

**INTERACTION OF NUCLEAR FACTORS
WITH THE PROMOTER OF THE
HUMAN PRO- α 1(I) COLLAGEN GENE**

RIEKELTJE KOEDOOD
B.Sc., B.Sc. Honns (Biochemistry)

**Thesis submitted in the Department of Biochemistry of the
Potchefstroomse Universiteit vir Christelike Hoër Onderwys in
partial fulfilment of the requirements for the degree Magister
Scientiae (Biochemistry).**

Supervisor: Prof. Dr. W.J. de Wet

**POTCHEFSTROOM
1993**

GENE

As jy die mens se weefsels penetreer
met instrumente selle dissekteer
om selfs die kerne oop te sny
om stringe gene te kan kry
dink jy
jy het dán die groot geheim ontdek?

As jy die gene-snoere neem
en tel en noem hul een vir een
soos mens krale kan bevoel, bevat
sodat jy weet wie volg na wat
dink jy
jy het dán die groot geheim ontdek?

As jy 'n geenfout kan bepaal
deur stringe uitmekaar te haal
en 'n siekte of gebrek
se oorsaak oop te vlek
dink jy
jy het dán die groot geheim ontdek?

Die groot geheim lê dieper as 'n geen:
Die Skepper-God het stof geneem
en gene in sy mense-skepsels ingeplant
om selfs verskil te maak by bloedverwant
maar die groot geheim het almal in gemeen
dis nie versteek in stringe-stringe geen ...
God het ons Sy Beeld verleen!

- Langenhoven

To my Father in Heaven without whom
none of this would have been possible.

INDEX

ABSTRACT	1
OPSOMMING	3
CHAPTER 1: INTRODUCTION	5
1.1 REGULATION OF GENE EXPRESSION	5
1.1.1 Promoters	5
1.1.2 Enhancers	6
1.1.3 Transcription factors	7
1.2 COLLAGEN	11
1.2.1 Collagen genes	11
1.2.2 Regulation of collagen synthesis	12
1.2.3 Transcriptional regulation of the pro- α 1(I) collagen gene	13
1.3 AIM OF THE STUDY AND APPROACH	15
CHAPTER 2: FUNCTIONAL ANALYSIS OF THE PRO- α 1(I) COLLAGEN PROMOTER	19
2.1 INTRODUCTION	19
2.2 CONSTRUCTION OF PLASMIDS	19
2.3 TRANSFECTION OF CELLS	28
2.3.1 Methods	28
2.3.2 Results	31
2.4 CONCLUSION	35
CHAPTER 3: BINDING OF NUCLEAR FACTORS TO THE PRO- α 1(I) COLLAGEN PROMOTER	37
3.1 INTRODUCTION	37
3.2 DNA MOTIFS IN THE PROMOTER	37
3.3 DNA-PROTEIN INTERACTIONS WITH THE PROMOTER	40
3.3.1 Electrophoretic mobility shift assay	40
3.3.2 DNAase I footprinting analysis	44
3.4 BINDING OF NUCLEAR FACTORS TO DNAase I PROTECTED ELEMENTS	47
3.4.1 Electrophoretic mobility shift assay	47
3.4.2 Southwestern blotting analyses	50
3.5 CONCLUSION	54

CHAPTER 4: DISCUSSION	56
4.1 INTRODUCTION	56
4.2 FUNCTIONAL ANALYSES OF THE PRO- α 1(I) PROMOTER	56
4.3 BINDING OF NUCLEAR FACTORS TO THE PRO- α 1(I) PROMOTER	59
4.3.1 Col-B	63
4.3.2 Pyrimidine rich region	65
4.3.3 CCAAT-boxes	70
4.3.4 Other elements	76
4.4 CONCLUSION	77
LIST OF ABBREVIATIONS	81
REFERENCES	83
ACKNOWLEDGEMENTS	97

ABSTRACT

INTERACTION OF NUCLEAR FACTORS WITH THE PROMOTER OF THE HUMAN PRO- α 1(I) COLLAGEN GENE.

At least thirteen types of collagen have been identified, each having a characteristic tissue distribution. Type I, the most abundant collagen, consists of two α 1 and one α 2 polypeptides. There is considerable evidence that expression of the genes coding for the type I subunits is regulated on the level of transcription. In the pro- α 1(I) collagen gene, regulatory regions have been identified both upstream of the transcription initiation site and in the first intron, with the minimal promoter contained in 250 bp immediately upstream of the start site for transcription. This 250 bp region has been shown to be sufficient to direct transcription of heterologous genes in different systems (Rossouw *et al*, 1987; Rippe *et al*, 1989; Liska *et al*, 1990 and Karsentry and de Crombrughe, 1990). In this study, the minimal promoter of the human pro- α 1(I) collagen gene is characterized. Transfection of human growth hormone under control of different promoters, revealed that even in the absence of an enhancer the pro- α 1(I) collagen promoter is a strong promoter. The results also suggest that the promoter sequences contain information for cell-specific modulation of expression of the gene.

Several DNA-protein binding assays performed with nuclear extracts from both expressing (fibroblasts) and non-expressing cells (lymphocytes) revealed differences in DNA-protein interactions with different cell extracts. These differences were observed at a pyrimidine stretch upstream of the CCAAT-boxes, around the duplicated and inverted CCAAT-boxes, and at a putative AP-2 site. Whereas DNAase I protection assays indicate that the pyrimidine stretch is bound by a factor present in fibroblasts, several DNAase I hypersensitive sites are induced here in the case of lymphocyte extracts. This DNAase I hypersensitivity suggests an alteration in DNA conformation in this region. The different but overlapping binding of nuclear factors to the CCAAT-boxes is suggestive of mutually exclusive binding of CCAAT-binding factors

as a possible mechanism for down-regulating the gene in non-expressing cells. In addition, we observed binding at a putative AP-2 site in the case of fibroblast extracts but not lymphocyte extracts. Similarities between fibroblast and lymphocyte extracts include binding to a putative Sp1 site and a factor of about 40 kd, named col-B, binding to a previously unidentified element upstream of the pyrimidine stretch.

OPSOMMING

INTERAKSIE VAN KERNFAKTORE MET DIE PROMOTOR VAN DIE MENS PRO- α 1(I) KOLLAGEENGEEN.

(Oorspronklike titel: Interaction of nuclear factors with the promoter of the human pro- α 1(I) collagen gene.)

Ten minste dertien kollageen tipes, elkeen met 'n karakteristieke weefselverspreiding, is tot dusver geïdentifiseer. Tipe I, die mees voorkomende kollageen, bestaan uit twee α 1 en een α 2 peptiedkettings. Daar is afdoende bewyse dat regulering van die tipe I gene onder andere op die vlak van transkripsie plaasvind. Reguleringsgebiede van die pro- α 1(I) kollageengeen kom beide opstroom van die inisiasie punt van transkripsie en in die eerste intron voor, terwyl 'n minimale promotor bevat is in die 250 bp direk opstroom van die transkripsionele beginpunt. Vorige studies het getoon dat die 250 bp promotor voldoende inligting bevat vir transkripsie van heteroloë gene in verskeie sisteme (Rossouw *et al*, 1987; Rippe *et al*, 1989; Liska *et al*, 1990 en Karsentry en de Crombrughe, 1990). Die doel van hierdie studie is die karakterisering van die minimale promotor van die mensgeen vir pro- α 1(I) kollageen. Transfeksie van die mens groeihoormoengeen onder beheer van verskillende promotors, het getoon dat die promotor van die pro- α 1(I) geen, selfs in die afwesigheid van enige aktiveringsvolgorde, 'n relatiewe sterk promotor is. Die resultate dui verder op die moontlikheid dat die promotor inligting bevat vir selspesifieke modulاسie van hierdie geen se uitdrukking.

Verskillende DNA-bindingsanalises, uitgevoer met kernekstrakte van selle waarin die geen uitgedruk (fibroblaste) of nie uitgedruk word nie (limfosiete), het selspesifieke verskille in DNA-protein interaksies aangedui. Hierdie verskille is waargeneem by 'n pirimidien-ryke volgorde opstroom van die CCAAT-homologië, romdom die gedupliseerde, omgekeerde CCAAT-homologië, en by 'n potensiële AP-2 herkenningsvolgorde. DNAase I afskermingsanalises het aangetoon dat 'n fibroblastfaktor aan die pirimidien-ryke element bind, terwyl verhoogde DNAase I

hipersensitiwiteit in die gebied waargeneem word in die geval van limfosietekstrakte. Hierdie DNAase I hipersensitiwiteit dui op moontlike veranderings in DNA-konformasie van hierdie element. Die verskillende maar oorvleuelende binding van kernfaktore aan die CCAAT-homologië dui op kompetisie vir binding in hierdie gebied as 'n moontlike meganisme vir inhibering van transkripsionele aktiwiteit in nie-uitdrukkende selle. Verder is ook binding by 'n potensiële AP-2 herkenningsvolgorde waargeneem in die geval van fibroblastekstrakte, maar nie limfosietekstrakte nie. Ooreenkomste tussen fibroblast- en limfosietkernekstrakte sluit binding aan 'n potensiële Sp1 volgorde in, asook 'n faktor met 'n molekulêre gewig van ongeveer 40 kd (col-B genoem), wat opstroom van die pirimidien element bind aan 'n volgorde-element wat nog nie voorheen geïdentifiseer is nie.

CHAPTER 1

INTRODUCTION

1.1 Regulation of gene expression

1.1.1 Promoters.

An important mode of regulation of gene expression is at the level of transcription. Mutational and deletional analyses have revealed two different *cis*-acting elements responsible for transcriptional regulation in eukaryotes (reviewed by Dynan and Tjian, 1985 and Maniatis *et al*, 1987). One element, termed promoter, is immediately upstream of the transcription initiation site, while enhancers can act from a distance of more than 1000 bp and can be located either upstream or downstream from the cap site.

Promoters of mammalian protein coding genes have been shown to consist of three basic elements: the initiator region (containing the start site of transcription), a TATA-box and upstream elements (reviewed by Dynan and Tjian, 1985; Maniatis *et al*, 1987). Most eukaryotic genes transcribed by polymerase II have a TATA-box about 30 bp upstream of the cap site. A few cases have been reported where mutation or deletion of the TATA-box significantly lowers promoter activity and/or causes transcription to start at heterologous sites (Smale and Baltimore, 1989 and references therein). The TATA-box provides a binding site for TFIID, one of the so-called general transcription factors which participate with Polymerase II in the formation of a pre-initiation complex (Nakajima *et al*, 1988).

The binding of TFIID to the TATA-box is the first step in the formation of a pre-initiation complex (Salzman and Weinmann, 1989). However, several polymerase II genes have no TATA-box. These TATA-less genes can be divided into two classes namely those with a GC-rich promoter (the housekeeping genes) and those with neither a TATA-box nor GC-rich promoter (including some of the *Drosophila* homeotic genes). It was shown that a wide variety of

sequences can functionally replace the TATA-box (Singer *et al*, 1990). In the absence of a defined TFIID binding site, as in the case of the housekeeping genes, scattered transcription start sites are observed. However, for TATA-less promoters TFIID is still involved in transcription initiation (reviewed by Sharp, 1992; White and Jackson, 1992). The TATA-box is therefore not absolutely required for TFIID-mediated transcription, but its presence increases the affinity of TFIID for the promoter and thus leads to higher levels of transcription (Hahn *et al*, 1989).

In the absence of a TATA box, the initiator region (Inr) containing the cap site is required for transcription (Smale and Baltimore, 1989). Mutating both the TATA-box and the Inr of the adenovirus major late promoter, completely abolished transcription, while mutating either one of these elements merely reduced the level of transcription (Concino *et al*, 1984). It has been shown that the Inr is weakly recognized by RNA polymerase II and association of the latter with the Inr is enhanced by the general transcription factors. Thus the Inr can recruit components of the pre-initiation complex to the transcription initiation site and plays a similar role as the TATA-box, though with lower efficiency than the latter (Carcamo *et al*, 1991).

Upstream elements, the third basic element of promoters, have been identified originally by mutational and deletional analysis. Many different elements have been identified up to date, some of which (like the CCAAT and GGGCGG-homologies) occur frequently in the promoters of different genes. These elements are binding sites for transcription factors which can either activate or repress transcription. The specific combination of upstream elements, as well as the presence or absence of active factors binding to these elements, determines whether or not a gene is transcribed in a certain cell (for review see Maniatis *et al*, 1987).

1.1.2 Enhancers

Enhancers were first identified in animal viruses as sequences activating transcription from a distance. Enhancers have three properties which distinguish them from promoters: (i) they activate transcription over large distances (more than 1000 bp); (ii) they are

orientation independent; and (iii) they can operate from either an upstream or downstream position (Serfling *et al*, 1985 and Müller *et al*, 1988).

Many different enhancers have already been identified, some of which (like the immunoglobulin enhancer) are tissue-specific. The study of a great variety of enhancers has revealed that enhancers and promoters can overlap both physically and functionally. Many transcription factors have binding sites in both enhancers and promoters (such as Sp1, Ap-1). In fact, multiple copies of binding sites for certain promoter factors, can act as an enhancer (Müller *et al*, 1988 and references therein). Analysis of the SV40 enhancer has shown that it consists of different functional elements (called enhansons), which can function on their own (when present in multiple copies) or in combination with each other. The specific combination of these enhansons determines the tissue-specificity of the enhancer as a whole (Schirm *et al*, 1987).

Two different models have been proposed to explain how enhancers can activate transcription at a distance namely the 'looping model' and the 'scanning model' (reviewed by Ptashne, 1986; Müller *et al*, 1988). According to the looping model, factors binding at the enhancer interact directly with the transcription machinery or factors binding at the promoter. The intervening sequences are thereby looped out bringing the promoter and enhancer into close proximity of one another. In the scanning model, the enhancer acts as an entry site for polymerase II from where it slides along the DNA until it reaches the promoter from where it starts transcribing.

1.1.3 Transcription factors.

Most, if not all promoter and enhancer elements involved in transcriptional regulation are binding sites for transcription factors. Many different transcription factors have been identified up to date, some of which are ubiquitous factors (such as Sp1 and Oct-1), while others are present only in certain tissues (such as AP-2 and Oct-2). Transcription of a given gene is not only dependent on the presence of the factors regulating it but also whether these factors are in an active conformation.

Certain sequence elements occur in many eukaryotic promoters, like the Sp1 binding site and the CCAAT-box. The consensus recognition sequence of Sp1 is GGGCGG but the same factor binds to other similar sequences. Since Sp1 has different binding affinities for different sites, this represents another level of transcriptional control (Kadonaga *et al*, 1986). Interestingly, the CCAAT-box which is present in many promoters, is recognized by a variety of factors (Dorn *et al*, 1987). Which factor actually binds to the CCAAT-box is determined by the flanking sequences. In addition, the relative abundance of the different CCAAT-binding factors, and their respective binding affinity to a specific CCAAT-box would determine which CCAAT factor would bind there preferentially. Different CCAAT factors have different characteristics, all of which would influence the transcription rate of a gene with a CCAAT-box in the promoter (for review, see Johnson and McKnight, 1989).

The isolation and subsequent cloning of different transcription factors revealed that they often consist of distinct domains responsible for DNA-binding, dimerization, activation and ligand binding (for review, see Mitchell and Tjian, 1989). Four different types of activation domains have been identified namely amphiphatic α -helices with a net negative charge, serine-threonine rich, glutamine rich and proline rich domains. These activation domains differ in their ability to activate transcription from remote (enhancer) positions. Acidic and glutamine domains activate strongly from both promoter and enhancer positions, serine-threonine and proline domains have low enhancer activity, while glutamine domains can activate only from a promoter position (Seipel *et al*, 1992). However, the activation domains of several transcription factors do not fit in any one of these classes.

The mere presence of transcription factors for which there are binding sites in the regulatory regions of a gene is sometimes not enough to ensure transcriptional activation. Where binding sites are overlapping and mutually exclusive, the respective factors compete for binding and transcriptional activity is dependent on the ratios of the factors present as well as their respective binding affinities (for examples see Akerblom *et al*, 1988; Barberis *et al*, 1987). Some

factors are present in the cell but are in an inactive conformation. This can be due to the phosphorylation status of the factor and/or the formation of a complex with another protein (Johnson and McKnight, 1989). For example, the transcription factor NF- κ B is present in a variety of cell types, but is associated with an inhibitor I κ B in cells not requiring transcriptional activation through NF- κ B (reviewed by Blank *et al*, 1992). Another mechanism of modulating the transcriptional activation potential of transcription factors is through the formation of heterodimers. In the case of fos and jun, homodimers as well as heterodimers can form, and each combination differs in binding affinity to AP-1 sites (Halazonetis *et al*, 1988; Busch and Sassone-Corsi, 1990).

The steroid receptors have an additional domain for ligand binding. These factors do not activate transcription unless they are bound by their respective ligands. The mechanisms by which the various hormone receptors are induced differ from each other. In the case of the glucocorticoid receptor, the receptor is complexed by heat shock protein 90 (hsp-90) and located in the cytoplasm. Upon binding of the hormone, the receptor dissociates from hsp-90 and is transported into the nucleus where it binds to the glucocorticoid responsive elements on the DNA (Johnson and McKnight, 1989). It has been shown that in the absence of the hormone binding domain, hsp-90 does not bind to the receptor, resulting in a constitutively active receptor (Godowski *et al*, 1987).

Different mechanisms for repression of transcription have been proposed (reviewed by Levine and Manley, 1989). In prokaryotes the predominant mechanism observed is through steric hindrance of the binding of transcription factors or even polymerase itself to the DNA. This has also been observed in certain cases in eukaryotes (Barberis *et al*, 1987, Akerblom *et al*, 1988). Heterodimerization is another way of preventing transcription factors from activating transcription. In the case of the helix-loop-helix (HLH) protein MyoD, another HLH protein (Id) can heterodimerize with MyoD. Id lacks the basic region necessary for DNA binding and upon dimerizing with MyoD, prevents the latter from binding to DNA and activating transcription (Benezra *et al*, 1990).

Other models which have been proposed are the quenching and direct model. According to the quenching model the repressor and activator both bind to the DNA, but the repressor somehow prevents functioning of the activator. There are indications that the α -specific genes in yeast α -mating type is repressed in this way (reviewed by Herskowitz, 1989). In the direct model the repressor interacts with some component of the pre-initiation complex or with polymerase itself. One such negatively acting factor has been identified as being part of a TFIID associated cofactor, USA (Meisterernst *et al.*, 1991).

How do transcription factors work? Seven general transcription factors have been identified that are required by RNA polymerase II for transcription initiation, namely TFIID, IIA, IIB, IIF, IIH and IIJ (Saltzman and Weinmann, 1989; Zawel and Reinberg, 1993). The activating regions of transcription factors can interact with any one of the components necessary for the formation of the pre-initiation complex or even polymerase II itself.

Several lines of evidence indicates that TFIID, which is necessary during the first stages of pre-initiation complex formation could be the target of transcription factors. Horikoshi *et al.* (1988) demonstrated that the transcriptional activator ATF interacts cooperatively with TFIID and that this interaction aids the recognition of the promoter by polymerase II and the other general initiation factors. In another study Stinger *et al.* (1990) showed that a column of the negatively charged activation domain of VP16 can selectively bind TFIID. In similar experiments it could be shown that the acidic domain of VP16 can also interact with TFIIB (Lin and Green, 1991).

Cloning the TATA binding protein (TBP) present in the TFIID fraction, revealed that this factor alone cannot account for transcriptional activation. This suggested the presence of co-activators or adaptors mediating the effect of transcription factors (Lewin, 1990 and references therein). In subsequent experiments it could be demonstrated that TFIID actually consists of a complex of TBP and at least six tightly bound TBP associated factors (TAFs). TBP alone is sufficient for basal transcription, but the TAFs are

necessary for activation of transcription by transcription factors (Dynlacht *et al.*, 1991). The discovery that there exists different sets of TAFs suggests that each transcription factor may interact with (a) different TAF(s) (reviewed by Sharp, 1992).

1.2 Collagen

1.2.1 Collagen genes

The collagens are a family of proteins which all have a triple-helical stretch consisting of repeating tripeptides with the sequence Gly-X-Y where X is often proline and Y hydroxyproline. There are at least 13 different types of collagen and they can be divided into two classes namely the non-fibrillar collagens and the fibrillar collagens. The latter consist of largely uninterrupted triple helices with a characteristic cross-striated pattern. Each collagen type has a characteristic tissue distribution and plays a different role in embryogenesis, development and the adult animal (reviewed by Vuorio and de Crombrughe, 1990).

The three major fibrillar collagens, types I, II and III have similar structures on both the gene and protein level. Type I collagen consists of two $\alpha 1$ and one $\alpha 2$ chain, while type II and type III are homotrimers. Comparing the exon - intron structure of the four genes coding for these proteins, reveals a basic structure consisting of 52 (± 1) exons. A large part of the genes consist of introns - 70% and 90% for the $\alpha 1(I)$ and $\alpha 2(I)$ genes respectively. (Chu *et al.*, 1984). The proteins are all synthesized as pro-collagens with globular N- and C-propeptides which border the central triple-helical domain. After association of the three strands forming one collagen fibre, specific N- and C-proteases remove the propeptides. Except for a few minor variations, the triple-helical domain of all four genes is encoded by 42 exons each starting with a complete codon for glycine and being either multiples of 54 bp or combinations of 54 and 45 bp. In addition, the N- and C-terminal joining exons code for the first and last few triplets of the triple helix (reviewed by Vuorio and de Crombrughe, 1990).

1.2.2 Regulation of collagen synthesis

Of the different types of collagen, most is known about types I - IV and knowledge of the regulation of collagen genes is largely limited to the genes of these types. However, here I will concentrate on the three major fibrillar collagens, types I, II and III.

Each of the three major fibrillar collagens has a characteristic tissue distribution. Type III and type I collagen are mostly associated with one another with type III being present in lesser amounts. Type I is the major protein constituent of tissues like skin, bones and tendons, while the aorta and other large blood vessels are enriched for type III collagen. On the other hand, type II collagen is the characteristic articular cartilage collagen, and occurs also in the eye and transiently during embryogenesis (Sandberg and Vuorio, 1987).

There is evidence for different levels of regulation of collagen synthesis (reviewed by Bornstein and Sage, 1989; Vuorio and de Crombrughe, 1990). That the regulation of transcription is an important mode of control of collagen synthesis has been suggested by experiments showing that the 2:1 ratio of $\alpha 1(I)$ to $\alpha 2(I)$ collagen chains is reflected by the same ratio of the respective mRNAs (de Wet *et al*, 1983). Different studies have furthermore shown that there is a good correlation between mRNA levels and changes in collagen synthesis.

The identification of polymorphic RNA transcripts for all the genes of the three major fibrillar collagens, has led to the suggestion that collagen synthesis may also be regulated at the mRNA level. This has been strengthened by the observation that TGF- β induced increase in $\alpha 1(I)$ collagen is associated with a change in the ratio of different mRNA forms of the $\alpha 1(I)$ gene (Bornstein and Sage, 1989 and references therein).

There has been some evidence for translational control of type I collagen such as occasional discrepancies between mRNA levels and collagen synthesis rates, as well as the finding that fragments of the amino-terminal globular domain of the pro- $\alpha 1$ subunit of type I collagen can inhibit the translation of collagen mRNA. Studies on

chondrocytes and their dedifferentiation suggested translational control of both type I and type II collagen. In this case, collagen synthesis doesn't correspond to the respective mRNA levels (Focht and Adams, 1984; Finer *et al*, 1985; Bennett and Adams, 1987).

The presence of a conserved inverted repeat around the translation initiation codon (AUG) of the type I and III genes as well as the invariant presence of two AUG codons followed by in-frame stop codons preceding the initiator AUG, are further suggestions of translational control. However, studies on mutants with disruptions of these conserved sequences, did not yield any information as to their function (Bornstein *et al*, 1988a).

1.2.3 Transcriptional regulation of the pro- α 1(I) collagen gene

The first clues to regulation of collagen gene transcription came from studies by Jaenisch and his colleagues of germ-line integration of Moloney murine leukemia virus in mice (Harbers *et al*, 1984). They found a recessive lethal mutant, Mov-13, in which transcription of the pro- α 1(I) collagen gene was prevented. Analysis of the site of integration revealed that the virus was inserted into the first intron of the α 1(I) gene. In another study, Barsh *et al* (1984) mapped DNAase I hypersensitive sites in the α 1(I) collagen gene. They found such sites near both the 3' and 5' ends of the gene, the latter being in the first intron.

The importance of the α 1(I) gene's first intron in regulation of its transcription, was confirmed by the identification of enhancing elements in this intron (Rossouw *et al*, 1987). Shortly after this report, enhancers were identified in the first introns of the α 2(I) and the α 1(II) genes (Rossi and deCrombrughe, 1987; Horton *et al*, 1987). Further studies on the human as well as mouse α 1(I) gene, revealed the presence of both positive and negative elements in the first intron (Bornstein *et al*, 1988b; Rippe *et al*, 1989).

Functional studies of the 5' untranscribed regions of both the α 1(I) and α 2(I) genes, have revealed that for both genes, 200 to 300 bp upstream of the cap site is sufficient to act as a promoter (Schmidt *et al*, 1986; Rossouw *et al*, 1987 and Rippe *et al*, 1989). Both

Rossouw *et al* (1987) and Rippe *et al* (1989) reported that in the case of the $\alpha 1(I)$ gene, addition of sequences more upstream of the minimal promoter results in changes in expression of reporter genes. This indicates that beside the regulatory elements in the first intron, this gene may also be regulated by remote upstream elements. In transgenic mice studies it was demonstrated that for the $\alpha 1(I)$ and $\alpha 2(I)$ collagen genes, about 2 kb of upstream sequences are sufficient to direct tissue specific expression of these genes in mice (Khillan *et al*, 1986; Wu *et al*, 1990; Slack *et al*, 1991; Pavlin *et al*, 1992).

All the genes of type I to III collagens have a TATA-box in their promoters, but only the type I genes have CCAAT-boxes.

heterodimeric factor binding to the CCAAT-box of the mouse $\alpha 2(I)$ gene has been identified and purified (Hatamochi *et al*, 1988). This factor was subsequently shown to be able to stimulate transcription from both the $\alpha 2(I)$ and $\alpha 1(I)$ promoters when added to fibroblast nuclear extract depleted of this factor (Maity *et al*, 1988). An NF-1 binding site has also been identified in the $\alpha 2(I)$ promoter, which is able to mediate stimulation of transcription by TGF- β (Rossi *et al*, 1988).

A few studies on sequence elements of the mouse $\alpha 1(I)$ gene have been published (Brenner *et al*, 1989; Karsentry and de Crombrughe, 1990). This promoter has, except for the TATA- and CCAAT-boxes also putative Sp1-sites. In the enhancer are additional Sp1 binding sites as well as an enhancer core element. One further element possibly involved in the regulation of this gene is a pyrimidine-rich region present in both the promoter and intron (Rossouw *et al*, 1987; Bornstein *et al*, 1987).

Brenner *et al* (1989) have studied the mouse $\alpha 1(I)$ promoter and have shown with DNAase I footprinting analysis that nuclear factors from fibroblasts bind to four regions in this promoter. They could also demonstrate that the most proximal footprinted sequence, containing a CCAAT-box, can activate transcription from a heterologous promoter. Karsentry and de Crombrughe (1990) postulated the binding of the $\alpha 2(I)$ CCAAT-binding factor as well as

two negatively acting factors to the $\alpha 1(I)$ promoter. In contrast to this data, Nehls *et al* (1991) demonstrated that NF-1 and not the $\alpha 2(I)$ factor binds to the CCAAT elements of the mouse $\alpha 1(I)$ promoter. DNAase I footprinting analyses were only made of a limited part of the first intron and revealed protection over at least one of the Sp1 binding sites as well as the enhancer core element (Bornstein *et al*, 1987). Experiments with purified Sp1 confirmed that this factors recognizes and binds to the Sp1 site in the first intron (Liska *et al*, 1992).

1.3 Aim of the study and approach

Initial studies revealed that 250 bp upstream of the transcription initiation site of both the human and mouse pro- $\alpha 1(I)$ collagen gene is sufficient to direct transcription of a reporter gene in different *in vivo* systems (Rossouw *et al*, 1987; Rippe *et al*, 1989). At the time of this study, very little was known about promoter elements involved in regulating expression of this gene. The aim of this study was the characterization of the 250 bp promoter of this gene. The intron of the pro- $\alpha 1(I)$ gene contains both negative and positive regulatory elements, and addition of upstream sequences to the minimal promoter modifies the rate of transcription. The question arose therefore how strong the 250 bp promoter is on its own, i.e. in the absence of intronic and far upstream sequences. In particular, we wanted to compare the strength of the pro- $\alpha 1(I)$ promoter with that of other known and well-characterized promoters.

Another aim was to identify sequence elements involved in transcriptional regulation of the gene. Two reports have been published on DNA-protein interactions of the murine pro- $\alpha 1(I)$ promoter with fibroblast nuclear factors (Brenner *et al*, 1989; Karsentry and de Crombrughe, 1990). In this study, similar experiments were done with the human promoter. In addition to comparing binding of fibroblast nuclear factors to the human promoter with those published for the murine promoter, we also tested extracts from cells not expressing type I collagen. This was done in order to identify promoter elements which might be involved in tissue-specific regulation of the gene.

There are two different approaches for characterizing regulatory regions of genes. The first is through functional assays in which the effect of the regulatory regions on the expression of a reporter gene is tested. This can be done either *in vitro*, using nuclear extracts from cells to provide the necessary transcription factors, or *in vivo* by transfection or micro-injection into cells. Mutational and deletional studies can then be done in the same system to refine the picture of which sequence elements are necessary for transcription and/or tissue-specific expression of genes. Transient transfections utilizing tissue culture cells has already been used successfully for determining regulatory regions in a variety of different genes, notably the globin and immunoglobulin genes (Müller *et al.*, 1988 and references therein). However, in order to obtain a complete picture of tissue-specific regulation of gene expression, the regulatory regions should be tested in the whole organism. This has become feasible through the techniques available for generating transgenic mice. In the past few years regulatory regions of several genes have been thus characterized (reviewed by Rusconi, 1991).

Different reporter genes can be used in such assays, some of which yield an easy measurable protein product. Commonly used reporter genes of the latter type, result in a product which can be detected by either an enzymatic assay, like chloramphenicol acetyl transferase (Gorman *et al.*, 1982) or by an immunological assay, like growth hormone (Selden *et al.*, 1986) and SV40 large T antigen (Benoist and Chambon, 1980). Another way of detecting transcription of a reporter gene is by RNAase protection or S1 analyses of the mRNA of the transfected cells with a probe hybridizing specifically to the mRNA transcribed from the reporter gene (Westin *et al.*, 1987).

The second approach is to map sequence elements involved in DNA-protein interactions. Different techniques are available for such studies:

(i) Electrophoretic mobility shift assays or bandshifts: With this assay one can detect whether an extract contains factors which bind to a radio-actively labeled fragment or oligonucleotide.

(ii) DNAase I footprinting analyses: This provides a means of mapping the sequences to which factors bind by comparing the DNAase I digestion pattern in the presence and absence of nuclear extract. Where factors are bound to the DNA, the nuclease is prevented from digesting the DNA and a 'footprint' appears when assayed by polyacrylamide gel electrophoresis (Galas and Schmitz, 1978).

(iii) Southwestern blotting analyses: In this assay nuclear extracts are separated on a denaturing protein gel and, after transfer to nitrocellulose, hybridized to a labeled fragment or oligonucleotide. This allows the determination of the molecular weights of factors binding to the fragment or oligonucleotide (Hübscher *et al*, 1987). This assay is limited however, in that only factors binding as monomers can be detected.

(iv) Methylation interference assay: Partially methylated DNA is bound to nuclear factors, the bound and unbound DNA is separated electrophoretically and isolated from the gel. After cleaving the DNA at the methylated residues, the DNA is analyzed by polyacrylamide gel electrophoresis. Comparing the cleavage pattern of the bound and unbound probe gives an indication of which residues are necessary for binding of factors (Siebenlist and Gilbert, 1980).

(v) *In vivo* footprinting: This assay is similar to DNAase I protection in that the areas bound by proteins are detected by virtue of their insensitivity to modification. In *in vivo* footprinting whole cells are treated with dimethyl sulfate (DMS) which methylates DNA wherever it is not protected by binding of proteins. The DNA is then extracted from the cells and subjected to chemical cleavage which cleaves DNA at all methylated residues. DNA from DMS treated and untreated cells are analyzed by gel electrophoresis and as in the case of DNAase I footprints, those areas bound by proteins show absence of cleavage. The advantage of this method is that DNA-interactions can be detected *in vivo* (Saluz and Jost, 1989)

Both approaches have been followed in this study. To determine to what extent the 250 bp promoter of the human pro- α 1(I) collagen gene can activate transcription in the absence of the gene's enhancer, transient transfections were performed with the human growth hormone gene under the control of different promoters. In

this way the activity of the pro- α 1(I) promoter could be compared to that of other promoters. The human growth hormone gene was chosen as a reporter gene because the product formed can be assayed in a simple, non-destructive way by means of a radio immuno assay. Since the culture medium is analyzed for the presence of growth hormone, the cells are still available for eventual RNA extraction and analysis.

To analyze DNA-protein interactions of the promoter, three different assays were employed, namely bandshifts, DNAase I footprinting assays and southwestern blotting analyses. In order to determine whether these interactions are tissue-specific, all assays were performed with nuclear extracts from different cell lines, i.e. both collagen expressing and non-expressing cells.

A fragment containing the promoter was used in bandshift assays to determine whether there are differences between factors from expressing and non-expressing cells binding to the promoter. The same fragment was subsequently used in DNAase I footprints for nailing down sequences where nuclear factors bind. Comparing the footprints done with extracts from fibroblasts (expressing type I collagen) and lymphocytes (non-expressing) gave an indication of which sequence elements might be involved in tissue specific regulation of the gene. Oligonucleotides were synthesized of the elements bound by nuclear factors as identified in the DNAase I footprinting assays. These oligonucleotides were tested in bandshift and southwestern blotting assays.

CHAPTER 2

FUNCTIONAL ANALYSIS OF THE PRO- α 1(I) COLLAGEN PROMOTER.

2.1. Introduction

The α 1(I) collagen gene has been reported to have regulatory elements in its first intron (Rossouw *et al*, 1987; Bornstein *et al*, 1988b; Rippe *et al*, 1989). This is also the case for the α 2(I) (Rossi & de Crombrughe, 1987) as well as the type II collagen genes (Horton *et al*, 1987). The question arose how dependent the pro- α 1(I) promoter is on the intronic enhancer elements or, in other words, what is the relative strength of the promoter on its own.

To address this question, the pro- α 1(I) minimal promoter as defined by Rossouw *et al* (1987) (250 bp upstream of the cap site), was compared to other previously characterized promoters. The promoters tested were fused to the human growth hormone in an expression vector and transiently transfected into different cell lines. Expression of the reporter gene was detected by testing the growth medium of the cells for the presence of human growth hormone. In this way we could obtain an indication of the relative strength of the pro- α 1(I) promoter.

2.2. Construction of plasmids

The construction and isolation of plasmids were performed as described by Vergeer (1988). All the plasmids that were used were derived from the plasmids included in the Allegro™ HGH transient gene expression assay system (Selden *et al*, 1986).

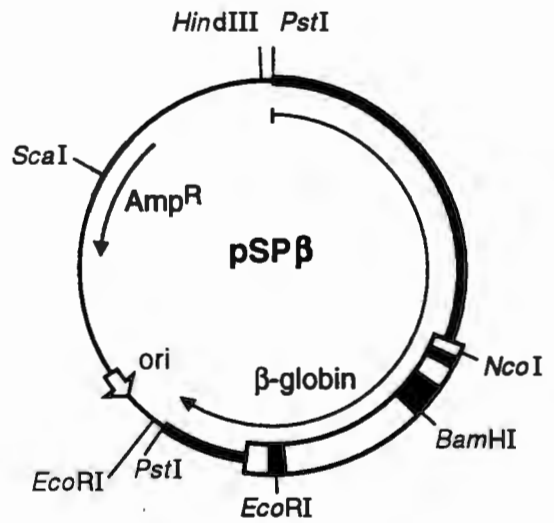
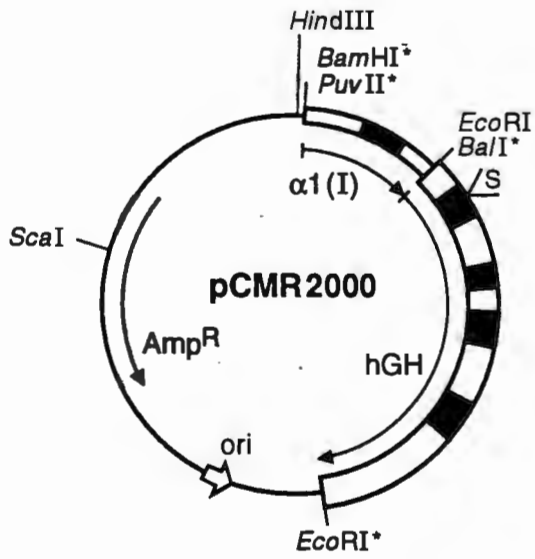
The human growth hormone sequences of the Allegro plasmid pOGH (fig. 2.4A), were cloned as a Bam HI - Eco RI fragment into Bam HI, Eco RI digested pSP64. In this plasmid (named pSP0GH) the human growth hormone's first exon was replaced by the first exon and promoter of the pro- α 1(I) collagen gene. Two plasmids were constructed namely pCMR2000 (fig. 2.1) and pCMR1000 (fig. 2.2)

with a short (250 bp) and long (2400 bp) promoter, respectively. The collagen and growth hormone sequences were fused in the first intron of both genes by ligating a filled in Eco RI site of pro- α 1(I) collagen to a Bal I site of the growth hormone. This ligation resulted in the regeneration of the Eco RI site and thus leaves the possibility of inserting enhancer fragments into the first intron.

Both human growth hormone and pro- α 1(I) collagen are normally secreted and therefore have signal peptides. The growth hormone peptidase site is located in the second exon and the α 1(I) collagen sequences of the fusion protein are removed during the process of secretion. Wild type human growth hormone would therefore be secreted into the medium allowing detection by anti-growth hormone antibodies.

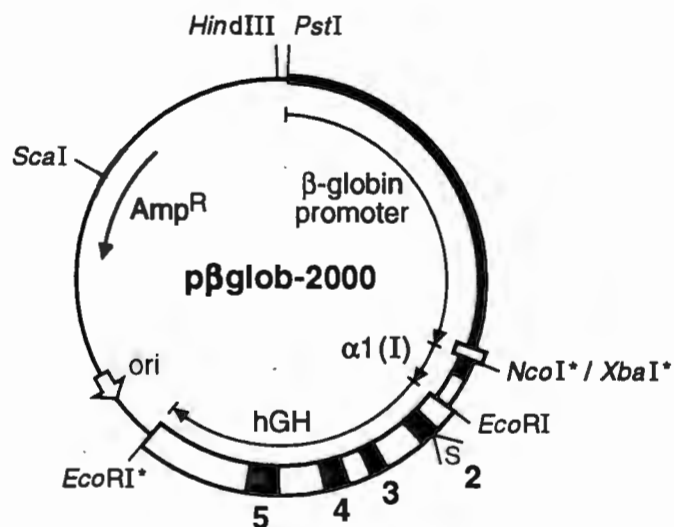
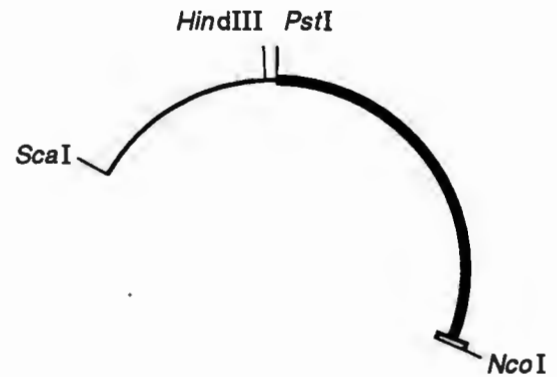
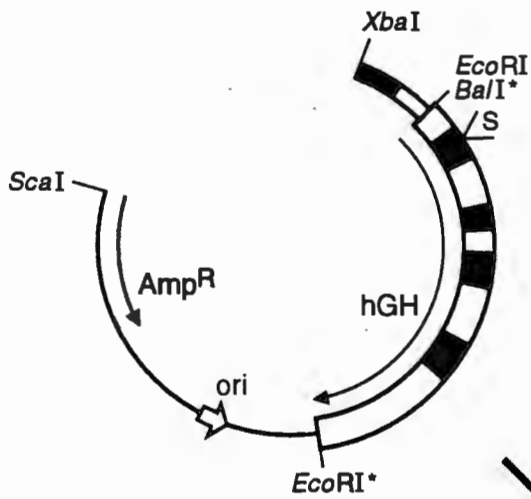
Figure 2.1: Construction of p β glob2000 - an example of cloning with pCMR2000.

The Nco I site of pSP β was filled in and the Nco I (blunt ended) - Sca I fragment was ligated to the Xba I (filled in), Sca I digested pCMR2000. Thin lines represent plasmid sequences, thick lines untranscribed and boxes transcribed regions of genes. The open boxes are non-coding regions and filled boxes exons, with pro- α 1(I) collagen sequences (α 1(I)) as smaller boxes and the growth hormone (hGH) and β -globin gene sequences as big boxes. The direction of transcription of the respective genes, as well as the ampicillin resistance gene (Amp^R) are indicated with arrows and the origin of replication with a thick arrow (ori). Restriction enzyme sites that have been destroyed, are marked with an asterisk. S indicates the signal peptidase site and numbers refer to the exons of the human growth hormone.



XbaI
Klenow
ScaI
CIP

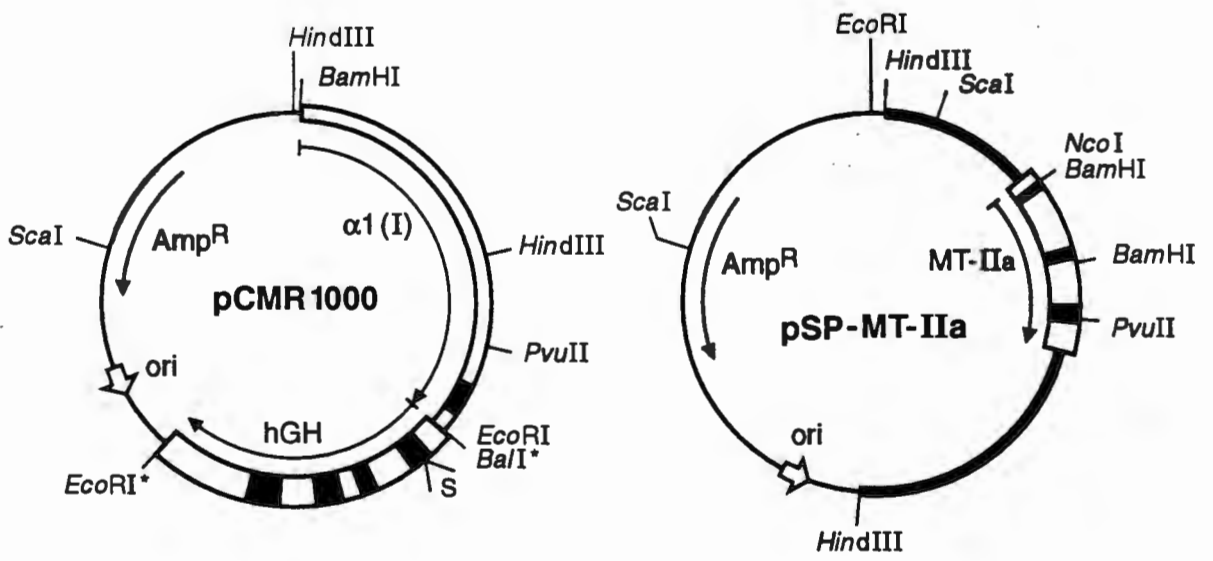
NcoI
Klenow
ScaI



Plasmids with different promoters were assembled by replacing the pro- α 1(I) collagen promoter of either pCMR2000 or pCMR1000 with other promoters. Examples of the construction using pCMR2000 and pCMR1000 are shown in figs. 2.1 and 2.2, respectively. Promoters of genes previously cloned into pSP-vectors were isolated by digestion of the plasmids with Sca I, which is situated in the β -lactamase gene, and another enzyme with a site between the transcription initiation site and the ATG codon of the specific gene. These fragments were ligated to a Sca I - blunt ended Xba I fragment of pCMR2000. The Xba I site is located between the transcription initiation site and the ATG codon of the pro- α 1(I) collagen gene. Translation of the mRNAs transcribed from these plasmids start, therefore, at the α 1(I) collagen initiation codon. As an example of this strategy, the cloning of p β glob2000 is shown in fig. 2.1. The origin of the promoter fragments used for constructing pSV2000 and pTK2000, which were constructed in a similar way, are shown in figs 2.3A and B respectively.

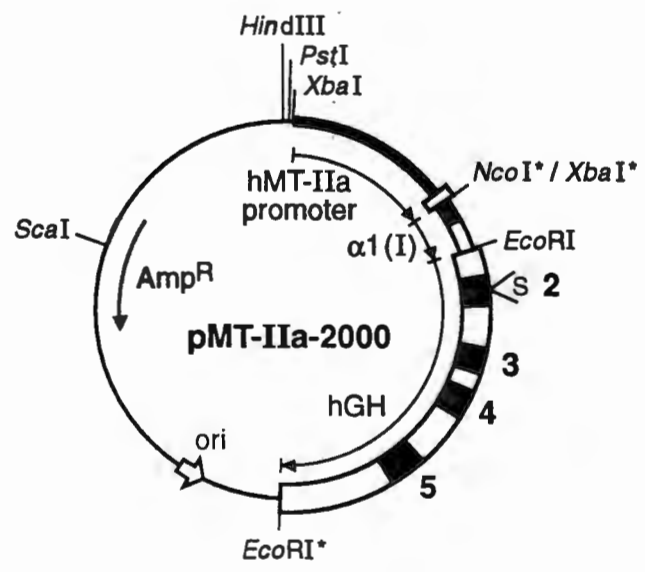
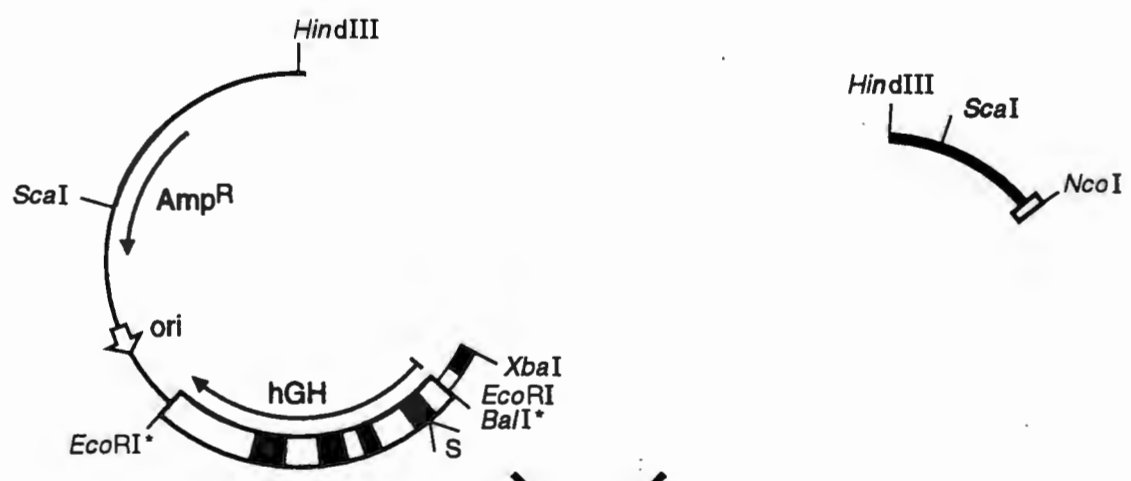
Figure 2.2: Construction of pMT-IIa-2000 - an example of cloning with pCMR1000.

The Nco I - Hind III fragment of pSP-MT-IIa was blunt ended with Klenow polymerase and cloned into pCMR1000 digested with Xba I and blunt ended. MT-IIa is the human metallothionein gene. All other symbols and abbreviations used are as described in fig. 2.1.



XbaI
Klenow
CIP

NcoI
HindIII
Klenow



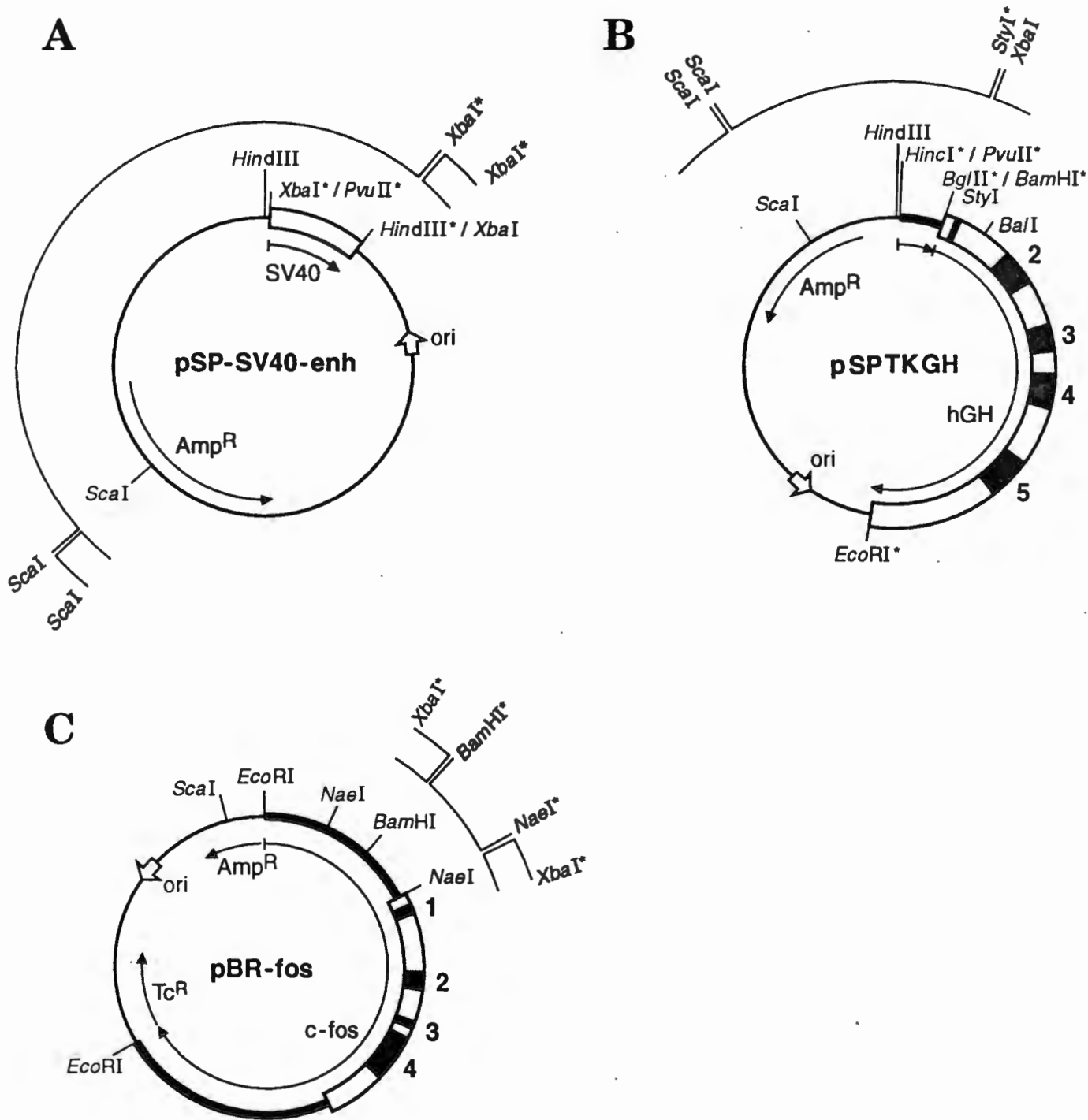


Figure 2.3: Origin of promoter regions used in cloning.

The actual fragments that have been used in the cloning are given above the plasmids, together with the restriction sites used.

A: The Sca I - Xba I blunt ended fragment of pSP-SV40-enh was ligated to the pCMR2000 fragment (fig. 2.1), resulting in pSV2000.

B: Of pSPTKGH the Sca I - blunt ended Sty I fragment was ligated to the pCMR2000 fragment (fig. 2.1), forming thus pTK2000.

C: The Bam HI - Nae I fragment of pBR-fos was blunt ended and cloned into Xba I digested pCMR1000 (fig.2.2) for pfos2000.

Tc^R indicates the tetracycline resistance gene and SV40 the SV40 enhancer-promoter, with all further symbols and abbreviations as described in fig.2.1.

Another strategy was employed for the construction of the plasmids with the promoters of the human metallothionein gene and the c-fos gene (pMT-IIa-2000 and pfos2000). These two promoters were isolated without sequences from pSP and inserted into pCMR1000 (fig. 2.2) digested with Xba I. The presence of an Xba I site in the pSP64 polylinker enabled the removal of the entire $\alpha 1(I)$ promoter by digestion with Xba I and allowed the ligation of other promoters to the same Xba I site as was used in the cloning with pCMR2000. In fig. 2.2 the construction of pMT-IIa-2000 is given as an example of this strategy, while the promoter fragment used for cloning pfos2000 is given in fig. 2.3C.

The following plasmids were constructed :

pTK2000:

A Hind III - Eco RI fragment containing the Herpes Simplex virus (HSV) thymidine kinase promoter and the human growth hormone gene from the plasmid pTKGH of the Allegro system (Selden *et al*, 1986) was cloned into the Hind III - Eco RI sites of pSP64. From this plasmid, named pSPTKGH (fig. 2.3B), a 2 kb Sca I - Sty I fragment was ligated to the 4 kb Xba I - Sca I fragment of pCMR2000. The Sty I site in pSPTKGH is located in the 5' untranslated region of the growth hormone (DeNoto *et al*, 1981).

p β glob2000:

The human β -globin gene, together with 2.3 kb upstream and 0.6 kb downstream sequences (Lawn *et al*, 1980), was isolated from plasmid β pstd as a 4.4 kb Pst I fragment and cloned into the Pst I site of pSP64. From the resulting plasmid, pSP β (fig. 2.1), the β -globin promoter was isolated as a Sca I - Nco I fragment and ligated to the 4 kb Xba I - Sca I fragment of pCMR2000 (fig. 2.1).

pSV2000:

The Pvu II - Hind III fragment of pSV2Neo (Southern and Berg, 1982) containing the enhancer and early promoter of the SV40 virus, was cloned into the Xba I site of pSP64 to create the plasmid pSP-SV40-enh (fig. 2.3A). Blunt end ligation of the Hind III site to the Xba I site resulted in the regeneration of the Xba I site and this

site was used to obtain a 1.6 kb Sca I - Xba I fragment which was ligated to the 4 kb pCMR2000 fragment.

pMT-IIa-2000:

For the construction of pSP-MT-IIa, a 3 kb Hind III fragment from pMTII-BPV (Karin *et al*, 1983) was ligated to Hind III digested pSP65. A 0.8 kb Hind III - Nco I fragment containing the human metallothionein IIa promoter was obtained from pSP-MT-IIa and ligated to Xba I digested pCMR1000 (fig. 2.2).

pfos2000:

The promoter of the human fos gene was obtained from the plasmid pBR-fos (fig. 2.3C) which consists of the fos gene with 2.3 kb and 6.5 kb of 5' and 3' flanking sequences, respectively (Sassone-Corsi *et al*, 1988, van Straaten *et al*, 1983). The 0.75 kb Bam HI - Nae I fragment containing the promoter was ligated to the 5.3 kb Xba I fragment of pCMR1000.

The plasmids p0GH and pXGH5 (fig. 2.4A and B) of the Allegro system were employed as controls. In pXGH5 the human growth hormone gene is under control of a mouse metallothionein-I promoter which results in relatively high expression of the growth hormone gene after transfection (Selden *et al*, 1986). On the other hand, p0GH where the growth hormone lacks any known eukaryotic promoter, served as a negative control. Selden *et al* (1986) reported that the amount of growth hormone secreted by cells transfected with this plasmid is about 0.5% of the amount secreted by cells transfected by pXGH5.

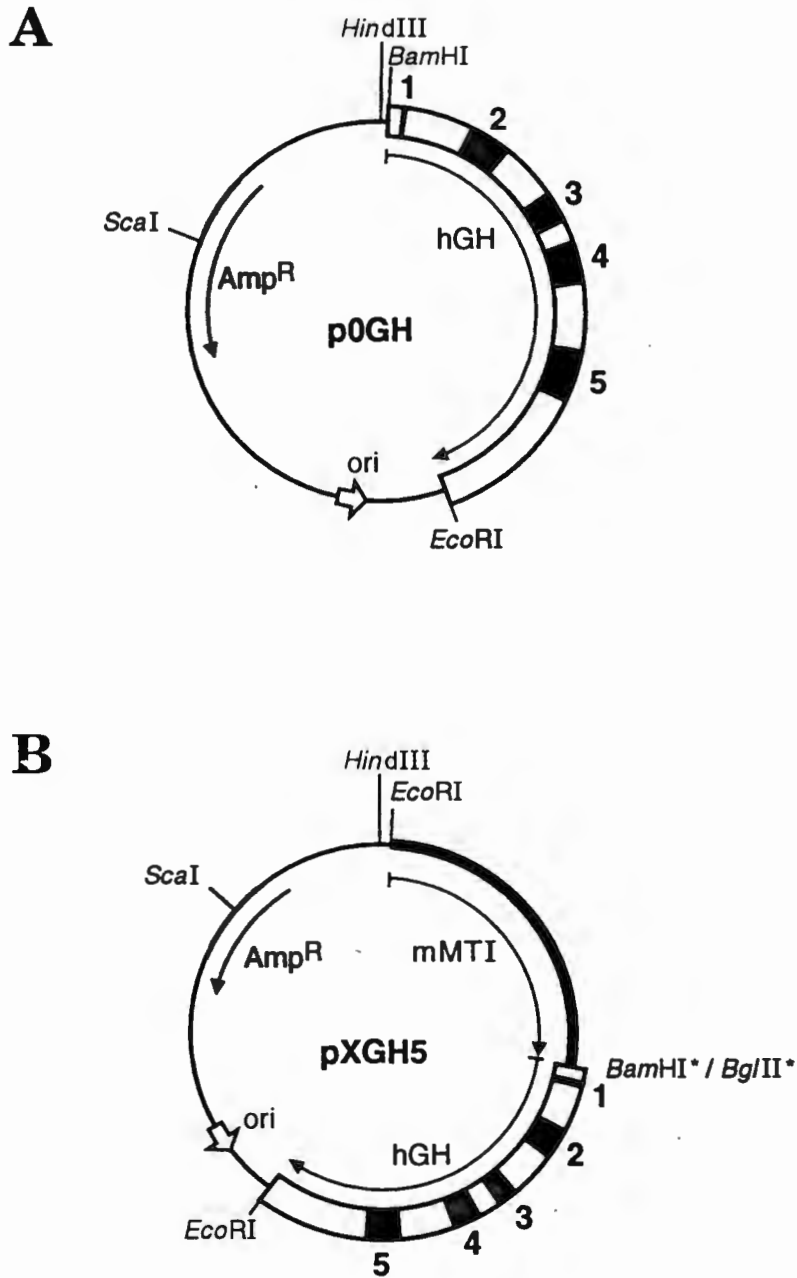


Figure 2.4: The Allegro plasmids

The two Allegro plasmids (described by Selden *et al*, 1986) were used as negative (p0GH) and positive (pXGH5) controls.

A: p0GH5: The human growth hormone with the promoter deleted, i.e. a promoterless plasmid.

B: pXGH5: The human growth hormone is driven by the mouse MT-I metallothionein promoter.

The symbols and abbreviations are as described in fig.2.1.

2.3. Transfection of cells.

2.3.1 Methods

Transfection was performed by the calcium phosphate method, essentially as described by Vergeer (1988). Cos-7 and C127 cells were cultured in DMEM with 10% fetal calf serum and NIH-3T3 cells in DMEM with 5% fetal calf serum. About 18 hours before transfection the cells were seeded so that the the cells were 60 - 80% confluent at the time of transfection.

For transfection of cells in a 75 cm² culture flask, the CaPO₄-DNA mix was prepared as follows: DNA was suspended in 1 ml HBS medium consisting of 0.14 M NaCl, 25 mM HEPES and 0.75 mM Na₂HPO₄ with a pH of exactly 7.1. The amount of DNA used for each reaction was determined using the following formula :
$$\mu\text{g DNA} = 10 \times (\text{size of plasmid} / \text{size of pBR})$$
. Of a 2 M CaCl₂ solution, 62 μ l was added slowly and after gentle mixing the samples were left at room temperature for 30 minutes.

After the CaPO₄-DNA precipitates had been added to the cells, they were incubated for 4 hours at 37°C. Thereafter, the medium with the precipitate was removed and the cells were rinsed with DMEM to remove any remainder of the precipitate. The cells were subjected to a glycerol shock by adding 3 ml of 15% (w/v) glycerol in HBS to the cells for 3 min, after which the cells were once again rinsed with DMEM. Total medium was added to the cells and they were incubated for another 4 days before medium was extracted for the growth hormone assay.

Transient expression of the transfected plasmids was detected by determining the amount of human growth hormone present in the medium after 4 days. Sleicher and Schuell's slot-blot machine was used to transfer 100 μ l of medium to a 0.1 μ m nitrocellulose membrane. A flow chart of the radio immuno assay, which was employed for the detection of growth hormone, is given in fig. 2.5 A. The nitrocellulose membrane was left to air dry before non-specific sites were blocked by incubating the membrane at room temperature for three hours in PBS (0.01 M Na₂HPO₄/NaH₂PO₄,

0.15 M NaCl, pH 7.2) with 0.5% Tween-20 and 6% Blotto non-fat dry milk. After the membrane was washed 3 times for 10 min in a wash buffer consisting of PBS and 0.1% Tween-20, it was hybridized for 1 hour with monoclonal anti-hGH IgG (Amersham, MCA 392, 1 mg/ml diluted 1:600 in the wash buffer). The membrane was washed again 3 x 10 min and then hybridized for another hour with biotinylated goat anti-mouse IgG (Amersham, RPN 1180). After 3 more washes, the membrane was incubated for 20 min with ³⁵S-Streptavidin (Amersham, SJ436, diluted 1:500 in the wash buffer) and washed two times. The membrane was left to air dry and exposed to Kodak XAR5 film.

For each assay a concentration series of commercially available human growth hormone (Boehringer Mannheim, 252964) was also transferred to the membrane. Autoradiograms were scanned with a densitometer and a standard curve was obtained by plotting the relative densities of the bands against the amount of growth hormone. An example of such a series of growth hormone standards as well as the resulting standard curve can be seen in fig. 2.5 B and C

Figure 2.5: Growth hormone assay.

A: Flow chart of the assay. Abbreviations: NC, nitrocellulose; hGH, human growth hormone; IgG, Immunoglobulin G.

B: Example of an autoradiogram of the standard series of purified human growth hormone assayed by the procedure in A.

C: Standard curve of the series in B, obtained by plotting the relative density of the bands (as determined by densitometric scanning of the autoradiogram) against the amount of human growth hormone blotted onto the membrane.

A

GROWTH HORMONE ASSAY

Slot-blot media onto NC membrane



Block non-specific sites



Monoclonal anti-hGH IgG



Biotinylated anti-mouse IgG



³⁵S-streptavidin



Autoradiography

B

ng hGH

0.5

1.0

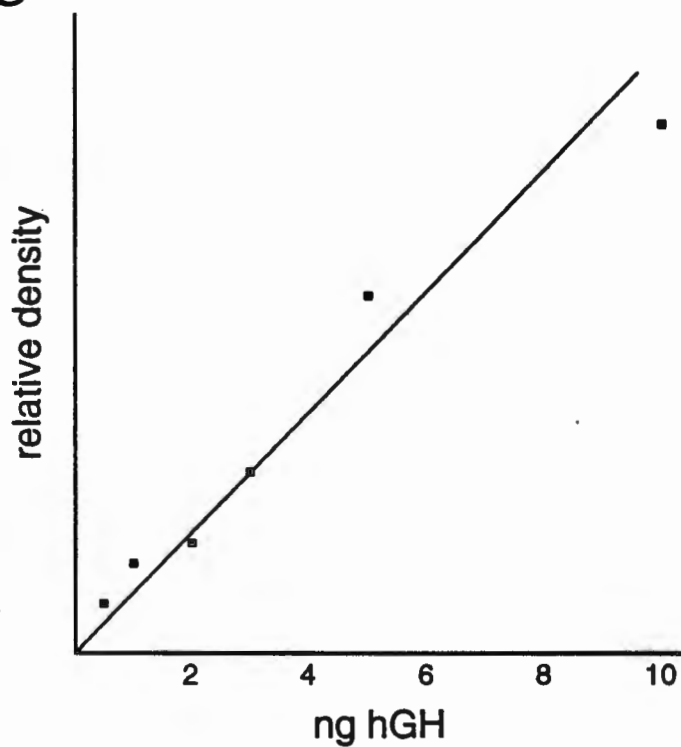
2.0



3.0

5.0

10.0

C

2.3.2 Results

The various plasmids derived from pCMR2000 and pCMR1000, as well as p0GH and pXGH5 were individually transfected into three different cell lines, namely cos-7, C127 and NIH-3T3 cells. Cos-7 cells are monkey kidney cells transformed by SV40 virus (Gluzman, 1981), C127 cells are mouse breast cancer cells (Lowy *et al*, 1978) and NIH-3T3 cells are mouse fibroblasts. These cell lines were chosen because they had previously been shown to express a reporter gene linked to the 250 bp pro- α 1(I) collagen promoter (Vergeer, 1988; Rossouw *et al*, 1987).

Autoradiograms of the growth hormone assay performed on the media of cells transfected with the constructed plasmids, are given in fig. 2.6. The autoradiograms were, as mentioned before, scanned with a densitometer to measure the relative density of each band. Standard curves were employed to determine the amount of growth hormone present in each sample. The values thus obtained, were corrected by subtracting the value for p0GH (which serves as a negative control). Both absolute and corrected values are given in table 2.1 and are depicted graphically in fig. 2.7.

Human growth hormone could be detected in the cell culture medium after transfection of the different plasmids. This indicates that the collagen-growth hormone fusion protein is successfully processed and secreted into the medium.

With all three cell lines tested, similar amounts of growth hormone was produced after transfection of pSV2000 and pTK2000. The values obtained for pMT-IIa-2000 and pfos2000 are in the same range as pTK2000 and pSV2000 in NIH-3T3 and C127 cells. However, in cos-7 cells very little growth hormone was present in the medium after transfection with these two plasmids. In all three cell lines p β glob2000 proved to be a good negative control, since transfection of this plasmid resulted in similar or lower levels of growth hormone as p0GH, the promoterless plasmid.

The amount of growth hormone detected in the media of cells transfected with pCMR2000 are, in all three cell lines tested, similar

or (in the case of NIH-3T3 cells) up to two-fold higher than that of pTK2000 and pSV2000.

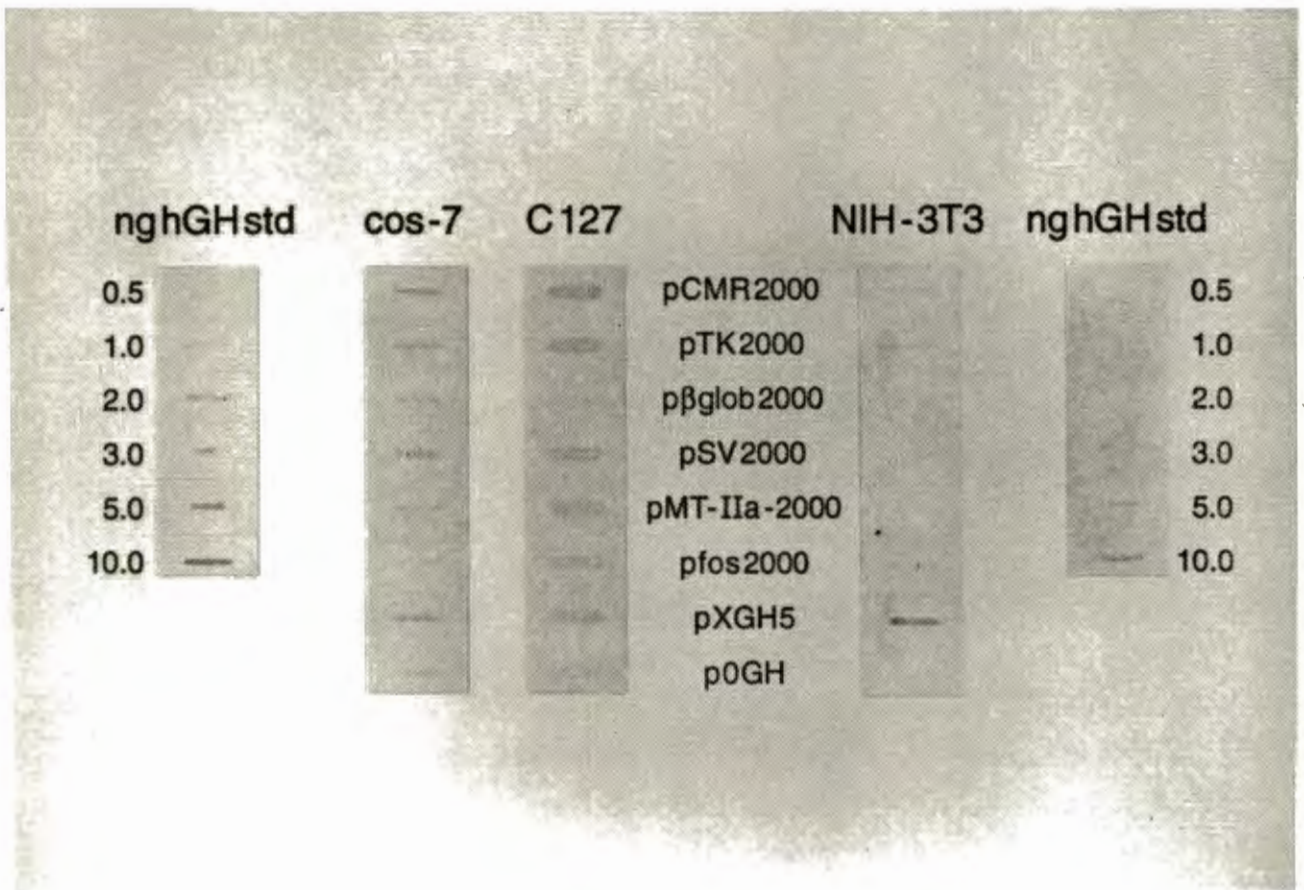


Figure 2.6: Growth hormone assay of medium of transfected cells.

The cells were transfected with 10 pBR units of the respective plasmids where a pBR unit is defined as the ratio of the plasmid's size to that of pBR in μg . Four days after transfection, 100 μl of medium was assayed for the presence of growth hormone using the assay described in section 2.3 and fig. 2.5.

The growth hormone standards on the left panel are from the membrane with the cos-7 and C127 samples, while the standards on the right refer to the NIH-3T3 samples.

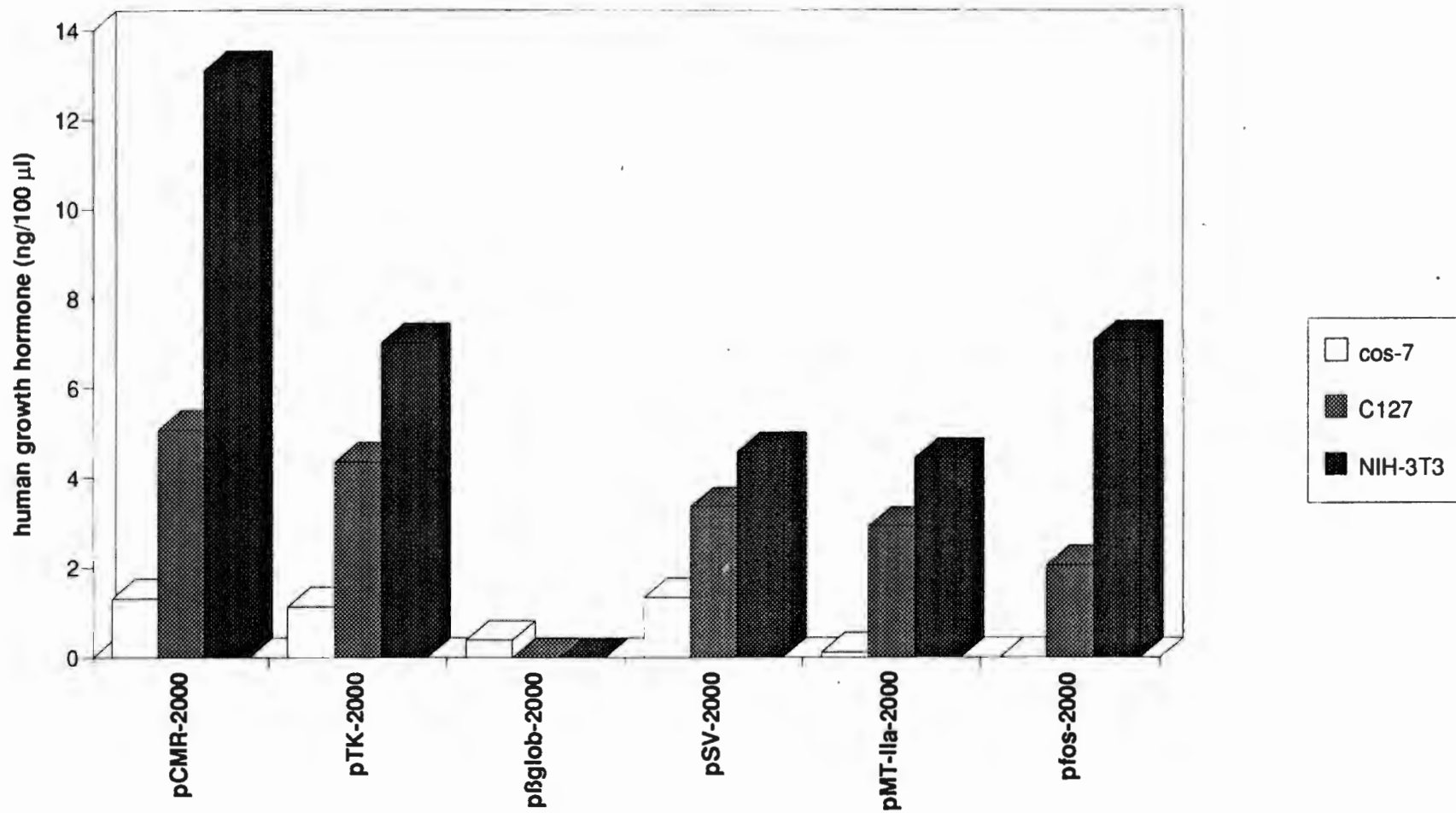
plasmid	Absolute values (ng hGH/100 μ l)			Corrected values (ng hGH/100 μ l)		
	cos-7	C-127	NIH-3T3	cos-7	C-127	NIH-3T3
pCMR2000	1.76	6.65	14.01	1.31	5.06	13.1
pTK2000	1.58	5.94	7.92	1.13	4.35	7.01
p β glob2000	0.84	1.58	0.66	0.39	0.0	0.0
pSV2000	1.78	4.49	5.48	1.33	3.35	4.57
pMT-IIa-2000	0.56	4.50	5.35	0.11	2.91	4.44
pfos2000	0.11	3.65	7.97	0.0	2.06	7.06
pXGH5	3.34	3.56	55.04	2.89	1.97	54.13
p0GH	0.45	1.59	0.91	0.0	0.0	0.0

Table 2.1: Amounts of growth hormone in medium of transfected cells.

The autoradiograms were scanned with a densitometer and the amount of growth hormone determined from the curve of the growth hormone standards. Corrected values were obtained by subtraction of the value of p0GH (the promoterless plasmid) from the absolute value.

Figure 2.7: Schematic presentation of corrected values of amount growth hormone produced by the transfected cells.

The data of table 2.1 presented graphically.



2.4. Conclusion

The constructs chosen, with the first exon of the pro- α 1(I) gene fused to the growth hormone gene in the first intron resulted in immunologically detectable growth hormone in the growth medium of the cells. This provided an easy and fast way to measure expression of the reporter gene under control of the different promoters.

However, since the growth hormone assay detects protein product, it is not necessarily a direct measure of transcription of the reporter gene. In order to compare transcriptional activity of different promoters more accurately, an analysis of the RNA should be performed. RNAase protection or S1 analyses of the cytoplasmic RNA can be used to determine the relative level of transcription and the ratio of correctly initiated vs readthrough transcripts. RNA analyses provide furthermore the possibility to simultaneously assay a reference gene with the reporter gene. Such a reference gene which is detected in the same assay as the reporter gene is the ideal internal control, since it allows one to correct for differences in both transfection efficiency and extraction and assaying of the RNA.

Since we assayed the presence of protein product and not RNA, and lack therefore an internal control, the values obtained can not be taken as absolute measures of the activity of the different promoters. Nevertheless, our results do indicate that expression of the growth hormone gene with the pro- α 1(I) collagen promoter is comparable to other strong promoters like the SV40 early promoter-enhancer, HSV thymidine kinase promoter (a house keeping gene) and the metallothionein IIa promoter. This suggests that the pro- α 1(I) promoter is a relatively strong promoter, even in the absence of an enhancer. In NIH-3T3 cells, which synthesize high amounts of type I collagen, expression of the plasmid with the pro- α 1(I) promoter was two fold higher than with the SV40 promoter-enhancer. This suggests the possibility that the pro- α 1(I) promoter might contain elements involved in tissue-specific regulation of the gene's expression.

As these results indicate that the 250 bp pro- α 1(I) promoter is a strong promoter, we would expect that it contains not only a TATA box and transcription initiation site but also promoter elements mediating transcriptional activation. At the time of this study, computer analyses of the promoter sequence had shown the presence of putative promoter elements such as a CCAAT-box and a Sp1 binding site as well as a pyrimidine rich element which could be involved in transcriptional regulation of the gene (Rossouw *et al*, 1987). In order to identify which promoter elements could play a role in transcriptional regulation of this gene, we decided to do an analysis of the binding of nuclear factors to the pro- α 1(I) collagen promoter.

CHAPTER 3

BINDING OF NUCLEAR FACTORS TO THE PRO- α 1(I) COLLAGEN PROMOTER

3.1. Introduction

The experiments described in the previous chapter led to the following questions:

- Do nuclear factors bind to elements in the promoter?
- Is there a difference in DNA-protein interactions when using nuclear extracts from different cell lines?
- To which DNA motifs do nuclear factors bind?

In order to answer these questions, nuclear extracts were prepared from cell lines either expressing or not expressing type I collagen. These extracts were used in bandshift and DNAase I footprinting experiments, as well as southwestern blotting analyses.

3.2. DNA motifs in the promoter

The pro- α 1(I) promoter sequence, as published by Bornstein *et al* (1987), differs in some aspects from that published earlier by Chu *et al* (1985) (fig. 3.1 B). We performed a computer search for known sequence elements in the promoter and found that a putative Sp1-site, present in the sequence reported by Chu *et al*, does not exist in the sequence of Bornstein *et al*. In the sequences of Chu *et al* the T residue at position -225 is absent, resulting in the sequence GGCGGG which is a potential binding site for Sp1 (fig. 3.1 B).

In order to determine which sequence elements were present in the promoter and are potentially bound by nuclear factors, we needed to establish the correct sequence of the promoter. To confirm the sequence, we subcloned a 350 bp Pvu II - Xba I fragment containing 260 bp of promoter sequences (the minimal promoter of Rossouw *et al*, 1987) as well as 90 bp 5' untranslated sequence (fig. 3.1 A) into M13. Single strand DNA was prepared and

used to perform dideoxy sequencing. This confirmed that the sequence published by Bornstein *et al* (1987) was correct.

The consensus elements identified by the subsequent computer search are indicated in fig. 3.1 B. A TATA-box is located at position -32 relative to the transcription initiation site (position +1). Both Rossouw *et al* (1987) and Bornstein *et al* (1987) indicated the sequence CCAAAT at position -142 as the CCAAT box. This does not correspond to the CCAAT-box consensus sequence. Instead, our search revealed the presence of two inverted CCAAT-boxes at positions -106 and -133. Further putative binding sites for known transcription factors are a Sp1 site at position -92 (consensus GGGCGG) and a sequence resembling the AP-2 binding site of the SV40 enhancer (Mitchell *et al*, 1987) at position -203 (TCCCCAGCC on the lower strand). Another interesting feature of the promoter is the presence of a pyrimidine rich sequence between -145 and -180 which is highly homologous to a similar region in the first intron of the gene (Rossouw *et al*, 1987).

Figure 3.1: Restriction map and sequence of the 5' end of the human pro- α 1(I) collagen gene.

A: Restriction map of the 5' untranscribed region and the first five exons. Abbreviations: B, Bam HI; Pv, Pvu II; Xb, Xba I; Nc, Nco I; Hd, Hind III, EV, Eco RV; Ap, Apa I; Sm, Sma I. Symbols: Line, 5' untranscribed region; shaded box, 5' untranslated region; filled boxes, exons; open boxes, introns.

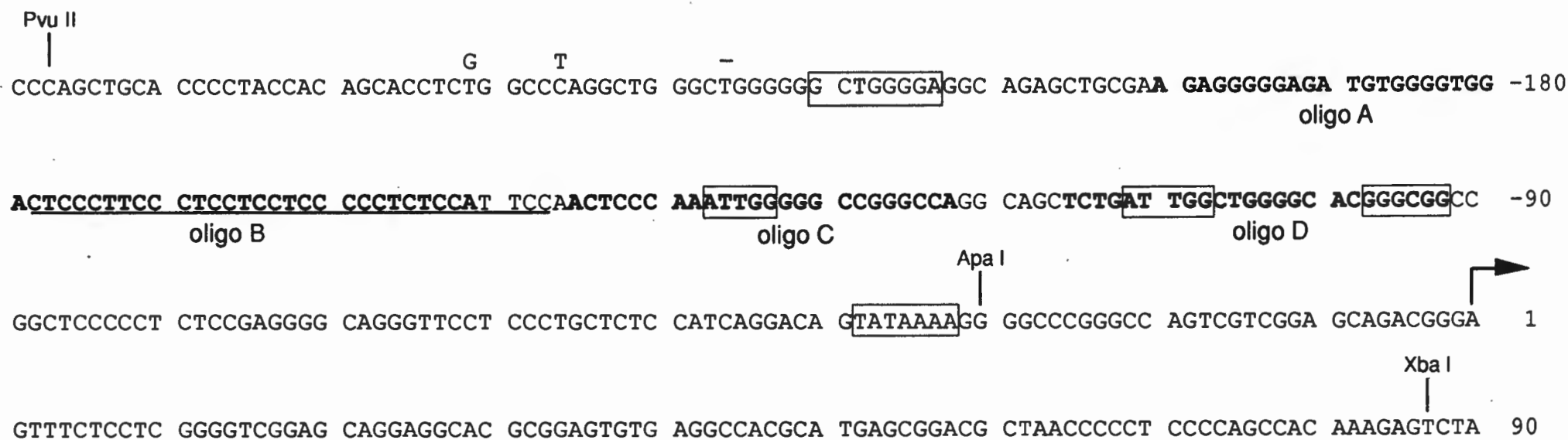
B: Sequence of the promoter from the Pvu II site (-263) to the Xba I site (+90). Numbers refers to nucleotide positions when the start site of transcription is taken as +1. Putative binding sites for AP-2 (GCTGGGGA), Sp1 (GGGCGG), the two CCAAT-boxes (ATTG) and the TATA-box are boxed and the pyrimidine stretch is underlined.

Bases where the sequence published by Chu *et al* (1985) differs from our own and that of Bornstein *et al* (1987) are indicated above the sequence. The four oligonucleotides used in bandshift and southwestern blotting are in bold letters and the identity of each oligonucleotide (oligo A, B, C, or D) is indicated below.

A



B



3.3. DNA-protein interactions with the promoter

3.3.1 Electrophoretic mobility shift assay.

For an overall indication of DNA-protein interactions with the promoter, electrophoretic mobility shift assays (bandshifts) were performed. To obtain a labeled fragment for use in the bandshift experiments, the 350 bp Pvu II - Xba I fragment used for sequencing, was subcloned into the Xba I site of pSP64. The resulting plasmid was digested with Hind III, dephosphorylated with calf intestine phosphatase and treated with polynucleotide kinase in the presence of γ -³²P ATP. After the kinasing reaction, the plasmid was digested with Bam HI and the labeled fragment purified by polyacrylamide gel electrophoresis.

Nuclear extracts were prepared essentially as described by Dignam *et al* (1983) and the protocol followed is summarized in the flow chart shown in fig. 3.2. Fibroblasts were scraped off in phosphate-buffered saline (PBS), while cells growing in suspension (Molt-4, BJA-B and HeLa cells) were collected by centrifugation. Cells were washed with 5 volumes cold PBS, centrifuged 15 minutes at 2 500 r.p.m. and resuspended in 5 volumes of buffer A (10 mM Hepes (pH 7.9), 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM DTT and 0.5 mM PMSF). After 10 minutes on ice, they were spun down (15 minutes at 2 000 r.p.m) and resuspended in buffer A to a volume of about 15 ml. After addition of protease inhibitors (Antipain, Leupeptin, Pepstatin A, and Chymotrypsin ABC) to 5 μ g/ml, the cells were homogenized with 24 strokes of a 'S' glass-glass Dounce homogenizer. The homogenate was centrifuged for 20 minutes at 10 000 g and the nuclear pellet was resuspended in buffer C (20 mM Hepes (pH 7.9), 1.5 mM MgCl₂, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT, 25% glycerol, 0.6 M KCl) to a volume of 12 ml. This was homogenized again with 10 strokes of a 'S' glass-glass Dounce homogenizer and mixed on a horizontal rotating table for 45 minutes. After centrifugation for 30 minutes at 125 000 g, the supernatant was dialyzed against 50 volumes buffer D (20 mM Hepes (pH 7.9), 50% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT) for 5-6 hours. The nuclear extract was centrifuged for 15 minutes in an

Eppendorf centrifuge, after which the supernatant was frozen in a dry ice-ethanol bath and stored at -70°C .

PREPARATION OF NUCLEAR EXTRACTS

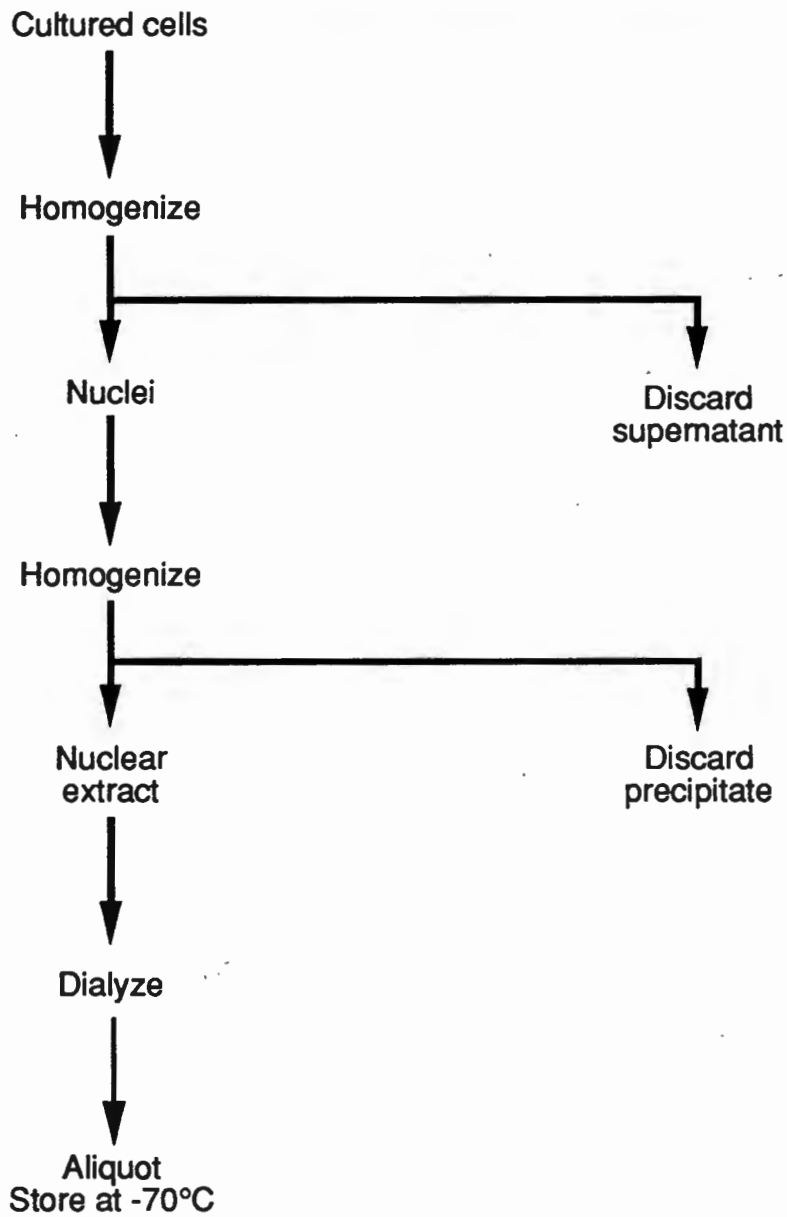


Figure 3.2: Preparation of nuclear extracts. Flowchart of strategy followed in the preparation of nuclear extracts from NIH-3T3, BJA-B, Molt-4 and HeLa cells.

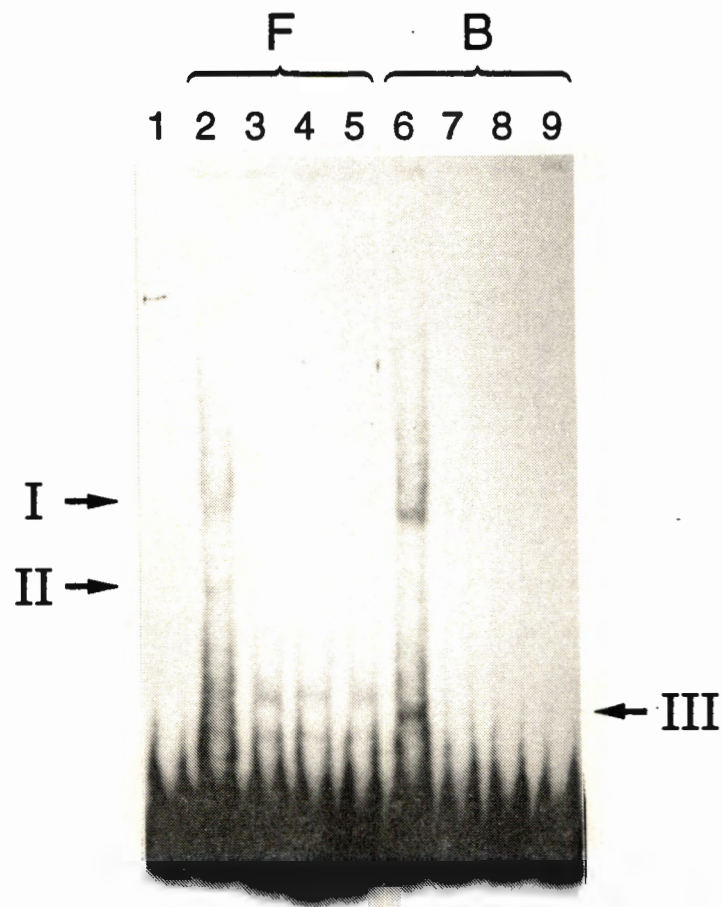
Bandshift experiments were performed as described by Schöler *et al* (1989). In short, the protocol was as follows. The end labeled DNA fragment and appropriate amounts of nuclear extract were incubated in the presence of 5 μ g BSA and 5 μ g poly(dI.dC) in a total volume of 20 μ l. The reaction mixture consisted of 14 mM Hepes, 1 mM spermidine, 5 mM MgCl₂, 60 mM KCl, 0.7 mM DTT, 0.1 mM EDTA, 0.1 mM ZnCl₂ and 5% glycerol. After an incubation of 15 min at 25°C, the samples were separated on a 3.5% native polyacrylamide gel with 0.1 M Hepes (pH 7.9) and 1 mM EDTA as running buffer.

As shown in fig. 3.3, bandshifts with nuclear extracts from fibroblasts and B-cells revealed the formation of specific DNA-protein complexes. Differences were observed between complexes formed with nuclear extracts from expressing and non-expressing cells. With extracts from NIH-3T3 cells, two complexes, I and II in lane 2 are formed which can be competed out by addition of an excess of unlabeled fragment (lanes 3 - 5). When using extracts from BJA-B cells, the same complexes, as well as another faster migrating complex, III in lane 6 are present. As in the case of NIH-3T3 extracts, complexes I and II, as well as III are competable with an excess of unlabeled fragment (lanes 7 - 9).

Although complexes I and II are formed with nuclear extracts of both cell lines, they are present in different ratios. With NIH-3T3 extracts equal amounts of both complexes were formed. This was not the case for BJA-B extracts, where complex I was favored.

Figure 3.3: Electrophoretic mobility shift assay with the 250 bp Pvu II - Xba I pro- α 1(I) promoter fragment.

A fragment containing the minimal promoter was end-labeled and incubated with nuclear extracts from either fibroblasts (lanes 2 to 5) or lymphocytes (lanes 6 to 9) in the absence (lanes 2 and 6) or presence of 100x (lanes 3 and 7), 200x (lanes 4 and 8) or 250x (lanes 5 and 9) molar excess of unlabeled fragment as competitor. Lane 1 is the fragment in the absence of nuclear extract. Specific complexes are indicated with arrows.



3.3.2 DNAase I footprinting analysis.

To define regions where nuclear factors interact with the promoter, DNAase I footprinting analyses were performed. A fragment stretching from the Pvu II site (position -263) to Apa I site (position -30) was excised from pIII α 1 (Vergeer, 1988) using restriction endonucleases Cla I (in the pBR322 sequences, 9 bp upstream of the Pvu II site) and Apa I. It was 3' end labeled at the Cla I site with the Klenow fragment of DNA polymerase I in the presence of α -³²P dCTP (Maniatis *et al.*, 1982).

DNAase I footprinting was performed as described by Ohlson and Edlund (1986) with minor variations. Standard reaction mixtures contained the following in a final volume of 50 μ l: 25 mM Hepes (pH 7.8), 25 mM KCl, 0.05 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF and 15% glycerol. Prior to addition of the labeled fragment, up to 60 ng of protein extract was incubated 20 minutes on ice with 125 or 250 ng poly(dI.dC). After this, 1 ng of the 3' end labeled DNA fragment was added and incubated on ice for 35 minutes and then at room temperature for 5 minutes. MgCl₂ was added to a concentration of 5 mM, followed by freshly diluted DNAase I (20 KU/ μ g DNA). DNAase I digestion was allowed to take place for 60 seconds at room temperature and was terminated by addition of 100 μ l stop solution consisting of 100 mM Tris.HCl (pH 8), 300 mM NaAc (pH 5), 0.5% sarkosyl, 10 mM EDTA, 100 μ g/ml proteinase K and 0.2 mg/ml tRNA. The reactions were incubated at 37°C for 15 minutes and at 95°C for 2 minutes after which they were extracted twice with phenol:chloroform (1:1) and once with chloroform:iso-amylalcohol (24:1). The fragments were ethanol precipitated and analyzed on 8% denaturing polyacrylamide gels (Maxam and Gilbert, 1980). To determine the boundaries of protection more precisely, Maxam and Gilbert sequencing (Maxam and Gilbert, 1980) was performed on the same 3' end labeled fragment and run alongside the samples.

As with the bandshift experiments, different protection patterns were obtained when using nuclear extracts from expressing or non-expressing cells (fig. 3.4 A and B). From the Maxam and Gilbert sequencing lanes run alongside the samples (not shown), the exact boundaries of the protection could be determined.

With fibroblast extracts, three areas were strongly protected (fig. 3.4 A), namely position -80 to -99 (over the Sp1-site), -116 to -152 (covering the distal CCAAT-box), and at position -174 to -203. Weaker protection was observed at position -156 to -173 (in the pyrimidine rich region) and at -206 to -219 (over the Ap-2 site). DNAase I hypersensitive sites were induced at positions -173, -205 and -220, thus marking the boundaries of some of the protected areas.

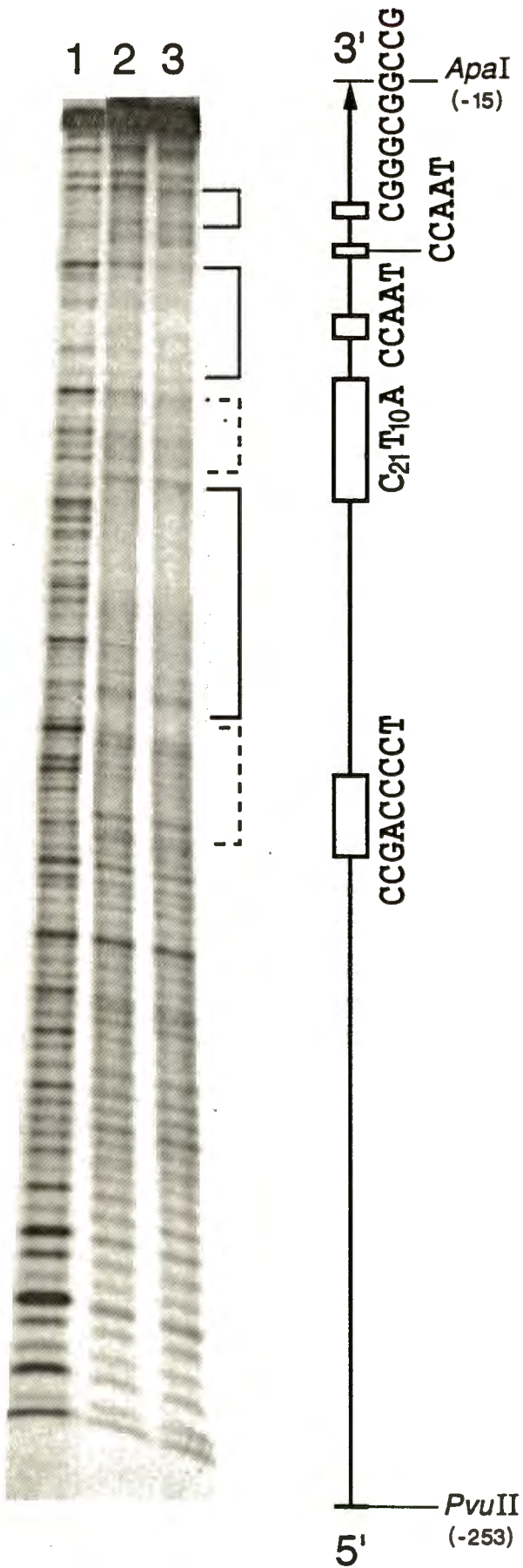
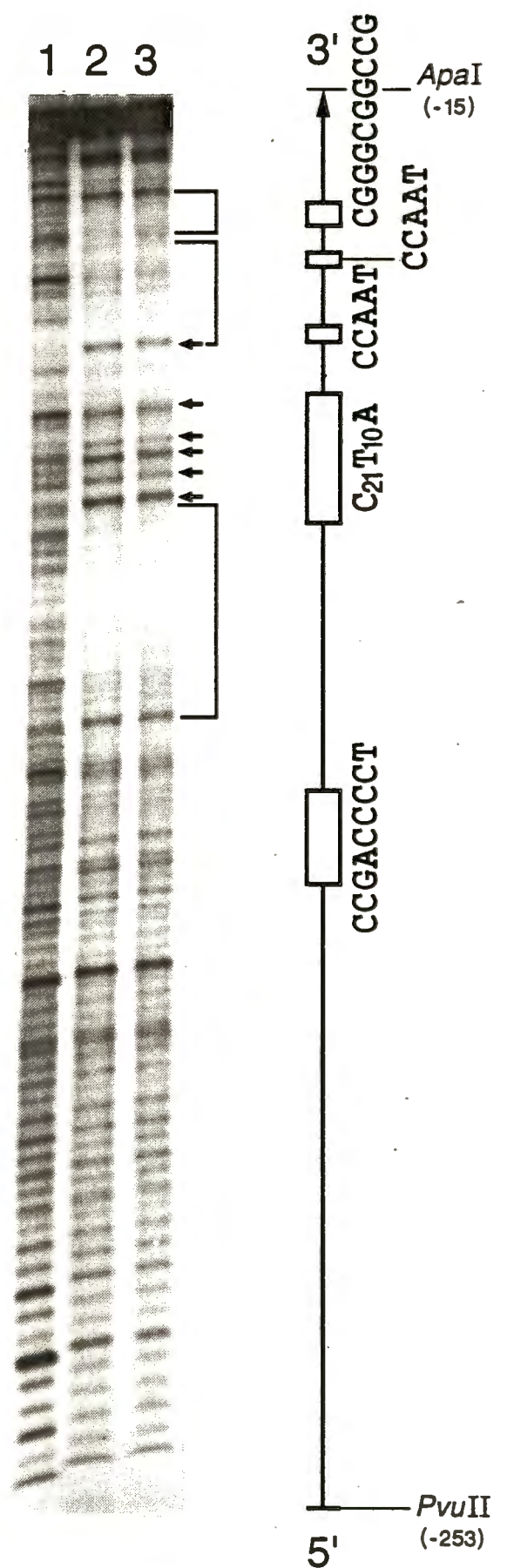
With lymphocyte extracts, strong protection was observed at positions -80 to -102 (over the Sp1 site), -105 to -137 (the CCAAT-boxes) and -174 to -203 (fig. 3.4 B). DNAase I hypersensitive sites were induced at the boundaries of some of the protected areas, at -138, -173 and -204 as well as several sites between -150 and -173 (in the pyrimidine rich area).

Figure 3.4: DNAase I footprinting analysis of the pro- α 1(I) promoter.

The Pvu II - Apa I promoter fragment was 5' end labeled and used in DNAase I footprinting assays. The precise areas of protection were determined by running Maxam and Gilbert sequencing reactions with the same fragment alongside the gels (not shown). The positions of putative sequence elements in the promoter (see fig. 3.1 B) are shown next to the gels. Protection against DNAase I is indicated by brackets while DNAase I hypersensitive sites that were induced are marked by arrows. Dotted brackets indicate weaker protection.

A: Fibroblast nuclear extracts. Lane 1 is in the absence of nuclear extract and lane 2 and 3 with 60 ng extract. Lanes 1 and 2: binding reactions were performed in the presence of 125 ng poli(dI.dC); lane 3: 250 ng poli (dI.dC)

B: Lymphocyte nuclear extracts. Lanes as in (A).

A**B**

The protection pattern with lymphocytes differs therefore in certain respects from that with fibroblast extracts. The protection over the CCAAT-boxes, which is centered on the distal CCAAT-box with fibroblast extracts, is shifted downstream when using lymphocyte extracts and covers both CCAAT-boxes. While the pyrimidine rich area is weakly protected by fibroblast extracts, strong DNAase I hypersensitive sites are induced here in the case of lymphocyte extracts. The weak fibroblast protection over the AP-2-site (called col-A) is not observed with lymphocyte extracts. Similarities between the the two cell lines include the protection over the Sp1 site (col-F) as well as the protection at position -174 to -203, immediately upstream of the pyrimidine rich area. In both cases the latter protection was the strongest and is bordered on both sides by the induction of DNAase I hypersensitive sites. This protected area contains no recognition site of a known transcription factor and we named it the col-B site.

3.4. Binding of nuclear factors to DNAase I protected elements.

3.4.1 Electrophoretic mobility shift assay.

In order to determine the significance of putative DNA elements as well as protected regions, oligonucleotides were synthesized to be used in bandshift and southwestern blotting experiments. The sequences of the oligonucleotides used are indicated on fig. 3.1 B. These include the col-B site which was strongly protected by nuclear extracts from both cell lines, as well as the pyrimidine rich area and the two CCAAT-boxes where we observed differences in the protection when using different cell lines. Oligonucleotide A stretches from -177 to -200 and thus covers the Col-B site. Oligonucleotide B contains the pyrimidine rich stretch from -151 to -180, and oligonucleotides C and D the distal and proximal CCAAT-boxes respectively (oligonucleotide C spans from -122 to -145 and oligonucleotide D from -92 to -115).

The oligonucleotides were end labeled by treating them with polynucleotide kinase in the presence of γ -³²P ATP. After gel-purifying the labeled oligonucleotides, they were used as probes in

bandshift experiments with nuclear extracts of fibroblasts and lymphocytes.

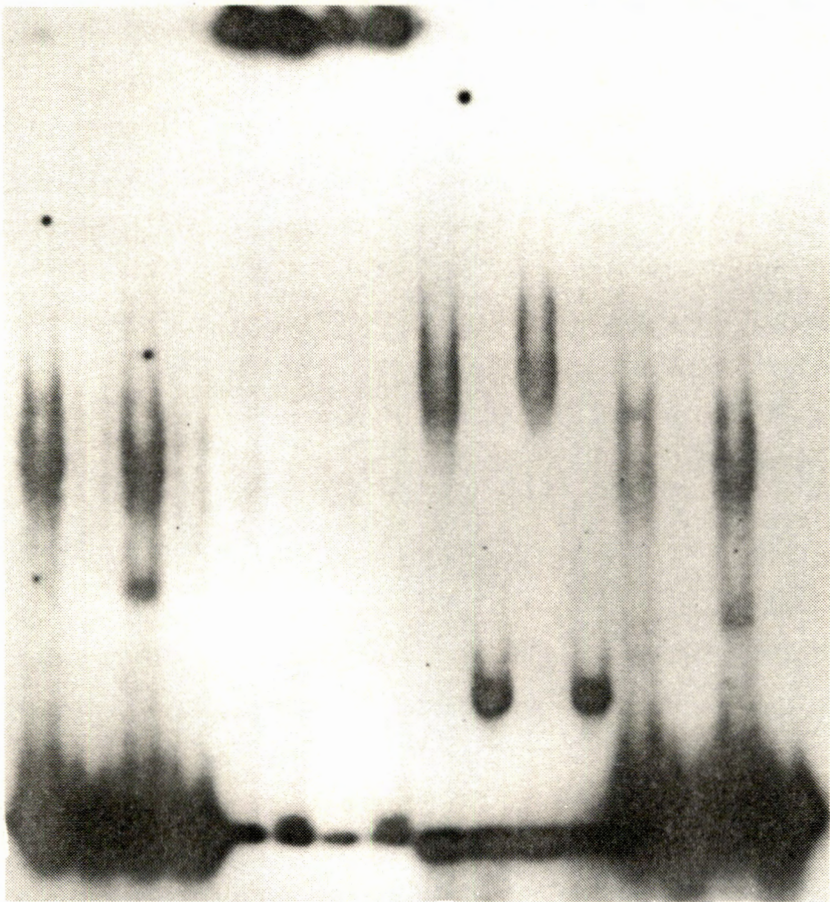
With both fibroblast and lymphocyte nuclear extracts, oligonucleotide A (Col-B site) was shifted (fig. 3.5 A). In addition to one protein-DNA complex common for both cell lines, there was another, faster migrating complex with lymphocyte extracts. While the latter, lymphocyte specific band was efficiently competed away by a 250 molar excess of unlabeled oligonucleotide A, the bigger, common band could only be partially competed out by the same amount. Additional experiments have shown that for total competition of this complex, a 500 molar excess of competitor is necessary (data not shown). The band common to both the fibroblast and lymphocyte lanes didn't run as a discrete band but rather as a smear, indicating that the formed complex(es) are not stable during electrophoresis.

In the bandshift of oligonucleotide B (the pyrimidine rich stretch) a light smear was observed which could be specifically competed away (fig 3.5 B). However, with this oligonucleotide most of the labeled fragment remained in the slots. This was also the case in the presence of specific competitor, suggesting the formation of non-specific complexes. In contrast to the DNAase I footprint data, which revealed a pronounced difference between the interaction of nuclear factors from fibroblasts and lymphocytes with the pyrimidine rich stretch, the bandshift with oligonucleotide B is the same with extracts from either cell line.

Figure 3.5: Electrophoretic mobility shift assay with oligonucleotides corresponding to defined sequence elements in the pro- α 1(I) promoter.

Oligonucleotides A to D (see fig. 3.1 A) were end-labeled and incubated with nuclear extract from either fibroblasts (F) or lymphocytes (L) in the absence (-) or presence (+) of 250 fold molar excess of the respective unlabeled oligonucleotides.

A		B		C		D	
F	L	F	L	F	L	F	L
-	+	-	+	-	+	-	+



The complexes formed with oligonucleotide C (distal CCAAT-box) was different when using fibroblast or lymphocyte extracts. With lymphocytes the complex formed migrated somewhat slower than the fibroblast complex. Both complexes could also be specifically competed away by a 250 molar excess of unlabeled oligonucleotide (fig. 3.5 C). Another faster migrating band was formed in the presence of competitor. This band was however also present in reactions where no nuclear extract had been added (data not shown) and is therefore not due to the formation of another DNA-protein interaction. This could be the result of some secondary structure assumed by the oligonucleotide itself.

The shift of the proximal CCAAT-box (oligonucleotide D) differs from that of the distal CCAAT (oligonucleotide C) and in this case there were also differences between the two extracts used. One complex, appearing as a smear on the gel, was present when incubating with either extract, though it was more pronounced with the lymphocyte extract (fig. 3.5 D). With fibroblast extracts, there was another slower migrating complex and with lymphocytes one that was smaller and thus migrating ahead of the common complex. As in the case of the other oligonucleotides, these complexes were all competed away by excess of specific competitor.

3.4.2. Southwestern blotting analyses.

Southwestern blotting was used to obtain an indication of the molecular weight of nuclear factors binding to the oligonucleotides. This analysis was performed as described by Hübscher (1987) and is in short as follows : Nuclear extracts were mixed with equal volumes of sample buffer which consisted of 5% SDS, 5 mM Tris.Cl (pH 6.8), 200 mM DTT, 20% glycerol and 0.05% pyronin y. The proteins were separated by SDS/PAGE on a 7% polyacrylamide gel (Laemlli, 1970). Amersham's rainbow protein markers were run alongside the samples as molecular weight markers. The proteins were electrophoretically transferred to nitrocellulose using a transfer buffer of 25 mM Trizma base, 190 mM glycine and 1 mM EDTA

The nitrocellulose blots were renatured for 24 hours at 4°C in a buffer consisting of 10 mM Hepes (pH 7.9), 1 mM DTT, 0.1 mM EDTA, 10 mM MgCl₂, 50 mM NaCl, 0.1 mM ZnSO₄, 10% glycerol and 5% non-fat dry milk. The blots were cut into strips of 3 mm each and incubated overnight at 4°C in the same buffer with less milk (0.25%) and with 25 ng of the labeled oligonucleotide and 5 µg poly(dI.dC) added. After hybridization with the probe, the strips were washed for 1 hour in two changes of binding buffer without probe and exposed to Cronex 4 film.

In order to obtain large scale nuclear extracts, HeLa cells growing in suspension cultures were used instead of NIH-3T3 cells for a cell line expressing type I collagen. Nuclear extracts of HeLa cells and lymphocytes were separated by polyacrylamide gel electrophoresis and hybridized with the same oligonucleotides as was used for the bandshifts.

The col-B oligonucleotide (oligonucleotide A) hybridized to a factor with apparent molecular weight of 80 kd present only in HeLa extracts as well as factors of 48 and 40 kd in both HeLa cells and lymphocytes (fig. 3.6 lanes a). Further bands that appear in the col-B lanes are also present in all the other lanes indicating that it may be the result of non-specific binding. These include factors of 120 kd and 55 kd present in all the lanes and a factor of 60 kd in all lymphocyte lanes. In order to verify which bands correspond to specific binding to any one of the oligonucleotides, binding has to be challenged by an excess of the relevant oligonucleotide as competitor. In the case of the col-B oligonucleotide, these experiments have been done in our laboratory (W.N. de Vries, unpublished results) and demonstrated that binding to the 40 kd factor is specific.

In addition to the non-specific bands, the pyrimidine-oligonucleotide (oligonucleotide B) hybridizes strongly to a factor of 48 kd present in both extracts (fig. 3.6 lanes b). With HeLa extracts this oligonucleotide hybridizes furthermore to a factor of 80 kd and with both extracts there is a weak signal at 40 kd.

The signals from the CCAAT-oligonucleotides (oligonucleotides C and D) was significantly weaker than those obtained with oligonucleotide A and B (compare lanes c and d with lanes a and b in fig. 3.6 A). Different exposures of the blots were made and aligned for comparable results (fig. 3.6 B). As already indicated in the bandshifts, the two CCAAT-oligonucleotides hybridized to different factors. The distal CCAAT-oligonucleotide recognizes a HeLa factor of 94 kd and factors of 48, 46, 40 and 35 kd on the blots of both HeLa cells and lymphocytes (fig. 3.6 lanes c). The proximal CCAAT-box is bound by a factor with a molecular weight of 38 kd present in both extracts, but it is more pronounced in the case of lymphocyte extracts. In addition, there is a 80 kd HeLa factor binding to oligonucleotide D (fig. 3.6 lanes d).

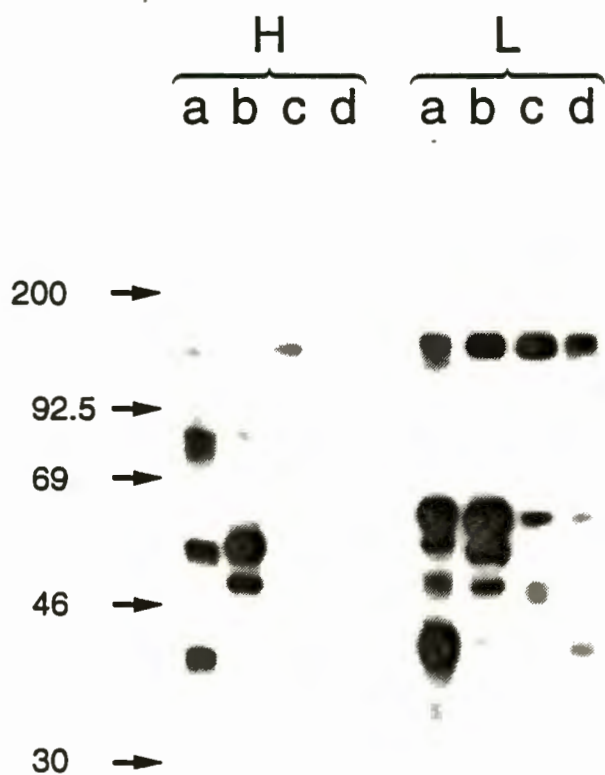
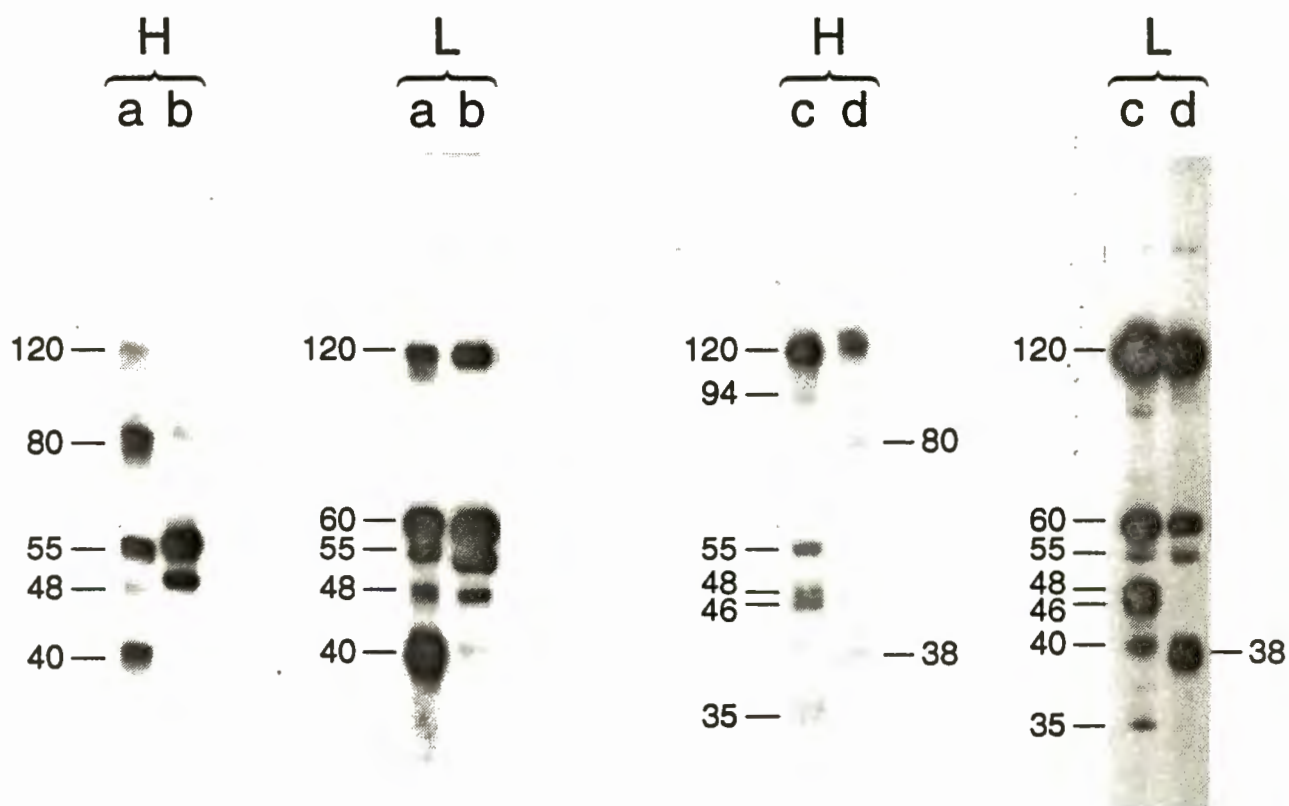
Figure 3.6: Southwestern blotting analyses with oligonucleotides corresponding to defined sequence elements in the pro- α 1(I) promoter.

HeLa (H) and lymphocyte (L) nuclear extracts were separated on a denaturing protein gel and transferred to a nitrocellulose membrane. The membrane was cut into strips which were then incubated with the labeled oligonucleotides.

A: Lanes a,b,c and d were incubated with oligonucleotides corresponding to sequences containing col-B, the pyrimidine element, and the distal and proximal CCAAT-boxes, respectively. Numbers indicate molecular weights of rainbow marker protein standards.

B: Shorter exposures of lanes a and b shown in panel A. Numbers indicate apparent molecular weights for individual bands.

C: Longer exposures of lanes c and d shown in panel A. Numbers indicate apparent molecular weights for individual bands.

A**B**

3.5. Conclusion

In summary, both bandshift and DNAase I footprinting experiments demonstrated specific binding of nuclear factors to promoter sequences. In addition, differences in DNA-protein interactions were observed when using nuclear extracts of expressing or non-expressing cells.

The DNAase I footprint data, in comparison with a computer analysis of the promoter sequences, indicated the potential significance of certain motifs present in the promoter. These motifs are putative AP-2 and Sp1 sites, two inverted CCAAT-boxes and a pyrimidine rich area homologous to a similar area in the intronic enhancer.

An additional binding site had been identified which was protected by (a) factor(s) present in nuclear extracts of both expressing and non-expressing cells. This site, named the col-B site, is immediately upstream of the pyrimidine rich area and does not contain the recognition sequence of any known transcription factor.

In three areas there were differences between the DNAase I protection patterns from fibroblasts or lymphocytes. In lymphocytes (non-expressing cells), the protection over the CCAAT-boxes was shifted downstream with respect to the protection by fibroblast factors. In the latter case, protection was centered on the distal CCAAT-box, while with lymphocyte extracts it stretches from the distal CCAAT-box till past the proximal one. In addition, the lymphocyte extracts lacked the protection observed with fibroblast extracts in the pyrimidine stretch and over the AP-2 site (Col-A). In fact, instead of protection of the pyrimidine stretch, several DNAase I hypersensitive sites were induced in this area after incubating the fragment with lymphocyte extracts.

DNAase I footprints were performed on fragments labeled at the Cla I site, 9 bp upstream of the Pvu II site of the $\alpha 1(I)$ promoter. For a more precise picture of the protection pattern of the promoter, DNAase I footprints should also be done on a fragment

labeled from the other side. This would show the boundaries of protection on the other strand and would furthermore result in a better resolution of the region immediately upstream of the start site of initiation. In the gels shown in fig. 3.4 the TATA-box (-32 to -38) can for example not be seen, since it is too far from the labeled end of the fragment.

Bandshift experiments with oligonucleotides of the col-B site, the pyrimidine stretch and the two CCAAT-boxes, showed that each of these motifs are indeed specifically recognized by nuclear factors. Except in the case of oligonucleotide B, the pyrimidine stretch, differences could be observed in binding of nuclear factors from expressing and non-expressing cells to all the oligonucleotides used.

Southwestern blotting analyses confirmed the binding of nuclear factors to these oligonucleotides and gave an indication of the apparent molecular weights of these factors. For all four oligonucleotides used, we could observe bands common to HeLa and lymphocyte extracts, as well as bands unique to either extract.

Although southwestern blotting is a useful tool for determining the molecular weight of factors binding to DNA, it is limited in that protein complexes are dissociated before performing gel electrophoresis. This means that only proteins which do not consist of multiple subunits and can bind DNA as monomers are detected by this method. It is therefore not a complete representation of all the factors that bind to a specific fragment of DNA. The denaturing conditions used for separating the proteins furthermore opens the possibility of artefacts, i.e. binding to factors which would not occur if these factors had been in their native conformation.

The significance of the results described here will be discussed in more detail in the next chapter.

CHAPTER 4

DISCUSSION

4.1 Introduction

In this study the minimal promoter as defined by Rossouw *et al* (1987), i.e. 250 bp upstream of the cap site, was compared to several other promoters. We could thus show that the pro- $\alpha 1$ (I) collagen promoter is a relatively strong promoter, comparable to SV40 promoter-enhancer and the HSV thymidine kinase promoter (chapter 2).

In addition, DNA-protein interactions with the promoter were studied using nuclear extracts from expressing as well as non-expressing cells. The promoter elements where DNA-protein interactions were observed were the CCAAT-boxes, pyrimidine stretch, putative Sp1- and AP-2 sites as well as a previously unidentified region, the col-B site. Nuclear extracts from expressing and non-expressing cells showed differences in the DNA-protein interactions in the region of the duplicated CCAAT-boxes, the pyrimidine element as well as the putative AP-2 site (chapter 3).

The significance of the results described in the previous chapters will be discussed and compared with other studies in this chapter.

4.2 Functional analysis of the pro- $\alpha 1$ (I) promoter

In order to determine the relative strength of the pro- $\alpha 1$ (I) collagen promoter, this promoter, as well as several other promoters, was cloned in front of a reporter gene and transfected into tissue culture cells.

Rossouw *et al* (1987) showed that 250 bp upstream of the transcription start site is sufficient to act as a promoter directing transcription of a heterologous reporter gene (α -globin, in their case). In this study the same promoter has been used, but linked to human growth hormone as a reporter. Other groups who have done similar experiments confirmed that the pro- $\alpha 1$ (I) promoter lies

within 250 bp upstream of the cap site (Rippe *et al*, 1989; Liska *et al*, 1990; Karsenty and deCrombrughe, 1990).

Vergeer (1988) demonstrated that the cell lines NIH-3T3, C127 and cos-7, all express α -globin when linked to the 250 bp pro- α 1(I) promoter. These cell lines were therefore used for the promoter comparison. The promoters used are previously characterized promoters and include the SV40 promoter-enhancer (a strong viral promoter-enhancer), HSV thymidine kinase promoter (a housekeeping promoter), β -globin promoter (a tissue-specific promoter, only active in erythroid cells), metallothionein promoter (strong and inducible by heavy metals) and the c-fos promoter (a tissue specific promoter).

As expected, both the plasmids with the SV40 enhancer-promoter and the thymidine kinase promoter were transcribed well in all the cell lines tested. p β -glob2000 (with the β -globin promoter) served as an additional negative control as it was not expressed at levels higher than p0GH, the promoterless plasmid. With the metallothionein promoter (pMT-IIa-2000) and c-fos promoter (pfos2000), expression was different in the different cell lines. pMT-IIa-2000 was expressed poorly in cos-7 cells but in C127 and NIH-3T3 cells its expression was comparable to that of pSV2000. No expression of pfos2000 could be detected in cos-7 cells, while in C127 cells it was similar or in NIH-3T3 cells slightly higher (1.5 fold) than pSV2000.

The experiments described in chapter 2 reveal furthermore that the pro- α 1(I) collagen promoter is as strong as the SV40 enhancer-promoter and the thymidine kinase promoter in both cos-7 and C127 cells. Transfection into NIH-3T3 cells resulted however in higher expression of pCMR2000 than any of the other plasmids tested. In this cell line, expression of pCMR2000 was about two-fold higher than pTK2000 and 2.5 fold higher than pSV2000. NIH-3T3 cells are fibroblasts and produce high amounts of type I collagen. The high expression of the pro- α 1(I) promoter in these cells (higher than SV40 promoter-enhancer) suggests that the 250 bp used as promoter may contain tissue-specific control elements.

Similar studies done by other groups confirm that the pro- $\alpha 1(I)$ promoter is a strong promoter. Rippe *et al* (1989) demonstrated that the chloramphenicol acetyl transferase gene (CAT gene) with 220 bp of the murine $\alpha 1(I)$ promoter is expressed at comparable levels to pSV2CAT and pMoCAT, with promoter-enhancers from SV40 and Moloney murine leukemia virus, respectively. Sherwood and Bornstein (1990) transfected collagen-CAT constructs with 331 bp of human promoter sequences into different collagen expressing cell lines. They also reported similar levels of expression of this construct and pSV2CAT. In their hands the SV40 promoter-enhancer is about 2-fold stronger than the $\alpha 1(I)$ promoter.

Rippe *et al* (1989) furthermore demonstrated tissue-specificity of the 220 bp promoter. Expression of their collagen-CAT construct mimicked the expression of endogenous type I collagen in different cell lines. In NIH-3T3 cells the collagen-CAT expression is similar to pSV2CAT. However, in *src* transformed NIH-3T3 cells, where $\alpha 1(I)$ collagen expression is reduced ten fold, CAT expression is similarly reduced with respect to untransformed NIH-3T3 cells. In the case of MOPC315 cells, which synthesize no type I collagen, the collagen-CAT gene was not expressed at all.

However, though we and others could demonstrate that the 250 bp promoter of the pro- $\alpha 1(I)$ gene is sufficient for strong expression and contains in addition information for type I collagen specific expression (Rippe *et al*, 1989), other regions, upstream as well as downstream of the promoter are also involved in regulation of this gene. The experiments of Rossouw *et al* (1987) showed that in *Xenopus laevis* oocytes expression of a reporter gene increases as more sequences are added upstream of the promoter. In addition they identified an enhancer in the first intron. Other groups reported that the region upstream of the promoter, as well as the first intron contain both positive and negative regulatory elements (Rippe *et al*, 1989; Bornstein *et al*, 1987; Bornstein *et al*, 1988b; Sherwood and Bornstein, 1990).

Transfection into tissue-culture cells is an easy and fast way to test transcriptional activity of regulatory regions of genes and can give

an indication of tissue-specificity of such regions. However, the cells used for these experiments are usually immortalized cells which have lost some of their original characteristics due to continuous passaging. Another limitation of this technique is that many different cell lines have to be tested before any conclusive remarks can be made about tissue-specificity of regulatory regions. In order to overcome these limitations, some groups have tested regulatory regions of pro- $\alpha 1(I)$ and pro- $\alpha 2(I)$ collagen genes in transgenic mice (Khillan *et al*, 1986; Wu *et al*, 1990; Slack *et al*, 1991; Pavlin *et al*, 1992).

Slack *et al* (1991) generated transgenic mouse lines using $\alpha 1(I)$ collagen-human growth hormone fusion constructs. With 2.3 kb of $\alpha 1(I)$ 5' flanking sequences, the transgene was expressed at high levels and in a tissue-specific manner, regardless of whether or not the first intron was present. In most tissues the expression was similar to the endogenous $\alpha 1(I)$ collagen gene. Constructs with the promoter shortened to 440 bp were expressed 15- to 500-fold lower than with the long promoter and in this case the first intron was necessary for tissue specific expression. Without the intron, constructs with the 440 bp promoter were poorly expressed in all tissues except lung. These results confirm that, in addition to the minimal promoter, both 5' upstream sequences and sequences in the first intron are involved in regulating the expression of the pro- $\alpha 1(I)$ collagen gene.

4.3 Binding of nuclear factors to the pro- $\alpha 1(I)$ promoter

For a closer characterization of the $\alpha 1(I)$ promoter, we performed *in vitro* DNA-protein binding studies. Although the transfection experiments described in chapter 2 do not include a cell line which expresses no type I collagen, the high expression from the pro- $\alpha 1(I)$ promoter in NIH-3T3 cells, suggested that the promoter might be involved in the tissue-specific regulation of expression of the gene. This led us to compare nuclear extracts from expressing (fibroblasts) and non-expressing cells (lymphocytes) for the presence of nuclear factors binding to the promoter.

Electrophoretic mobility shift assays (bandshifts) with a fragment containing the $\alpha 1(I)$ promoter from -260 to +90 revealed differences in binding of nuclear factors from fibroblasts as compared to lymphocytes. Whereas two DNA-protein complexes (complexes I and II) were present when using either extract, there was one complex unique to the lymphocyte extract (complex III). In addition, the relative amounts of complexes I and II differed between the two cell lines. With fibroblast extracts they are present in similar amounts, while complex I is favored when using lymphocyte extracts. The three complexes could all be competed out with excess of unlabeled fragment indicating that they were formed by factors binding to the promoter fragment specifically.

This experiment revealed cell-specific binding of nuclear factors to the promoter but the question remained where on the promoter do nuclear factors bind and which sequence elements may be involved in tissue-specific regulation of the gene. Computer analyses of the DNA-sequence revealed the presence of sequence elements which are known to be recognized by specific transcription factors. These include the TATA-sequence which is recognized by the general transcription factor TFIID, two inverted CCAAT-boxes, as well as putative Sp1 and AP-2 binding sites. In addition, there is a pyrimidine-rich sequence with remarkable homology to a similar region in the first intron (Rossouw *et al*, 1987).

For a more detailed analysis of binding of nuclear factors to the promoter, DNAase I footprinting assays were performed. With nuclear extracts of fibroblasts five areas were protected (fig. 4.1). These include the distal CCAAT-box (CCAAT-F in fig. 4.1), the putative AP-2- and Sp1 sites (COL-A and COL-F respectively), and the pyrimidine rich area (PYR). In addition, strong protection was observed immediately upstream of the pyrimidine stretch (COL-B).

In the case of lymphocyte extracts the DNAase I pattern was markedly different over the pyrimidine rich area and over the duplicated CCAAT-boxes (fig. 4.1). Whereas the protection upstream of the pyrimidine-rich area (COL-B) and over the putative Sp1-site (COL-F) was similar to that of fibroblast extracts, the protection over the CCAAT-boxes was shifted downstream (CCAAT-L). Instead

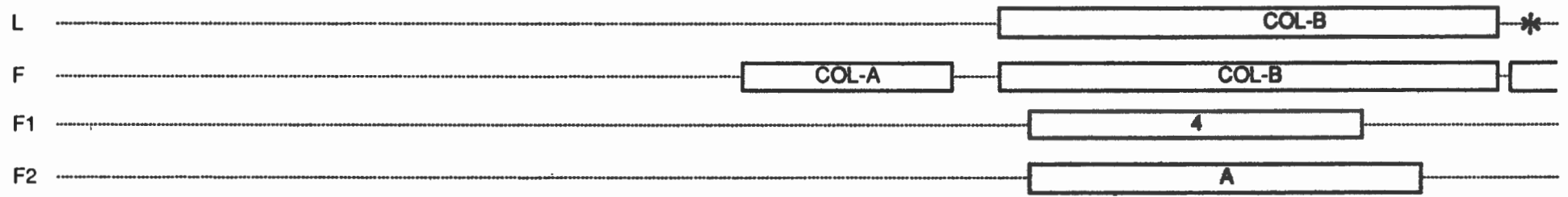
of the distal CCAAT-box (as for fibroblasts), it rather covered the area in between the two CCAAT-boxes and included the proximal CCAAT-motif. In addition, this protection was bordered by a DNAase I hypersensitive site that was induced upon incubation with lymphocyte nuclear extracts and coincided with the T residue of the distal CCAAT-box.

Significant differences were observed in the pyrimidine rich area. Whereas there was weak protection in this area when incubated with fibroblast extracts, several DNAase I hypersensitive sites were induced here in the case of lymphocyte extracts. Lastly, the fibroblast protection over the putative AP-2 site (COL-A) could not be observed with lymphocyte extracts.

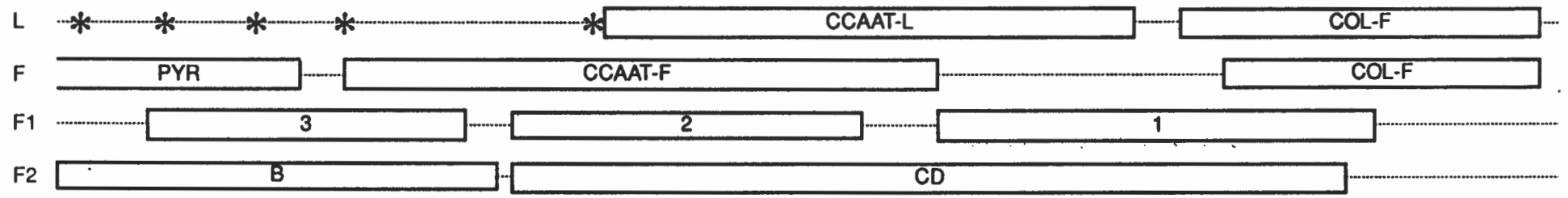
Figure 4.1: DNAase I protection of the 250 bp pro- α 1(I) promoter.

Comparison of the DNAase I protection pattern of the human pro- α 1(I) promoter obtained with nuclear extracts from lymphocytes (L) and fibroblasts (F) and of the murine pro- α 1(I) promoter with fibroblast extracts as published by Brenner et al (F1) and Karsentry and de Crombrughe (F2). Nucleotides where the mouse sequence (Mu) differs from that of the human sequence (Hu) are indicated below the human sequence. Numbers refer to nucleotide positions with the transcription initiation site taken as +1. The TATA-box and inverted CCAAT-boxes are boxed (solid lines) as well as the putative AP-2 and Sp1-sites of the human sequence (dotted lines). The pyrimidine-rich stretch is underlined. Protected regions are indicated by boxes and labeled as published (F1: footprints 1 to 4; F2: footprints A to D) or as in the text (in the case of L and F). PYR is the protection in the pyrimidine-rich stretch and CCAAT-L and CCAAT-F the protection over the CCAAT-boxes by nuclear factors from lymphocytes and fibroblasts respectively. Asterisks indicate DNAase I hypersensitive sites that were induced in the presence of nuclear extract.

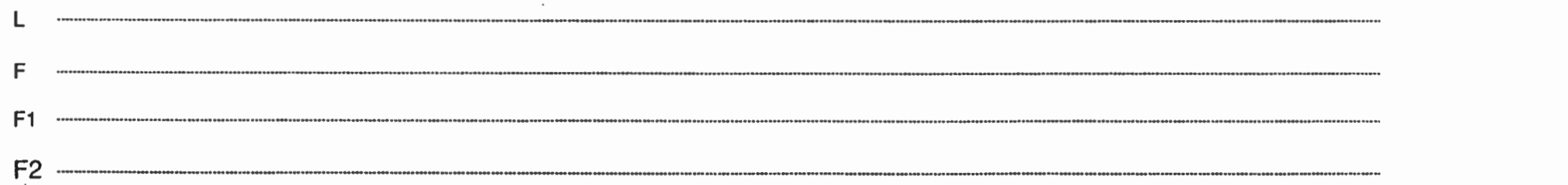
-260 CCCCTACCAC AGCACCTCTG GCCCAGGCTG GGCTGGGGGG CTGGGGAGGC AGAGCTGCGAA GAGGGGGAGA TGTGGGGTGG ACTCCCTTCC Hu
 T AGTG T TA AT AA C T -G G C CT T Mu



-170 CTCCTCCTCC CCCTCTCCAT TCCAACCTCCC AAATTGGGG CCGGGCCAGG CAGCTCTGAT TGGCTGGGGC ACGGGCGGCC GGCTCCCCCT Hu
 T TTC T C CC TTCG T T T G C T T Mu



-80 CTCCGAGGGG CAGGGTTCCT CCCTGCTCTC CATCAGGACA GTATAAAAAGG GGCCCGGGCC AGTCGTCGGA GCAGACGGGA +1 Hu
 A A A TG -A Mu



Two groups have published DNAase I footprinting analyses of the murine $\alpha 1(I)$ promoter using NIH-3T3 nuclear extracts (Brenner *et al.*, 1989; Karsentry and de Crombrughe, 1990). The DNAase I protection pattern as described by these two groups corresponds to the NIH-3T3 (fibroblast) pattern we observed (see fig. 4.1 for a comparison of the three studies). As in this study, using the human gene, there is protection at the col-B site, in the pyrimidine rich area and over the distal CCAAT-box.

There is however a difference in that the fourth protected area of the mouse gene covers the proximal CCAAT-box and is therefore shifted about ten nucleotides upstream of our col-F protection. This difference may be due to differences in the sequence of the mouse and human gene and will be discussed later in more detail. In addition, neither group describes the col-A protection. However, in both reports the col-A region was beyond the range of the DNAase I footprints shown.

The significance of the different areas of protection will be discussed separately.

4.3.1 Col-B

With nuclear extracts from both fibroblasts and lymphocytes, an area immediately upstream of the pyrimidine rich region was strongly protected against DNAase I digestion. The sequences in this area do not correspond to any known sequence element and may therefore be caused by binding of an as yet unidentified factor. In the first intron protection of the same region (i.e. immediately upstream of the pyrimidine stretch) was also observed (W.N de Vries, unpublished data). Closer inspection revealed a similarity in the sequences protected in the promoter and the first intron.

An oligonucleotide corresponding to this region (named the col-B site) was synthesized and tested in bandshift assays. Both fibroblast and lymphocyte extracts contain (a) factor(s) with the same electrophoretic mobility binding to this oligonucleotide. In addition there was another, faster migrating complex formed when using lymphocyte extracts. In southwestern blotting analyses the labeled

oligonucleotide was bound to several proteins. Some of these proteins bound to all the oligonucleotides used, suggesting that they correspond to non-specific binding. The apparent molecular weight of the remaining proteins bound to this oligonucleotide were 40, 48 and 80 kd in the case of HeLa extracts and 40 and 48 kd for lymphocytes. Competition experiments have shown that the 40 kd factor is specific for this sequence (W.N. de Vries, unpublished results).

The possibility remains that the 40 kd factor of fibroblasts and lymphocytes is not the same factor, but the similarity in the DNAase I protection of the col-B site leads us to suspect that the same factor is present in both cell lines and binds there. Attempts to clone the 40 kd factor binding to the col-B site are currently underway in our laboratory and will lead to further insight into the nature of this factor. In the southwestern blotting analyses we did not observe a specific factor present only in lymphocytes and not in fibroblasts as was observed in the electrophoretic mobility shift assays. This may be due to the southwestern blotting analyses being limited to detecting proteins dissociated into their different subunits and retaining binding activity after having been separated electrophoretically under denaturing conditions..

In the mouse pro- α 1(I) promoter the col-B site is also protected by (a) factor(s) present in NIH-3T3 fibroblasts (Brenner *et al*, 1989; Karsentry and de Crombrugge, 1990). These two groups also did transfection experiments in which this element was either mutated or deleted. Brenner *et al* could show that deleting their footprint 4 (the col-B site) didn't have an effect on transcriptional activity, suggesting that this element is not needed for expression of the gene.

Karsentry and de Crombrugge mutated the col-B site such that binding was abolished (binding to mutated oligonucleotides was tested in bandshift assays). Transfections using the α 1(I) promoter with this element thus mutated, resulted in an 1.7 fold increase of transcription. This led them to conclude that an inhibitory factor (IF1) binds here. In addition, they claimed that the factor binding to the col-B site and the pyrimidine rich element is the same factor.

4.3.2 pyrimidine rich region

In our hands there was weak DNAase I protection in the pyrimidine rich area with fibroblast nuclear extracts. However, with lymphocyte extracts, there was no protection but instead several strong DNAase I hypersensitive sites were induced. In bandshifts no differences could be observed between extracts from the two cell lines used, while in southwestern blotting analyses there is a fibroblast factor of 80 kd binding to the pyrimidine oligonucleotide which is not present in lymphocytes. The other factors binding to this oligonucleotide (40 kd and 48 kd) are present in both lymphocyte and fibroblast extracts.

Karsentry and de Crombrughe (1990) showed with bandshifts that the pyrimidine rich stretch and the col-B site of the mouse pro- α 1(I) promoter are recognized by the same factor. Since the col-B site is very purine-rich on the coding strand, it shows some similarities to the non-coding strand of the pyrimidine rich stretch. In our bandshift and southwestern blotting experiments we did indeed observe that proteins of the same electrophoretic mobility bind to these two elements. As in bandshifts of de Crombrughe's group, binding to the pyrimidine rich element was weaker than that of the col-B oligonucleotide. The same was true for the 40 and 80 kd factors binding to these oligonucleotides in the southwestern blots. The col-B oligonucleotide bound preferentially to these two factors and weakly to the 48 kd factor. The reverse was true for the pyrimidine element where the 48 kd band was more intense while the 40 and 80 kd bands were relatively weak.

This suggests that although the two oligonucleotides have sequence similarities and seem to be recognized by the same proteins, there are differences in the binding affinity of different factors to these elements. Therefore, it can not be excluded that the DNAase I footprints observed at the col-B site and the pyrimidine rich stretch were the result of two different proteins binding to these elements.

Karsentry and de Crombrughe (1990) showed that, as for the col-B site, mutations in the pyrimidine element abolishing binding causes an increase in transcriptional activity as tested in transfection

experiments. The 4 fold increase in expression is however higher than in the case of the Col-B site (1.7 fold). Constructs in which both elements are mutated behaved the same as those where only the pyrimidine element had been mutated. They concluded that the same factor (inhibitory factor 1 or IF1) binds to these two elements and inhibits transcription. However, Brenner *et al* (1989) demonstrated that deletion of both elements results in a 50% decrease in transfection, while deleting only the col-B site has no significant effect on transcription.

The apparent discrepancy of the data concerning importance of the pyrimidine rich element can be explained by the difference in binding affinity of factors binding to the col-B site and pyrimidine element. In the bandshifts of Karsentry and de Crombrughe both elements bind to (a) factor(s) migrating at the same electrophoretic speed. The complex formed with the pyrimidine element is however less pronounced than the one with the col-B element. This was the same in our own bandshifts and the southwestern blotting assay suggests that the factor that binds strongly to the col-B site and less strong to the pyrimidine rich element may be the 40 kd factor. However the pyrimidine element binds preferentially to another factor of apparent molecular weight of 48 kd.

Since the mutations that Karsentry and de Crombrughe chose were mutations that abolished binding to the factor observed in the bandshifts with either the col-B site or the pyrimidine element, it remains to be proven if these mutations also prevents binding of other factors. The results of Brenner *et al* (1989) suggest that the pyrimidine element is necessary for full activity of the $\alpha 1(I)$ promoter. It is therefore possible that the mutations of Karsentry and de Crombrughe (1990) do indeed prevent the binding of IF1, but not that of another positively acting protein. If the inhibitory factor and such a transcriptional activator would compete for binding to the same site, mutations abolishing the binding of the former would indeed lead to higher transcriptional activity as was observed.

The notion that the pyrimidine element is bound by a transcriptional activator rather than an inhibitory factor is

strengthened by the DNAase I footprint data showing no protection of this element when using lymphocyte extracts. In fact, several strong DNAase I hypersensitive sites were induced in this region indicating that in the case of lymphocyte nuclear extracts there is no factor binding this element.

Homopurine-homopyrimidine stretches are frequently located in the 5' flanking regions of many eukaryotic genes as well as sites involved in genetic recombination (Mirkin *et al*, 1987 and references therein). Several groups studying promoters with such purine- or pyrimidine rich elements have shown that (a) nuclear factor(s) binds to this region or that it is necessary for full transcriptional activity of the promoter (Davis *et al*, 1989; Gilmour *et al*, 1989; Hoffman *et al*, 1990; and Mavrothalassitis *et al*, 1990). The 29 bp pyrimidine stretch of the *c-Ki-ras* promoter is bound by one or more nuclear factors and when deleted transcription from this promoter decreases (Hoffman *et al*, 1990). Interestingly, analyzing the binding of factors to the 125 bp pyrimidine rich element in the promoter of the *c-myc* gene, revealed that different motifs occurring in this region are bound by either a nuclear factor or a ribonucleoprotein (Davis *et al*, 1989).

Homopurine-homopyrimidine sequences are often sites of S1 nuclease hypersensitivity suggesting that unusual, non-B DNA structures are formed by these sequences. Lee *et al* (1979) and Lyamichev *et al* (1985, 1986) have shown in *in vitro* experiments that homopurine-homopyrimidine stretches of DNA can assume a triple helical structure under low pH conditions or negative supercoiling. In this structure, named H-DNA, about half of the pyrimidine strand folds back on itself forming Hoogsteen base pairing with the purines. This results in a triple helical structure with half of the purine strand being present as single stranded DNA (reviewed by Wells *et al*, 1988).

Low pH conditions are necessary for protonation of the cytosine residues without which this structure can not be formed. Increase in the length of the homopyrimidine-homopurine stretches or a higher level of supercoiling stabilizes the H-DNA structure making it less pH dependent (Collier and Wells, 1990). In addition, the

formation of H-DNA can be achieved at a neutral pH when the DNA contains 5-methyl cytosine (Lee *et al*, 1984). Another form, *H-DNA where the triple helix consists of two purine strands and a pyrimidine strand, forms at neutral pH in the presence of zinc-ions (Bernuès *et al*, 1989). These results make the possibility that DNA can naturally assume such structures more feasible.

In experiments with plasmids containing poly(dG)-poly(dC) sequences of varying length, Kohwi and Kohwi-Shigematsu (1991) showed that there is a sharp transition between dG tracts forming H-DNA structures or not. In PBS supercoiled plasmids with dG tracts of 32 bp and longer form triplexes whereas those of 30 bp and less do not. Interestingly, a similar length dependency was observed with regard to these dG stretches' ability to augment transcription of a reporter gene when placed upstream of a promoter. In this case, dG tracts of 27 to 30 bp increase transcription of the CAT gene driven by the HSV thymidine kinase gene while longer dG stretches (33 bp or more) have no effect on transcriptional activity. Co-transfecting plasmids containing dG tracts of up to 30 bp (and no longer) inhibits this transcriptional augmentation. They postulate that a transcriptional activator binds to poly(dG) stretches of up to 30 bp. However, dG tracts of more than 30 bp form triple helical structures and thus prevents the transcriptional activator from binding to the DNA and activating transcription.

In this study we could show that NIH-3T3 fibroblasts contain (a) factor(s) that bind to the pyrimidine stretch of the pro- α 1(I) collagen promoter. In addition, Brenner *et al* (1989) demonstrated with deletions that this element is necessary for transcriptional activation. Bandshifts and southwestern blotting analyses show that lymphocyte and fibroblast extracts contain the same factors binding to an oligonucleotide of the pyrimidine rich region. However, there was no DNAase I protection in this region when using nuclear extracts of lymphocytes but instead several DNAase I hypersensitive sites were induced.

This DNAase I hypersensitivity is strongly reminiscent of the nuclease sensitivity observed in homopyrimidine-homopurine stretches caused by the DNA assuming an alternate structure. For

H-DNA the pyrimidine or purine stretch should contain a mirror repeat which is not the case for the pyrimidine element. Since the formation of H-DNA also requires superhelical tension, it is unlikely that we could detect this structure in our experiments using linearized DNA. However, Evans and Efstratiadis (1986) could demonstrate non-B, non-Z DNA structure in $(dG-dA)_n \cdot (dC-dT)_n$ which is independent of Hoogsteen G-CH⁺ base-pairing in both supercoiled and linear DNA.

These results suggest that the pyrimidine rich element may be involved in the tissue-specific activity of the pro- $\alpha 1(I)$ promoter. Fibroblasts (cells expressing type I collagen) contain a factor(s) binding to this region. The data from Brenner *et al* (1989) on the mouse pro- $\alpha 1(I)$ promoter and the reports of factors binding to similar regions in other promoters suggest that this factor(s) may activate transcription. In contrast, with lymphocyte extracts we observe increased DNAase I hypersensitivity suggesting the formation of unusual DNA structure. The presence of similar factors binding an oligonucleotide containing the pyrimidine element in both lymphocyte and fibroblast nuclear extracts suggests the possibility that the formation of an alternative DNA structure may prevent the binding of a transcriptional activator and thus lead to a decrease in transcription in cells which don't express type I collagen.

Another interesting possibility opened up after comparing the pyrimidine rich stretches present in the promoter and the first intron. Alignment of the two regions reveals that, given the right conditions, these two regions might hybridize to assume a H-DNA conformation (Fig 4.2). Such an interaction between the promoter and the intronic enhancer would of course bring the two important regulatory regions into close proximity of each other and could thus provide a mechanism through which the enhancer may exert its action. This would be in accordance with the looping model for enhancer action (for review see Müller and Schaffner, 1990) and would furthermore explain why the orientation of the enhancer has an effect on its efficiency - different orientation would result in a less perfect alignment.

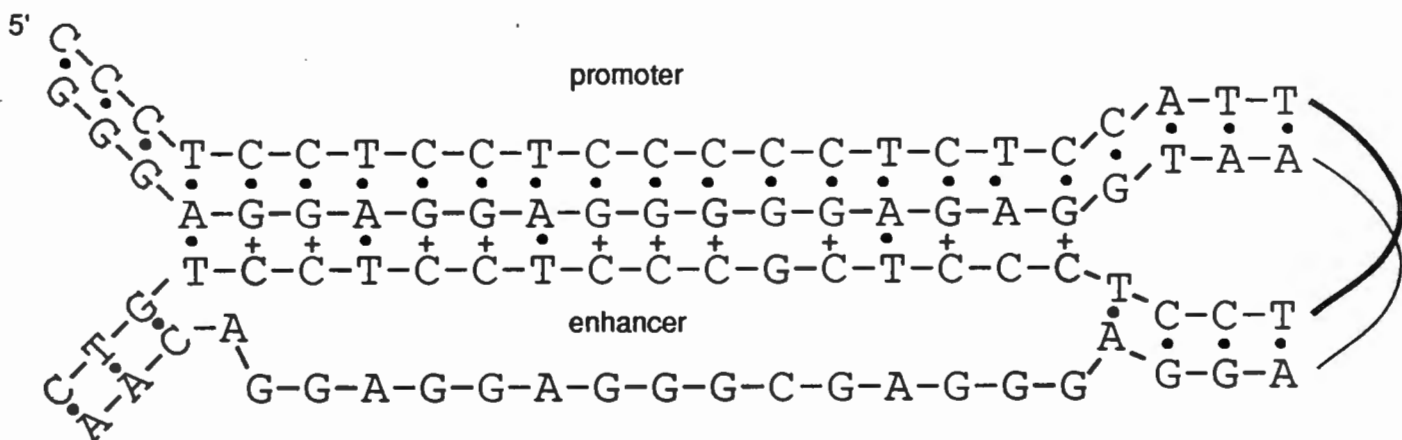


Figure 4.2: Alignment of the pyrimidine stretches of the pro- α 1(I) promoter and enhancer.

Aligning the pyrimidine elements of the promoter and intron reveals the possibility of H-DNA formation by Hoogsteen base pairing between the pyrimidine strand of the enhancer element and the purine strand of the promoter element.

However, the experiments described in chapter 3 as well as those characterizing the mouse pro- α 1(I) promoter (Brenner *et al*, 1989; Karsentry and de Crombrughe, 1990) were all performed with the promoter fragment in absence of the enhancer. It would therefore be interesting to test whether the DNA-protein interactions with the pyrimidine rich element would be different when linked in *cis* with the gene's enhancer. Alternatively, the pyrimidine rich element in the intron could react the same as the one in the promoter upon binding of nuclear factors to this region and its flanking sequences. Thus it could have a similar role in the regulation of the pro- α 1(I) gene as the corresponding promoter element.

4.3.3 CCAAT-boxes

The CCAAT-element is an essential element of many mRNA-coding genes (Breathnach & Chambon, 1981). Several different factors binding to this element and activating transcription have been identified like C/EBP (Graves *et al*, 1986), CTF/NF-1 (Jones *et al*, 1987), NF-Y and NF-Y* (Dorn *et al*, 1987) and CP-1 and CP-2

(Chodosh *et al*, 1988). The type I collagen genes also contain this element in their promoters and in the pro- α 1(I) promoter the CCAAT-box is duplicated and inverted (fig. 4.1).

Comparing the DNAase I protection pattern in the region of the CCAAT-boxes, reveals that not only are there differences in DNA-protein interactions between nuclear factors of fibroblasts and lymphocytes, but also between the human and mouse promoter (fig. 4.1). In our hands, fibroblast nuclear factors protect the distal CCAAT-box (CCAAT-F in fig. 4.1) while the proximal CCAAT-box remains unprotected. In both reports on the mouse pro- α 1(I) promoter (Brenner *et al*, 1989 and Karsentry and de Crombrughe, 1990) nuclear extracts from NIH-3T3 cells contain factors protecting both the proximal and distal CCAAT motifs. The protection of the CCAAT-boxes in the murine gene includes two 12 bp direct repeats occurring downstream of the CCAAT-boxes. However, the second of these repeats (downstream of the proximal CCAAT-box) is not fully conserved in the human promoter. This may explain the difference in protection of the proximal CCAAT-box between mouse and human genes.

Hatamochi *et al* (1988) reported the isolation of a CCAAT binding factor (CBF) present in NIH-3T3 and liver extracts, using an oligonucleotide containing the CCAAT-box of the pro- α 2(I) collagen promoter. This factor was shown to bind also to the proximal but not the distal CCAAT-box of the murine pro- α 1(I) promoter (Maity *et al*, 1988). They demonstrated furthermore that addition of CBF to a depleted NIH-3T3 nuclear extract can increase transcription from the mouse pro- α 1(I) promoter threefold. The extracts were depleted for CBF by mixing it with the DNA affinity resin used to purify the factor.

Upon incubation of the murine pro- α 1(I) promoter with fibroblast nuclear extract, Karsentry and de Crombrughe (1990) observed a DNAase I footprint covering both CCAAT-boxes together with their adjacent 12 bp repeats. However, in bandshift assays using mutated and non-mutated oligonucleotides as competitors, they could demonstrate that this extended region is bound by at least two different factors. The footprint CD could thus be divided into two

parts similar to footprints 1 and 2 of Brenner *et al* (1989) (see fig. 4.1).

Since CBF (the $\alpha 2(I)$ CCAAT binding factor) can bind to the proximal CCAAT-box of the $\alpha 1(I)$ promoter, they concluded that binding to footprint D (the proximal CCAAT-box) is due to CBF. In support of this theory, they showed that an oligonucleotide of footprint D is bound by a factor with the same electrophoretic mobility as CBF. That the proximal CCAAT-box is important was indicated by the 4 fold decrease in transcription when mutating this motif. Mutations in the distal 12 bp repeat caused a 10 fold increase in transcription, leading them to suggest that an inhibitory factor (IF2) binds to footprint C (the distal CCAAT-box and adjacent 12 bp repeat).

Footprints 1 and 2 observed by Brenner *et al* (1989) were further characterized by the same group (Nehls *et al*, 1991). In both bandshift assays and DNAase I footprints oligonucleotides of footprint 1 and 2 could compete with each other for binding of nuclear factors. Mutating either CCAAT-box changed the DNAase I footprint pattern and in transfections these mutations resulted in a two to three fold decrease in transcription. In contrast, mutating the distal 12 bp repeat had no effect on transcription, though this mutation abolished binding to footprint 2.

Nehls *et al* (1991) showed convincingly that both footprint 1 and 2 are bound by the CCAAT-binding factor NF-1 and Sp1, with NF-1 binding predominantly to the proximal CCAAT-box and Sp1 to the distal 12 bp repeat. In transfection experiments they could demonstrate that overexpression of NF-1 leads to an increase of transcription of a gene driven by the pro- $\alpha 1(I)$ promoter, while overexpression of Sp1 inhibits transcription. Co-transfecting expression vectors of both NF-1 and Sp1 revealed that overexpression of Sp1 can inhibit the increase of transcription caused by NF-1. The binding sites of NF-1 and Sp1 are overlapping and this result can therefore be explained by mutually exclusive binding of the two factors. Since the 12 bp repeat contains an imperfect Sp1 consensus sequence, GGGCCGGG in stead of GGGCGGG, the binding affinity of Sp1 to this site would be low, favoring thus

the binding of NF-1. Competition between Sp1 and NF-1 for binding to overlapping sites would not only explain the inhibitory effect of overexpression of Sp1 but also the increase in transcription upon mutating the distal 12 bp repeat reported by Karsentry and de Crombrughe (1990). Since these mutations abolish the binding of Sp1, there is no hindrance for NF-1 binding and activating transcription.

In competition experiments with oligonucleotides containing the consensus sequence of different CCAAT binding factors, Nehls *et al* (1991) could demonstrate that the two protected regions over the CCAAT-boxes are not bound by CBF, CP1, CP2 or C/EBP. This is in contrast to the publication of Karsentry and de Crombrughe (1990) where they attributed the footprint over the proximal CCAAT-box to CBF. However, although CBF can bind to the proximal CCAAT motif of the pro- α 1(I) promoter, they provided no proof that this is indeed the fibroblast factor protecting this region from DNAase I digestion. The above mentioned experiments of Hatamochi *et al* (1988) with CBF depleted NIH-3T3 nuclear extracts are not conclusive either. The oligonucleotide they used for depleting the extracts, contained a CCAAT-box and could therefore retain also other factors recognizing the CCAAT element. If the transactivator normally binding to the pro- α 1(I) CCAAT-boxes would thus be absent or greatly reduced, the proximal CCAAT-box would be available for binding of CBF. This would explain the enhancement of transcription from the pro- α 1(I) promoter upon addition of purified CBF.

The proximal 12 bp repeat is not conserved in the human pro- α 1(I) promoter. Instead of the imperfect Sp1 consensus sequence present in the distal repeat and in both repeats of the murine gene, there is actually a perfect Sp1 consensus sequence downstream of the proximal CCAAT-box of the human gene (fig. 4.1). With extracts of both fibroblasts and lymphocytes this Sp1 site is protected against DNAase I digestion while the proximal CCAAT-box remains unprotected. This is in accordance with the theory of Nehls *et al* (1991) that binding of Sp1 prevents the binding of NF-1 to the overlapping NF-1 site.

Competition of the DNAase I footprints of Nehls *et al* suggested that NF-1 binds predominantly to the proximal CCAAT motif and Sp1 to the distal 12 bp repeat. However, their transfection data with mutations in the promoter show that while both CCAAT-boxes are necessary for full transcriptional activity, mutations abolishing the Sp1 footprint have no effect on transcription. In addition, mutating either CCAAT-box resulted in a change of the DNAase I footprinting pattern. This, together with the fact that the sequence and bandshift binding pattern of footprint 1 and 2 is the same, suggests that both footprints are caused by NF-1 rather than Sp1. It would then follow from this that the fibroblast protection we observe over the distal CCAAT-box is caused by the binding of NF-1. However, it remains to be proven that it is indeed the binding of NF-1 we observe and that this binding is necessary for full transcriptional activity of the human promoter.

The data of Nehls *et al* (1991) and Karsentry and deCrombrughe (1990) indicate that the CCAAT-boxes of the pro- α 1(I) gene are important for full transcriptional activity. Therefore it was particularly interesting that we could observe differences in the binding of nuclear factors from collagen expressing or non-expressing cells in this region. Both bandshifts and southwestern blots revealed differences in the factors present in lymphocyte or fibroblast extracts binding to the two CCAAT containing oligonucleotides. In addition, the DNAase I protection pattern differed when using the respective extracts. With fibroblasts (expressing cells), protection centered on the distal CCAAT-box while protection with lymphocytes (non-expressing cells) was shifted to include both CCAAT-boxes. In particular, a strong DNAase I hypersensitive site was induced in the region protected by fibroblast extracts, coinciding with the T residue of the distal CCAAT element (fig. 4.1).

The presence of the duplicated, inverted CCAAT-boxes and the DNAase I protection when using extracts from expressing and non-expressing cells, is reminiscent of the report by Barberis *et al* (1987) on the sperm H2B-1 histone gene of the sea urchin. They could show that non-expressing cells contain a dominant factor which can, through mutually exclusive binding, prevent the binding

of a CCAAT binding factor. The vertebrate homolog of this factor, named the CCAAT displacement protein (CDP), binds to at least two other genes with duplicated CCAAT-boxes in their promoters. These are the A γ -globin gene (Superti-Furga *et al*, 1988) and the gp91-phox gene (Skalnik *et al*, 1991).

Methylation interference studies and bandshifts with mutated oligonucleotides showed that CDP recognizes the two CCAAT-boxes and a recurring TGA trinucleotide in the promoter of the A γ -globin gene. As in the case of the latter, this trinucleotide occurs three times in the promoters of the sea urchin's sperm H2B-1 histone gene as well as the human gp91-phox gene. Inspection of the sequence of the pro- α 1(I) collagen promoter reveals that though the spacing of the two CCAAT-boxes is similar to that of the histone H2B and the γ -globin gene, the TGA trinucleotide occurs only once with the A being the A of the proximal inverted CCAAT-box. However, though methylation interference have shown that CDP interacts with the G of the TGA trinucleotide, mutating this G in all three TGA motifs of the γ -globin gene didn't completely abolish binding of CDP (Superti-Furga *et al*, 1988). This indicates that the three trinucleotides are not necessary for binding of CDP.

The DNAase I protection by extracts of expressing and non-expressing cells lead us to suspect that the same factor could be binding to the pro- α 1(I) promoter, acting as a repressor. Clearly two different factors are binding to the CCAAT-boxes when using nuclear extracts from lymphocytes or fibroblasts. The region protected by lymphocyte extracts overlaps with the fibroblast protection. The 5' boundary of the lymphocyte binding is in the distal CCAAT-box, and stretches past the proximal CCAAT-box. This is similar to the mutually exclusive binding of CDP and CCAAT-box binding factors of the above mentioned promoters with duplicated CCAAT motifs. Superti-Furga *et al* (1989) showed that the lymphocyte cell line BJA-B does indeed contain CDP binding to the γ -globin promoter.

The reports on the mouse α 1(I) promoter suggest that in collagen expressing cells a transcriptional activator binds to one or both of the CCAAT-boxes. The binding of CDP to the promoter in the case of

non-expressing cells could be a mechanism through which the expression from this promoter is prevented or decreased. In the case of the gp91-phox promoter, it was found that deleting the CDP/CP1 binding region can relieve repression of transcription in non-expressing cells. This derepression is independent of the CP1 (CCAAT binding protein) binding site, suggesting that in addition to blocking binding of CP1, CDP may also repress transcription by another mechanism (Skalnik *et al*, 1991).

Since the oligonucleotides used for bandshifts and southwestern do not include an oligonucleotide containing both CCAAT-boxes, the binding of CDP could not be observed in these experiments. It remains to be proven, therefore whether CDP does indeed bind to the pro- α 1(I) collagen promoter and whether this is involved in the tissue-specific regulation of the gene.

4.3.4 Other elements

The putative AP-2 site at position -203 is weakly protected with fibroblast but not with lymphocyte extracts. This area was out of range in the experiments done on the mouse promoter and was therefore not observed in the murine pro- α 1(I) promoter studies (Brenner *et al*, 1989 and Karsentry and de Crombrughe, 1990). Since the sequence of this protected area is not fully conserved between mouse and human, it would be interesting to know whether there is also a factor binding here to the mouse promoter. DNAase I footprints using purified AP-2 showed that this factor can indeed bind to the site at -203 (W.P. Vergeer, unpublished data). Since AP-2 is highly expressed in NIH-3T3 cells (Mitchell *et al*, 1991), we are tempted to conclude that it is indeed AP-2 binding here. However, the col-A protection was not quite centered on the putative AP-2 site and it remains to be tested by way of competition experiments with an AP-2 binding site oligonucleotide whether the factor binding here is indeed AP-2.

The GGGCGG-box which is the core recognition sequence for the transcriptional activator Sp1 (Gidoni *et al*, 1984) is protected in DNAase I footprints when using either fibroblast or lymphocyte nuclear extracts. As already mentioned in the discussion of the

binding of factors to the region of the CCAAT-boxes, it is likely that this site is bound by Sp1. For the mouse promoter it was shown that Sp1 can bind to oligonucleotides of footprints 1 and 2 (see fig. 4.1; Nehls *et al*, 1991). However, since the mouse promoter doesn't contain a consensus Sp1 binding site, and the possible sites are overlapping with NF-1 sites, it is not sure whether Sp1 binds to this promoter *in vivo*. In the case of the human gene there is actually a consensus Sp1 site and it is likely that the protection we observe here is caused by Sp1. As already mentioned before, occupation of the Sp1 site may be the reason for absence of binding to the proximal CCAAT-box. The Sp1 site was contained in the proximal CCAAT-oligonucleotide (oligonucleotide D). However, the southwestern blots with this oligonucleotide wasn't bound by the 95 kd or 105 kd isoforms of Sp1. Whether any of the bands observed in the bandshift with this oligonucleotide corresponds to binding of Sp1, will have to be verified through competition experiments with an Sp1 consensus binding site.

4.4 Conclusion

The following conclusions can be made from this study:

The pro- α 1(I) collagen promoter is a strong promoter, comparable to the HSV thymidine kinase promoter and the SV40 enhancer-promoter.

A factor present in nuclear extracts of both fibroblasts and lymphocytes binds strongly to the promoter at a sequence which does not correspond to the consensus sequence of any known transcription factor. This factor, named col-B, has an apparent molecular weight of about 40 kd.

Differences between DNA-protein interactions with nuclear extracts of collagen expressing or non-expressing cells could be observed in three regions, namely the pyrimidine rich element, around the duplicated CCAAT-boxes and the putative AP-2 site. This suggests that these regions may play a role in the tissue-specific regulation of this gene.

DNAase I protection assays indicate that whereas the pyrimidine rich element is bound by a factor present in fibroblast nuclear extracts, the same region is not bound by any factor from lymphocytes. Instead, the induction of DNAase I hypersensitive sites here is suggestive of a change to a non-B DNA conformation in the presence of lymphocyte nuclear extracts.

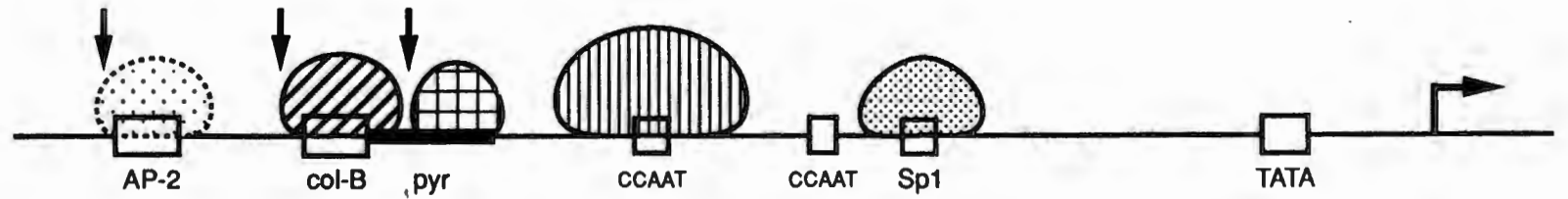
Different factors bind to the CCAAT-boxes when incubating with extracts from expressing or non-expressing cells. Experiments done on the mouse pro- $\alpha 1(I)$ promoter suggest that NF-1 binds to one or both of the CCAAT-boxes and activates transcription of the gene (Nehls *et al.*, 1991). The DNAase I protection pattern and sequence of the duplicated CCAAT-box region is reminiscent of the case of the CCAAT displacement protein (CDP) which represses transcription by competing for binding with transcriptional activators recognizing the CCAAT-box.

A summary of the binding of nuclear factors from expressing and non-expressing cells to the pro- $\alpha 1(I)$ collagen promoter is presented in fig. 4.3. Our experiments, together with data published by various other groups, lead us to suspect that the fibroblast factors binding to the pyrimidine element and the distal CCAAT-box are activating transcription in expressing cells. The DNA-protein interactions with these two elements in the presence of lymphocyte extracts, present two possible mechanisms for down-regulating the activity of the pro- $\alpha 1(I)$ collagen gene in non-expressing cells. In the case of the pyrimidine element, the formation of an alternative DNA conformation may prevent the binding of (a) positively acting factor(s). In the CCAAT-box region the binding of a potential trans-activating factor might be prevented by the binding of CDP or another factor with a similar function.

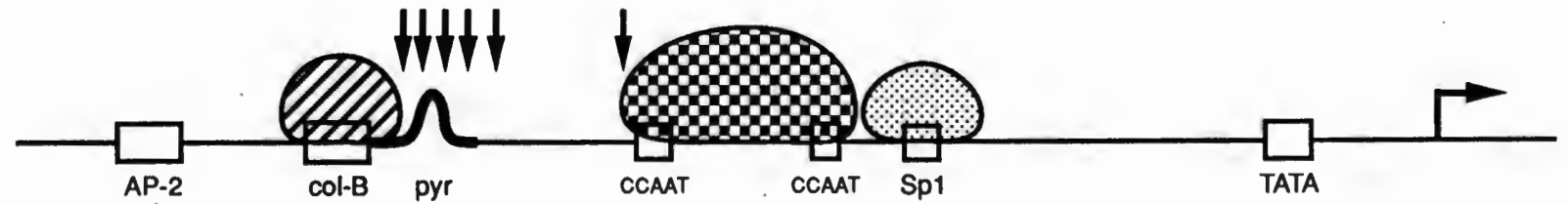
Figure 4.3: Proposed model for regulation of pro- α 1(I) collagen gene through binding of nuclear factors to the promoter.

Summary of in vitro binding studies of proteins to the promoter. Symbols: Boxes, binding sites for transcription factors; thick line, pyrimidine-rich region; bent arrow, transcription initiation site; half circles, binding of factors as observed in DNAase I footprints; vertical arrows, DNAase I hypersensitive sites induced by binding of nuclear factors; bend thick line, possible DNA conformation change as indicated by induction of DNAase I hypersensitive sites.

Fibroblasts
(expressing cells)



Lymphocytes
(non-expressing cells)



LIST OF ABBREVIATIONS

bp	base pairs
CAT	chloramphenicol acetyl transferase
CBF	CCAAT binding factor
CDP	CCAAT displacement factor
dATP	deoxyadenosine triphosphate
dC	deoxycytidine
dCTP	deoxycytosine triphosphate
dI	deoxyinosine
DMEM	Dulbecco's modified Eagle's medium
DMS	dimethylsulfate
DNA	deoxyribonucleic acid
DTT	1,4-dithio-DL-threitol
EDTA	ethylenediaminetetraacetic acid
Hepes	4-(2-hydroxyethyl)-piperazine-1-ethane-sulfonic acid
hsp	heat shock protein
HSV	Herpes simplex virus
Ig	immunoglobulin
Inr	initiator
kb	kilobase pairs
kd	kilodalton
KU	Kunitz units
mRNA	messenger ribonucleic acid
MT-IIa	metallothionein IIa
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PMSF	phenylmethyl sulfonylfluoride

Py	pyrimidine
R	resistant
RNA	ribonucleic acid
r.p.m.	revolutions per minute
SDS	sodium dodecyl sulfate
SV40	simian virus 40
TGF- β	transforming growth factor β
TM	trade mark
tris	tris(hydroxymethyl)aminomethane

REFERENCES

Akerblom, I.E., Slater, E.P., Beato, M., Baxter, J.D. and Mellon, P.L. (1988). Negative regulation by glucocorticoids through interference with a cAMP responsive enhancer. *Science* 241, 350-353.

Barberis, A., Superti-Furga, G. and Busslinger, M. (1987). Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell* 50, 347-359.

Barsh, G.S., Roush, C.L. and Gelinas, R.E. (1984). DNA and chromatin structure of the human $\alpha 1(I)$ collagen gene. *J. Biol. Chem.* 259, 14906-14913.

Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L. and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* 61, 49-59.

Bennett, V.D. and Adams, S.L. (1987). Characterization of the translational control mechanism preventing synthesis of $\alpha 2(I)$ collagen in chicken vertebral chondroblasts. *J. Biol. Chem.* 262, 14806-14814.

Benoist, C. and Chambon, P. (1980). Deletions covering the putative promoter region of early mRNAs of simian virus 40 do not abolish T-antigen expression. *Proc. Natl. Acad. Sci. USA* 77, 3865-3869.

Bernués, J., Beltrán, R., Casasnovas, J.M. and Azorín, F. (1989). Structural polymorphism of homopurine-homopyrimidine sequences: the secondary DNA structure adopted by a d(GA.CT)₂₂ sequence in the presence of zinc ions. *EMBO J.* 8, 2087-2094.

Blank, V., Kourilsky, P. and Israël, A. (1992). NF- κ B and related proteins: Rel/dorsal homologues meet ankyrin-like repeats. *Trends Biochem. Sci.* 17, 135-140.

Bornstein, P., McKay, J., Morishima, K., Devarayalu, S. and Gelinas, R.E. (1987). Regulatory elements in the first intron contribute to transcriptional control of the human $\alpha 1(I)$ collagen gene. *Proc. Natl. Acad. Sci. USA* 84, 8869-8873.

Bornstein, P., McKay, J., Devarayalu, S. and Cook, S.C. (1988a). A highly conserved, 5' untranslated, inverted repeat sequence is ineffective in translational control of the $\alpha 1(I)$ collagen gene. *Nucl. Acids Res.* 16, 9721-9736.

Bornstein, P., McKay, J., Liska, D.J., Apone, S. and Devaralayalu, S. (1988b). Interactions between the promoter and first intron are involved in transcriptional control of $\alpha 1(I)$ collagen gene expression. *Mol. Cell. Biol.* 8, 4851-4857.

Bornstein, P. and Sage, H. (1989). Regulation of collagen gene expression. *Progress in nucleic acid research and molecular biology* 37, 70-105.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50, 349-383.

Brenner, D.A., Rippe, R.A. and Veloz, L. (1989). Analysis of the collagen $\alpha 1(I)$ promoter. *Nucl. Acids Res.* 17, 6055-6064.

Busch, S.J. and Sassone-Corsi, P. (1990). Dimers, leucine zippers and DNA-binding domains. *Trends Genet.* 6, 36-40.

Carcamo, J., Buckbinder, L. and Reinberg, D. (1991). The initiator directs the assembly of a transcription factor IID-dependent transcription complex. *Proc. Natl. Acad. Sci. USA* 88, 8052-8056.

Concino, M.F., Lee, R.F., Merryweather, J.P. and Weinmann, R. (1984). The adenovirus major late promoter TATA box and initiation site are both necessary for transcription *in vitro*. *Nucl. Acids Res.* 12, 7423-7433.

Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988). Human CCAAT-binding proteins have heterologous subunits. *Cell* 53, 11-24.

Chu, M-L., De Wet, W., Bernard, M., Ding, J-F., Morabito, M., Myers, J., Williams, C. and Ramirez, F. (1984). Human pro α 1(I) collagen gene structure reveals evolutionary conservation of introns and exons. *Nature* 310, 337-340. ✓

Chu, M-L., De Wet, W., Bernard, M. and Ramirez, F. (1985). Fine structural analysis of the human pro- α 1(I) collagen gene. *J. Biol. Chem.* 260, 2315-2320.

Collier, D.A. and Wells, R.D. (1990). Effect of length, supercoiling, and pH on intramolecular triplex formation. *J. Biol. Chem.* 265, 10652-10658.

Davis, T.L., Firulli, A.B. and Kinniburgh, A.J. (1989). Ribonucleoprotein and protein factors bind to an H-DNA-forming *c-myc* DNA element: possible regulators of the *c-myc* gene. *Proc. Natl. Acad. Sci. USA* 86, 9682-9686. ✓

DeNoto, F.M., Moore, D.D. and Goodman, H.M. (1981). Human growth hormone DNA sequence and mRNA structure: possible alternative splicing. *Nucl. Acids Res.* 9, 3719-3730.

De Wet, W.J., Chu, M-L. and Prockop, D.J. (1983) The mRNAs for the pro- α 1(I) and pro- α 2(I) chains of type I procollagen are translated at the same rate in normal human fibroblasts and in fibroblasts from two variants of Osteogenesis Imperfecta with altered steady state ratios of the two mRNAs. *J. Biol. Chem.* 258, 14385-14389.

Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucl. Acids Res.* 11, 1475-1489.

Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987). A multiplicity of CCAAT box-binding proteins. *Cell* 50, 863-872. ✓

Dynan, W.S. and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* 316, 774-778.

Dynlacht, B.D., Hoey, T. and Tjian, R. (1991). Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* 66, 563-576.

Evans, T. and Efstratiadis, A. (1986). Sequence-dependent S1 nuclease hypersensitivity of a heteronomous DNA duplex. *J. Biol. Chem.* 261, 14771-14780.

Evans, R.M. and Hollenberg, S.M. (1988). Zinc fingers: Gilt by association. *Cell* 52, 1-3.

Finer, M.H., Aho, S., Gerstenfeld, L.C., Boedtker, H. and Doty, P. (1985). Unusual DNA sequences located within the promoter region and the first intron of the chicken pro- α 1(I) collagen gene. *J. Biol. Chem.* 262, 13323-13332.

Focht, R.J. and Adams, S.L. (1984). Tissue specificity of type I collagen gene expression is determined at both transcriptional and post-transcriptional levels. *Mol. Cell. Biol.* 4, 1843-1852.

Galas, D.J. and Schmitz, A. (1978). DNAase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucl. Acids Res.* 5, 3156-3170.

Gidoni, D., Dynan, W.S. and Tjian, R. (1984). Multiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature*, 312, 409-413.

Gilmour, D.S., Thomas, G.H. and Elgin, S.C.R. (1989). *Drosophila* nuclear proteins bind to regions of alternating C and T residues in gene promoters. *Science* 245, 1487-1490.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* 23, 175-182.

Godowski, P.J., Rusconi, S., Miesfeld, R. and Yamamoto, K.R. (1987). Glucocorticoid receptor mutants that are constitutive activators of transcriptional enhancement. *Nature* 325, 365-368.

Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2, 1044-1051.

Graves, B.J., Johnson, P.F. and McKnight, S.L. (1986). Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. *Cell* 44, 565-576.

Hahn, S., Buratowski, S., Sharp, P.A. and Guarente, L. (1989). Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc. Natl. Acad. Sci. USA* 86, 5718-5722.

Halazonetis, T.D., Georgopoulos, K., Greenberg, M.E. and Leder, P. (1988). c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities. *Cell* 55, 917-924.

Harbers, K., Kuehn, M., Delius, H. and Jaenisch, R. (1984) Insertion of retrovirus into the first intron of $\alpha 1(I)$ collagen gene leads to embryonic lethal mutation in mice. *Proc. Natl. Acad. Sci. USA* 81, 1504-1508.

Hatamochi, A., Golumbek, P.T., Van Schaftingen, E. and De Crombrughe, B. (1988). A CCAAT DNA binding factor consisting of two different components that are both required for DNA binding. *J. Biol. Chem.* 263, 5940-5947.

Herskowitz, I. (1989). A regulatory hierarchy for cell specialization in yeast. *Nature* 342, 749-757.

Hoffman, E.K., Trusko, S.P., Murphy, M. and George, D.L. (1990). An S1 nuclease-sensitive homopurine/homopyrimidine domain in the c-Ki-ras promoter interacts with a nuclear factor. *Proc. Natl. Acad. Sci. USA* 87, 2705-2709.

Horikoshi, M., Hai, T., Lin, Y-S., Green, M.R. and Roeder, R.G. (1988). Transcription factor ATF interacts with the TATA factor to facilitate establishment of preinitiation complex. *Cell* 54, 1033-1042.

Horton, W., Miyashita, T., Kohno, K., Hassell, J.R. and Yamada, Y. (1987). Identification of a phenotype specific enhancer in the first intron of the rat collagen II gene. *Proc. Natl. Acad. Sci. USA* 84, 8864-8868.

Hübscher, U. (1987). Double replica southwestern. *Nucl. Acids Res.* 15, 5486.

Johnson, P.F. and McKnight, S.L. (1989). Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* 58, 799-839.

Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. and Tjian, R. (1987). A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* 48, 79-89.

Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986). Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends Biochem. Sci.* 11, 20-23.

Karin, M., Cathala, G. and Nguyen-Huu, M.C. (1983). Expression and regulation of a human metallothionein gene carried on a autonomously replicating shuttle vector. *Proc. Natl. Acad. Sci. USA* 80, 4040-4044.

Karsenty, G. and de Crombrughe, B. (1990). Two different negative and one positive regulatory factors interact with a short promoter segment of the $\alpha 1(I)$ collagen gene. *J. Biol. Chem.* 17, 9934-9942.

Khillan, J.S., Schmidt, A., Overbeek, P.A., De Crombrughe, B. and Westphal, H. (1986). Developmental and tissue-specific expression directed by the $\alpha 2$ type I collagen promoter in transgenic mice. *Proc. Natl. Acad. Sci. USA* 83, 725-729. ✓

Kohwi, Y. and Kohwi-Shigematsu, T. (1991). Altered gene expression correlates with DNA structure. *Genes Dev.* 5, 2547-2554.

Laemlli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988). The leucine zipper: a Hypothetical structure common to a new class of DNA binding proteins. *Science* 240, 1759-1764.

Lawn, R.M., Efstradiatis, A., O'Connell, C. and Maniatis, T. (1980). The nucleotide sequence of the human β -globin gene. *Cell* 21, 647-651.

Lee, J.S., Johnson, D.A. and Morgan, A.R. (1979). Complexes formed by (pyrimidine)_n.(purine)_n DNAs on lowering the pH are three-stranded. *Nucl. Acids Res.* 6, 3073-3091.

Lee, J.S., Woodsworth, M.L., Latimer, L.J.P. and Morgan, A.R. (1984). Poly(pyrimidine).poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucl. Acids Res.* 12, 6603-6614.

Lewin, B. (1990). Commitment and activation at pol II promoters: a tail of protein-protein interactions. *Cell* 61, 1161-1164.

Levine, M. and Hoey, T. (1988). Homeobox proteins as sequence-specific transcription factors. *Cell* 55, 537-540.

Levine, M. and Manley, J.L. (1989). Transcriptional repression of eukaryotic promoters. *Cell* 59, 405-408.

Lin, Y-S. and Green, M.R. (1991). Mechanism of action of an acidic transcriptional activator *in vitro*. *Cell* 64, 971-981.

Liska, D.J., Slack, J.L. and Bornstein, P. (1990). A highly conserved intronic sequence is involved in transcriptional regulation of the $\alpha 1(I)$ collagen gene. *Cell Regulation* 1, 487-498.

- Liska, D.J., Robinson, V.R. and Bornstein, P. (1992). Elements in the first intron of the $\alpha 1(I)$ collagen gene interact with Sp1 to regulate gene expression. *Gene Expression* 2, 379-389.
- Lowy, D.R., Rands, E. and Scolnick, E.M. (1978). Helper-independent transformation by unintegrated Harvey sarcoma virus DNA. *J. Virol.* 26, 291-298.
- Lyamichev, V.I., Mirkin, S.M. and Frank-Kamenetskii, M.D. (1985). A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. *J. Biomol. Struc. Dyn.* 3, 327-338.
- Lyamichev, V.I., Mirkin, S.M. and Frank-Kamenetskii, M.D. (1986). Structures of homopurine-homopyrimidine tract in superhelical DNA. *J. Biomol. Struc. Dyn.* 3, 667-669.
- Maity, S.N., Golumbek, P.T., Karsenty, G. and De Crombrughe, B. (1988). Selective activation of transcription by a novel CCAAT binding factor. *Science* 241, 582-585. ✓
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
- Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987). Regulation of inducible and tissue-specific gene expression. *Science* 236, 1237-1245.
- Manor, H., Rao, B.S. and Martin, R.G. (1988). Abundance and degree of dispersion of genomic $d(GA)_n.d(TC)_n$ sequences. *J. Mol. Evol.* 27, 96-101.
- Mavrothalassitis, G.J., Watson, D.K. and Papas, T.S. (1990). Molecular and functional characterization of the promoter of *ETS2*, the human *c-ets-2* gene. *Proc. Natl. Acad. Sci. USA* 87, 1047-1051. ✓
- Maxam, A.M. and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Meth. Enzymol.* 65, 499-560.

Meisterernst, M., Roy, A.L., Lieu, H.M. and Roeder, R.G. (1991). Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. *Cell* 66, 981-993.

Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A. and Frank-Kamenetskii, M.D. (1987). DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature* 330, 495-497.

Mitchell, P.J., Wang, C. and Tjian, R. (1987). Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. *Cell* 50, 847-861.

Mitchell, P.J. and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245, 371-378.

Mitchell, P.J., Timmons, P.M., Hèbert, J.M., Rigby, P.W.J. and Tjian, R. (1991). Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. *Genes Dev.* 5, 105-119.

Müller, M.M., Gerster, T. and Schaffner, W. (1988) Enhancer sequences and the regulation of gene transcription. *Eur. J. Biochem.* 176, 485-495.

Müller, H-P. and Schaffner, W. (1990) The enhancer mechanism - *cis*-elements can act in *trans*. *Trends Genet.* 6, 300-304.

Nakajima, N., Horikoshi, M. and Roeder, R.G. (1988). Factors involved in specific transcription by mammalian RNA polymerase II: Purification, genetic specificity, and TATA box-promoter interactions of TFIID. *Mol. Cell. Biol.* 8, 4028-4040.

Nehls, M.C., Rippe, R.A., Veloz, L. and Brenner, D.A. (1991). Transcription factors Nuclear factor I and Sp1 interact with the murine collagen $\alpha 1(I)$ promoter. *Mol. Cell. Biol.* 11, 4065-4073.

- Ohlson, H. and Edlund, T. (1986). Sequence-specific interactions of nuclear factors with the insulin gene enhancer. *Cell* 45, 35-44.
- Parkhurst, S.M., Bopp, D. and Ish-Horowicz, D. (1990). X:A ratio, the primary sex-determining signal in *Drosophila*, is transduced by helix-loop-helix proteins. *Cell* 63, 1179-1191.
- Pavlin, D., Lichtler, A.C., Bedalov, A., Kream, B.E., Harrison, J.R., Thomas, H.F., Gronowicz, G.A., Clark, S.H., Woody, C.O. and Rowe, D.W. (1992). Differential utilization of regulatory domains within the $\alpha 1(I)$ collagen promoter in osseous and fibroblastic cells. *J. Cell. Biol.* 116, 227-236.
- Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. *Nature* 322, 697-701.
- Rippe, R.A., Lorenzen, S-I., Brenner, D.A. and Breindl, M. (1989). Regulatory elements in the 5'-flanking region and the first intron contribute to transcriptional control of the mouse alpha 1 type I collagen gene. *Mol. Cell. Biol.* 9, 2224-2227.
- Rossi, P. and De Crombrughe, B. (1987). Identification of a cell-specific transcriptional enhancer in the first intron of the mouse $\alpha 2$ (type I) collagen gene. *Proc. Natl. Acad. Sci. USA* 84, 5590-5594.
- Rossi, P., Karsenty, G., Roberts, A.B., Roche, N.S., Sporn, M.B. and De Crombrughe, B. (1988). A nuclear factor 1 binding site mediates the transcriptional activation of a type 1 collagen promoter by transforming growth factor- β . *Cell* 52, 405-414.
- Rossouw, C.M.S., Vergeer, W.P., Du Plooy, S.J., Bernard, M.P., Ramirez, F. and De Wet, W.J. (1987). DNA sequences in the first intron of the human pro- $\alpha 1(I)$ collagen gene enhance transcription. *J. Biol. Chem.* 262, 15151-15157.
- Rusconi, S. (1991). Transgenic regulation in laboratory animals. *Experientia* 47, 866-877.

Saltzman, A.G. and Weinmann, R. (1989). Promoter specificity and modulation of RNA polymerase II transcription. *FASEB J.* 3, 1723-1733. ✓

Saluz, H.P. and Jost, J.P. (1989). Genomic footprinting with *Taq* polymerase. *Nature* 338, 277.

Sandberg, M. and Vuorio, E. (1987). Localization of types I, II and III collagen mRNAs in developing human skeletal tissues by in situ hybridization. *J. Cell Biol.* 104, 1077-1084.

Sassone-Corsi, P., Sisson, J.S. and Verma, I.M. (1988). Transcriptional autoregulation of the proto-oncogene *fos*. *Nature* 334, 314-319. ✓

Schirm, S., Jiricny, J. and Schaffner, W. (1987). The SV40 enhancer can be dissected into multiple segments, each with a different cell type specificity. *Genes Dev.* 1, 65-74.

Schmidt, A., Rossi, P. and De Crombrughe, B. (1986). Transcriptional control of the mouse $\alpha 2(I)$ collagen gene: Functional deletional analysis of the promoter and evidence for cell-specific expression. *Mol. Cell. Biol.* 6, 347-354.

Schöler, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N. and Gruss, P. (1989). Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *EMBO J.* 8, 2551-2557.

Seipel, K., Georgiev, O. and Schaffner, W. (1992). Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. *EMBO J.* 11, 4961-4968.

Selden, R.F., Howie, K.B., Rowe, M.E., Goodman, H.M. and Moore, D.D. (1986). *Mol. Cell. Biol.* 6, 3173-3179. ✓

Serfling, E., Jasin, M. and Schaffner, W. (1985) Enhancers and eukaryotic gene transcription. *Trends Genet.* 1, 224-230.

Sharp, P.A. (1992). TATA-binding protein is a classless factor. *Cell* 68, 819-821. ✓

Sherwood, A.L. and Bornstein, P. (1990). Transcriptional control of the $\alpha 1(I)$ collagen gene involves orientation- and position-specific intronic sequences. *Biochem. J.* 265, 895-897.

Siebenlist, U. and Gilbert, W. (1980). Contact between *Escherichia coli* RNA polymerase and an early promoter of phage T7. *Proc. Natl. Acad. Sci.* 77, 122-126.

Singer, V.L., Wobbe, C.R. and Struhl, K. (1990). A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes Dev.* 4, 636-645.

Skalnik, D.G., Strauss, E.C. and Orkin, S.H. (1991). CCAAT displacement protein as a repressor of the myelomonocytic-specific gp91-phox gene promoter. *J. Biol. Chem.* 266, 16736-16744.

Slack, J.L., Liska, D.A. and Bornstein, P. (1991). An upstream regulatory region mediates high-level, tissue-specific expression of the human $\alpha 1(I)$ collagen gene in transgenic mice. *Mol. Cell. Biol.* 11, 2066-2074. ✓

Smale, S.T. and Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell* 57, 103-113.

Southern, P.J. and Berg, P. (1982). Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J. Mol. Appl. Gen.* 1, 327-341.

Stinger, K.F., Ingles, C.J. and Greenblatt, J. (1990). Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* 345, 783-786. ✓

Superti-Furga, G., Barberis, A., Schaffner, G. and Busslinger, M. (1988). The -117 mutation in greek HPFH affects the binding of three nuclear factors to the CCAAT region of the γ -globin gene. *EMBO J.* 7, 3099-3107.

Superti-Furga, G., Barberis, A., Schreiber, E. and Busslinger, M. (1989). The protein CDP, but not CP1, footprints on the CCAAT region of the γ -globin gene in unfractionated B-cell extracts. *Biochim. Biophys. Acta* 1007, 237-242.

Treisman, R. (1985). Transient accumulation of *c-fos* RNA following serum stimulation requires a conserved 5' element and *c-fos* 3' sequences. *Cell* 42, 889-902. ✓

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

Vergeer, W.P. (1988). Transfeksie van soogdiersellyne met pendelvektore vir die bestudering van transkripsionele reguleringsvolgordes in die mensgeen vir pro- α 1(I)-kollageen. M.Sc. thesis, PU for CHE. ✓

Vuorio E. and de Crombrughe B. (1990). The family of collagen genes. *Ann. Rev. Biochem.* 59, 837-872.

Wells, R.D., Collier, D.A., Hanvey, J.C., Shimizu, M. and Wohlrab, F. (1988). The chemistry and biology of unusual DNA structures adopted by oligopurine-oligopyrimidine sequences. *FASEB J.* 2, 2939-2949. ✓

Westin, G., Gerster, T., Müller, M.M., Schaffner, G., and Schaffner, W. (1987). OVEC, a versatile system to study transcription in mammalian cells and cell-free extracts. *Nucl. Acids Res.* 17, 6787-6798.

White, R.J. and Jackson, S.P. (1992). The TATA-binding protein: a central role in transcription by RNA polymerases I, II and III. *Trends Genet.* 8, 284-288.

Wu, H., Bateman, J.F., Schnieke, A., Sharpe, A., Barker, D., Mascara, T., Eyre, D., Bruns, R., Krimpenfort, P., Berns, A. and Jaenisch, R. (1990). Human-mouse interspecies collagen I heterotrimer is

functional during embryonic development of Mov13 mutant mouse embryos. *Mol. Cell. Biol.* *10*, 1452-1460.

Zawel, L. and Reinberg, D. (1993). Initiation of transcription by RNA polymerase II: a multi-step process. *Prog. Nucl. Acids Res. Mol. Biol.* *in press*.

ACKNOWLEDGEMENTS

I want to express my gratitude towards the following persons:

- Prof. Dr. Wouter J. de Wet for supervising this study.
- Mimi de Vries for introducing me into the laboratory and teaching me the essentials of experimental molecular biology, as well as providing many of the materials used in this study. I am particularly indebted to her for the large scale preparations of nuclear extracts and her help with the DNA-protein binding assays.
- Charlotte Mienie for establishing the growth hormone assay and assisting me especially in this part of the work.
- All my other colleagues at the Department of Biochemistry, Potchefstroom and the Institute of Molecular Biology II, Zürich for their help, support and friendship.
- Prof. Dr. Walter Schaffner and Dr. Pamela J. Mitchell for providing the means for finishing this study while working under them.
- Fritz Ochsenbein for his expert help with the figures.
- Dr. Christopher Hovens for reading the manuscript and correcting language mistakes.
- My parents, brothers and all the many friends for the love with which they continued to support and encourage me.
- The CSIR for their financial support.