.-H., Fields,K.L. (1981) J. Cell Biol. 88:115-126.
9. Bennett,G.S., Fellini,S.A., Toyama,Y., Holtzer,H. (1979) J. Cell Biol. 82:577-584.
10. Franke,W.W., Schmid,E., Winter,S., Osborn,M., Weber,K. (1979) Exp. Cell Res. 123:25-46.
11. Traub,U.E., Nelson,W.J., Traub,P. (1983) J. Cell Sci. 62:129-147.
12. Virtanen,I., Lehto,V.-P., Lehtonen,E., Vartio,T., Stenman,S., Kurki,P., Wager, O., Small, J., Dahl, D., Badley, R.A. (1981) J. Cell Sci. 50:45-63.
13. Zehner,Z.E., Li,Y., Roe,B.A., Paterson,B.M., Sax,C.M. (1987) J. Biol. Chem. 262:8112-8120.
14. Zehner,Z.E., Paterson,B.M. (1983) Proc. Natl. Acad. Sci. 80:911-915.
15. Zehner,Z.E., Paterson,B.M. (1983) Nucl. Acids Res. 11:8317-8332.
16. Goldberg,M. (1979) Ph.D. Thesis, Stanford University, Stanford, CA.
17. Hardison,R.C., Butler,E.T., Lacy,E., Maniatis,T., Rosenthal,N., Efstratiadis,A. (1979) Cell 18:1285-1297.
18. Jones,K.A., Kadonaga,J.T., Luciw,P.A., Tjian, R. (1986) Science 232:755-759.
19. Kadonaga,J.T., Jones,K.A., Tjian,R. (1986) Trends Biochem. Sci. 11:20-23.
20. Gidoni,D., Dynan,W.S., Tjian,R. (1984) Nature 312:409-413.
21. Kim,C.H., Heath,C., Bertuch,A., and Hansen,U. (1987) Proc. Natl. Acad. Sci. 84:6025-6029.
22. Dynan,W.S., Tjian,R. (1985) Nature 316:774-778.
23. Quax,W., Egberts, W.V., Hendriks, W., Quax-Jeuken, Y., Bloemendal, H. (1983) Cell 35:215-223.
24. Quax,W., van den Broek,L., Egberts,W.V., Ramaekers,F., Bloemendal,H. (1985) Cell 43:327-338.
25. Herr,W., Gluzman,Y. (1985) Nature 313:711-714.
26. Manley,J.L., Fire,A., Cano,A., Sharp,P.A., Gefter,M.L. (1980) Proc. Natl. Acad. Sci. 77:3855-3859.

27. Dynan,W.S., Tjian,R. (1983) Cell 32:669-680.
28. Kadonaga,J.T., Tjian,R. (1986) Proc. Natl. Acad. Sci. 83:5889-5893.
29. Dente,L., Cesareni,G., Cortese,R. (1983) Nucl. Acids Res. 11:1645-1655.
30. Zasloff,M. (1983) Proc. Natl. Acad. Sci. 80:6436-6440.
31. Graham,F.L., Van Der Eb,A.J. (1973) Virology 52:456-467.
32. Gorman,C. (1985) in Glover, D.M. (eds), DNA Cloning: A Practical Approach. IRL Press, Washington, D.C., Vol II, pp. 143-190.
33. Gorman,C.M., Moffat,L.F., Howard,B.H. (1982) Mol. Cell. Biol. 2:1044-1051.
34. Strauss,F., Varshavsky,A. (1984) Cell 37:889-901.
35. Graves,B.J., Johnson,P.F., McKnight,S.L. (1986) Cell 44:565-576.
36. Weiher,H., Konig,M., Gruss,P. (1983) Science 219:626-631.
37. Lilienbaum,A., Legagneux,V., Portier,M.-M., Dellagi,K., Paulin,D. (1986) EMBO J. 5:2809-2814.
38. Franke,W.W., Mayer,D., Schmid,E., Denk,H., Borenfreund,E. (1981) Exp. Cell Res. 134:345-365.
39. Eisenberg,S.P., Coen,D.M., McKnight,S.L. (1985) Mol. Cell. Biol. 5:1940-1947.
40. Krieg,T.M., Schafer,M.P., Cheng,C.K., Filpula,D., Flaherty,P., Steinert, P.M., Roop,D.R. (1985) J. Biol. Chem. 260:5867-5870.
41. Singh,H., LeBowitz,J.H., Baldwin,A.S., Sharp,P.A. (1988) Cell 52:415-423.
42. Ishii,S., Kadonaga,J.T., Tjian,R., Brody,J.N., Merlino,G.T., Pastan, I. (1986) Science 232:1410-1413.
43. McKnight,S.L., Kingsbury,R. (1982) Science 217:316-325.
44. Lee,W., Haslinger,A., Karin,M., Tjian,R. (1987) Nature 325:368-372.
45. Bader,B.L., Magin,T., Hatzfield,M., Franke,W.W. (1986) EMBO J. 5:1865-1875.
46. Marchuk,D., McCrohon,S., Fuchs,E. (1985) Proc. Natl. Acad. Sci. 82:1609-1613.
47. Rittling,S.R., Baserga,R. (1987) Mole. and Cell. Biol. 7:3908-3915.
48. Davidson,I., Fromental,C., Augereau,P., Wildeman,A., Zenke,M., Chambon, P. (1986) Nature 323:544-548.
49. Larsen,P.R., Harney,J.W., Moore,D.D. (1986) Proc. Natl. Acad. Sci. 83:8283-8287.
50. Dynan,W.S., Tjian,R. (1983) Cell 35:79-87.
51. Maxam,A.M., Gilbert,W. (1980) Meth. Enzymol. 65:499-650.
52. Jones,K.A., Yamamoto,K.R., Tjian,R. (1985) Cell 42:559-572.
53. Kadonaga,J.T., Carner,K.R., Masiarz,F.R., Tjian,R. (1987) Cell 51:1079-1090.
54. Briggs,M.R., Kadonaga,J.T., Bell,S.P., Tjian,R. (1986) Science 234:47-52.